

Descriptive Statistics

Presentation / Grouping of Data

Can be adopted two formats for presenting data sets:

- A data list, which is an explicit listing of all the individual measurements. Display with space between the individual measurements, or in set notation with individual measurements separated by commas.

Case (1) Data obtained by measuring age of 21 randomly selected students enrolled in G230 course could be presented as data list.

18 18 19 19 19 18 22 20 18 18 17 19 18 24 18 20 18 21 20 17 19


OR

18,18,19,19,19,18,22,20,18,18,17,19,18,24,18,20,18,21,20,17,19

1) Frequency Distribution:

A data set can also be presented by data frequency table, in which each *distinct* value x is listed in the first row and its frequency f (which is the number of times the value x appears in the data set) is listed below it in the second row. It is especially convenient when data sets are large and the number of distinct values is not too large.

Can be represent the same data in Case (1) by Frequency Table:



x	17	18	19	20	21	22	24
f	2	8	5	3	1	1	1

Key:

Data sets can be presented either by listing all the elements or by giving a table of values and frequencies.

Case (2) 30 measurements represent random sample of 300 population of sand sediments hand specimen from **specific geologic outcrop (Already known the size range of sand)**. Need to organize the following data by frequency and Relative distribution.

Fine Very Coarse Very Coarse Coarse Medium Coarse Coarse Very Fine Medium Fine
 Medium Fine Very Coarse Coarse Fine Coarse Medium Coarse Fine Medium
 Very Coarse Coarse Very Fine Medium Coarse Very Fine Coarse Fine Very Coarse Coarse

Solution:

As the data above are categorical (qualitative), discrete classes can be used (In the column of the Class Limit. Five classes of groups can be distributed in the frequency table.

Class / Class Limit	Frequency	Relative Frequency (%)
Very Coarse	5	0.166 ≈ 17%
Coarse	10	0.333 ≈ 33%
Medium	6	0.200 ≈ 20%
Fine	6	0.200 ≈ 20%
Very Fine	3	0.100 ≈ 10%
	30	1 ≈ 100%

2) Relative Frequency Distribution:

The frequency divided by the total frequency. This gives the percent of values falling in that class.

By adopting the frequency and relative distribution can be recognize that the predominant class of sediments is **Coarse which has 10 frequencies and 33% from total.**

Case (3) 50 measurements represent random sample of rainfall monthly averages in mm collected from Basra Meteorological Record Station along last 12 years. Need to organize the following data by frequency distribution and Relative Frequency.

112 100 127 120 134 118 105 110 109 112 110 118 114 118 117 120 111
 117 116 118 112 114 114 105 109 107 112 114 115 114 110 120 111 119
 118 117 118 122 102 110 116 108 110 121 113 120 112 105 113 104

Solution:

- Before make a table, need to sort the raw data (values) ascendingly to find the minimum (100) and the maximum (134) values.
- To find Range by maximum value subtracts minimum: (R = 134 - 100 = 34).
- To find number of classes: Number of Classes = 1+(3.3*logN)
 No. of Classes = 1+(3.3*log50) = 6.6 ≈ 7 (must rounded to integer number).
- To find Class Width: Class Width = $\frac{\text{Range}}{\text{No.of Classes}} = \frac{34}{7} = 4.9 \approx 5$
- Then select a starting point (The lowest value or any convenient number less than the lowest value); add the width to get lower limits.

Class / Class Limit	Frequency	R. Frequency (%)
100 – 104	2	0.040 ≈ 4%
105 – 109	8	0.160 ≈ 16%
110 – 114	18	0.360 ≈ 36%
115 – 119	13	0.260 ≈ 26%
120 – 124	7	0.140 ≈ 14%
125 – 129	1	0.020 ≈ 2%
130 – 134	1	0.020 ≈ 2%
	50	1 ≈ 100%

As data above is quantitative (numerical) and continuous, then can be make the limit of class (From minimum value to maximum using Class Width between each class/group. Can be recognize predominant class (110 – 114) of rainfall (mm) along last 12 years.

➤ Can be adopted the data used in the Case (2) above to make **ungrouped** frequency distribution:

Observation (X)	Frequency (f)	Relative Frequency (%)	Observation (X)	Frequency (f)	Relative Frequency (%)
Very Coarse	1	0.0333 ≈ 3%	Medium	1	0.0333 ≈ 3%
Very Coarse	1	0.0333 ≈ 3%	Medium	1	0.0333 ≈ 3%
Very Coarse	1	0.0333 ≈ 3%	Medium	1	0.0333 ≈ 3%
Very Coarse	1	0.0333 ≈ 3%	Medium	1	0.0333 ≈ 3%
Very Coarse	1	0.0333 ≈ 3%	Medium	1	0.0333 ≈ 3%
Coarse	1	0.0333 ≈ 3%	Medium	1	0.0333 ≈ 3%
Coarse	1	0.0333 ≈ 3%	Fine	1	0.0333 ≈ 3%
Coarse	1	0.0333 ≈ 3%	Fine	1	0.0333 ≈ 3%
Coarse	1	0.0333 ≈ 3%	Fine	1	0.0333 ≈ 3%
Coarse	1	0.0333 ≈ 3%	Fine	1	0.0333 ≈ 3%
Coarse	1	0.0333 ≈ 3%	Fine	1	0.0333 ≈ 3%
Coarse	1	0.0333 ≈ 3%	Fine	1	0.0333 ≈ 3%
Coarse	1	0.0333 ≈ 3%	Very Fine	1	0.0333 ≈ 3%
Coarse	1	0.0333 ≈ 3%	Very Fine	1	0.0333 ≈ 3%
Coarse	1	0.0333 ≈ 3%	Very Fine	1	0.0333 ≈ 3%
30 of Observations				30	1 ≈ 100%

➤ Also, can be adopted the data used in the Case (3) above to make **ungrouped** frequency distribution:

Observation (X)	Frequency (f)	Relative Frequency (%)	Observation (X)	Frequency (f)	Relative Frequency (%)	Observation (X)	Frequency (f)	Relative Frequency (%)
100	1	0.020 ≈ 2%	112	1	0.020 ≈ 2%	118	1	0.020 ≈ 2%
102	1	0.020 ≈ 2%	112	1	0.020 ≈ 2%	118	1	0.020 ≈ 2%
104	1	0.020 ≈ 2%	112	1	0.020 ≈ 2%	118	1	0.020 ≈ 2%
105	1	0.020 ≈ 2%	112	1	0.020 ≈ 2%	118	1	0.020 ≈ 2%
105	1	0.020 ≈ 2%	113	1	0.020 ≈ 2%	118	1	0.020 ≈ 2%
105	1	0.020 ≈ 2%	113	1	0.020 ≈ 2%	118	1	0.020 ≈ 2%
107	1	0.020 ≈ 2%	114	1	0.020 ≈ 2%	119	1	0.020 ≈ 2%
108	1	0.020 ≈ 2%	114	1	0.020 ≈ 2%	120	1	0.020 ≈ 2%
109	1	0.020 ≈ 2%	114	1	0.020 ≈ 2%	120	1	0.020 ≈ 2%
109	1	0.020 ≈ 2%	114	1	0.020 ≈ 2%	120	1	0.020 ≈ 2%
110	1	0.020 ≈ 2%	114	1	0.020 ≈ 2%	120	1	0.020 ≈ 2%
110	1	0.020 ≈ 2%	115	1	0.020 ≈ 2%	121	1	0.020 ≈ 2%
110	1	0.020 ≈ 2%	116	1	0.020 ≈ 2%	122	1	0.020 ≈ 2%
110	1	0.020 ≈ 2%	116	1	0.020 ≈ 2%	122	1	0.020 ≈ 2%
110	1	0.020 ≈ 2%	117	1	0.020 ≈ 2%	127	1	0.020 ≈ 2%
111	1	0.020 ≈ 2%	117	1	0.020 ≈ 2%	134	1	0.020 ≈ 2%
111	1	0.020 ≈ 2%	117	1	0.020 ≈ 2%	50 Observations	50	1 ≈ 100%

3) Cumulative Frequency / Relative (Percentage) Distribution:

- ❖ Cumulative Frequency Distribution shows the total frequency through the upper real limit of each class.
- ❖ Cumulative Relative (Percentage) Distribution shows the total percentage through the upper real limit of each class.

It has been adopted the data from Case (2) and Case (3) above to represent the Cumulative and Percentage Frequency Distributions.

Class / Class Limit	Frequency	Cumulative	Cumulative Percentage (%)
Very Coarse	5	5	17
Coarse	10	15	50
Medium	6	21	70
Fine	6	27	90
Very Fine	3	30	100
	30		

Class / Class Limit	Frequency	Cumulative	Cumulative Percentage (%)
100 – 104	2	2	4
105 – 109	8	10	20
110 – 114	18	28	56
115 – 119	13	41	82
120 – 124	7	48	96
125 – 129	1	49	98
130 – 134	1	50	100
	50		

EXERCISES

1. List all the measurements for the data set represented by the following data frequency table.

x	31	32	33	34	35
f	1	5	6	4	2

2. List all the measurements for the data set represented by the following data frequency table.

x	97	98	99	100	101	102	103	105
f	7	5	3	4	2	2	1	1

3. Construct the data frequency table for the following data set.

22 25 22 27 24 23
26 24 22 24 26

4. Construct the data frequency table for the following data set.

$\{1,5,2,3,5,1,4,4,4,3,2,5,1,3,2,$
 $1,1,1,2\}$

ANSWERS

1. $\{31,32,32,32,32,32,33,33,33,33,33,33,34,34,34,34,35,35\}$.

- 3.

x	22	23	24	25	26	27
f	3	1	3	1	2	1

Break

Who used the word statistics for first time??

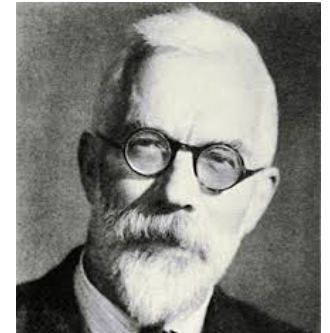
- The **word 'Statistics'** was **first** introduced by a German Scholar, Gottfried Achenwall (1719 – 1772), in the middle of the 18th century. From the very name, it is felt that it must be related to the administrative functioning of a state supplying facts regularly and quantitatively regarding its various fields of administration.



Gottfried Achenwall

Who is father of statistics??

- Sir Ronald Fisher. Touted as the greatest scientist of his time, Sir Ronald Fisher (1890 - 1962) was a British statistician and biologist who was known for his contributions to experimental design and population genetics. He is known as the **father** of modern **statistics** and experimental design.



Sir Ronald Fisher

Display / Graphing of Data

There are many effective ways to present /display data graphically that gives a sense of data distribution across the range of values that appear. Three graphical representations of sets of data: stem and leaf diagram, frequency histogram, relative frequency histogram and others.

❑ Stem and Leaf Diagram

Case (4) Suppose 30 students in a statistics class took a test and made the following scores:

Q/ How did the class do on the test? A quick glance at the set of 30 numbers does not immediately give a clear answer!!

86		80	25	77	73	76	100	90	69	93
90		83	70	73	73	70	90	83	71	95
40		58	68	69	100	78	87	97	92	74

Original Data

Solution:

By adopting a stem and leaf diagram which provide a quick display / easy visibility interpretation of how the data are distributed across the range of their values.

Next slide shows how to produce/organize it.

- The numbers in the tens place, from 2 through 9, and number 10, are the "stems," and are arranged in numerical order from top to bottom to the left of a vertical line (Step 1).

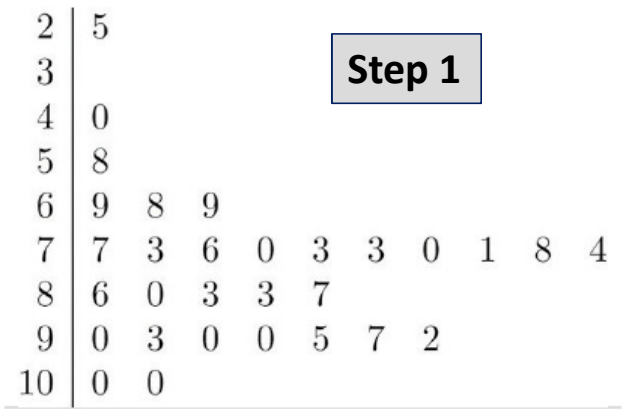
- The number in the units (Ones) place in each measurement is a "leaf," and is placed in a row to the right of the corresponding stem (Step 1).

- Thus the three leaves 9, 8, and 9 in the row headed with stem 6 correspond to the three exam scores in the 60s: (69 in the first row of data), (68 and 69 in the third row), (Step 1).

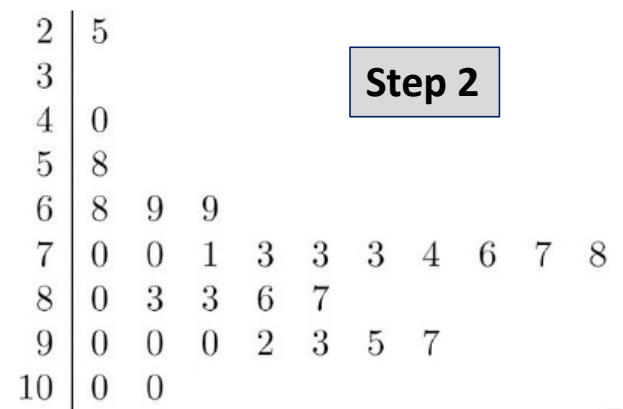
- The display is made even more useful for some purposes by rearranging the leaves in numerical order, as shown in Step 2 "Ordered Stem and Leaf Diagram". There are two perfect scores; three students made scores under 60; most students scored in the 70s, 80s and 90s; and the overall average is probably in the high 70s or low 80s.

- Can be spread the diagram out by splitting each tens place number into lower and upper categories. For example, all the scores in the 80s may be represented on two separate stems, lower 80s and upper 80s:

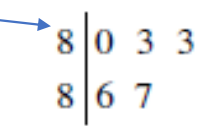
➤ Note that all of the original data can be recovered from the stem and leaf diagram.



Step 1



Step 2



Splitting

❑ Frequency Histogram

The stem and leaf diagram is not practical for large data sets, so we need a different, purely graphical way to represent data.

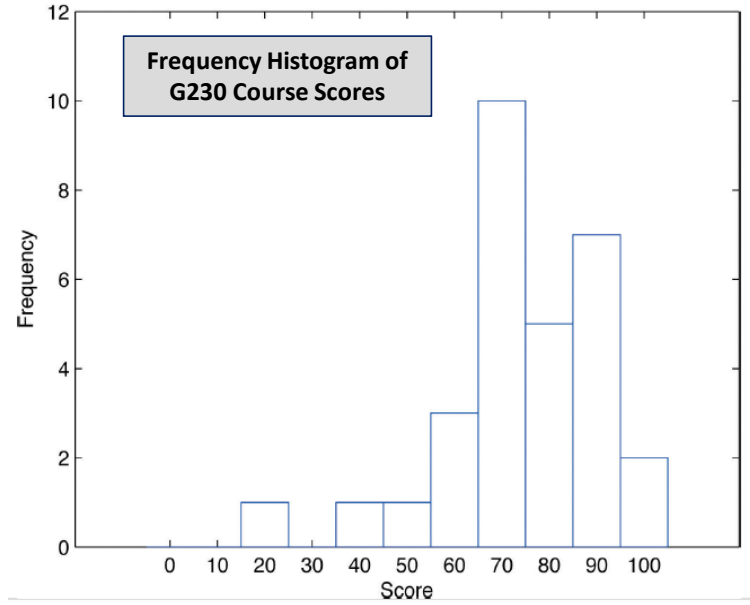
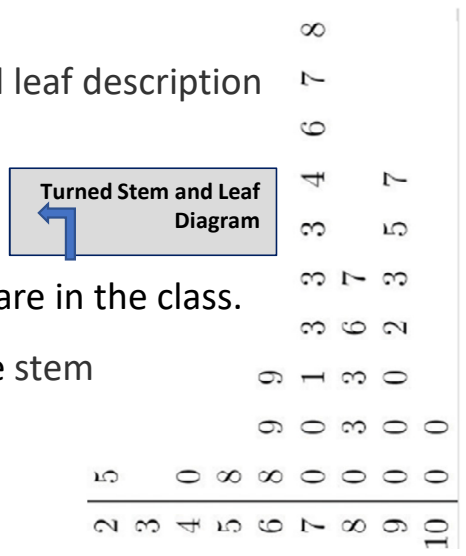
Frequency Histogram is a graphical technique showing how data are grouped or/and distributed across the range of their values by collecting them into classes and indicating the number of measurements in each class. Can be adopted the discrete or continuous data in histogram. It is visual interpretation for large data set and show the gaps as well as determine Median.

- It is similar to a vertical bar graph. However, a histogram, unlike a vertical bar graph, shows no gaps between the bars.
- Finally the Histogram is a form of Bar Graph used with interval or ratio scaled variables.

- Can be used the same data adopted in stem and leaf description to show the histogram construction.

- Frequency got from a class of measurements / the number of measurements in the data set that are in the class.

➤ Noted the similarity between Histogram and the stem and leaf diagram if turned the last one left 90 degrees / quarter counterclockwise.



❑ Relative Frequency Histogram

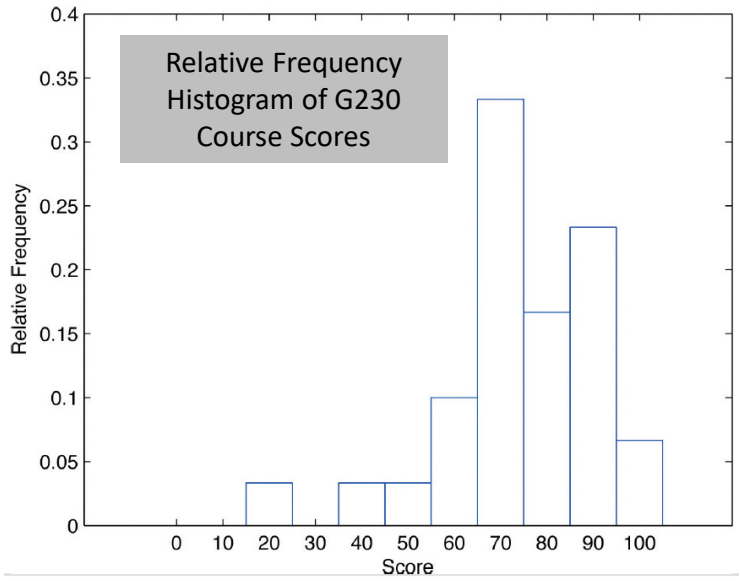
The frequency of each class/group divided by the total frequency. This gives the percent of values falling in that class. We can thus construct a diagram by drawing for each group/class, a vertical bar whose length is the relative frequency of that group.

It is exactly the same as the frequency histogram except that the vertical axis in the relative frequency histogram is not frequency but relative frequency.

✓ Relative frequency histogram is important because the heights can be interpreted as probabilities. These probability histograms provide a graphical display of a probability distribution, which can be used to determine the likelihood of certain results to occur within a given population.

✓ The relative frequency histogram is important because the labeling on the vertical axis reflects what is important visually: the relative sizes of the bars.

✓ A key point is that now if each vertical bar has width 1 unit, then the total area of all the bars is 1 or 100%.

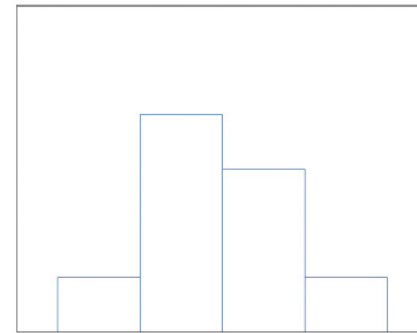
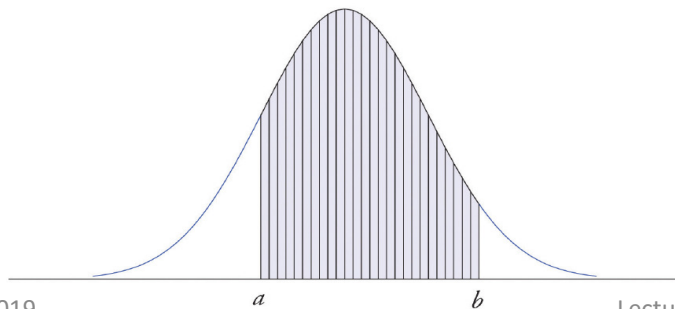


✓ When the size of a sample is small only a few classes can be used in constructing a relative frequency histogram. Very large sample has very fine relative frequency histogram. If the sample size were to increase indefinitely then the relative frequency histogram would be so fine that it would look like a smooth curve especially in population/very large data set.

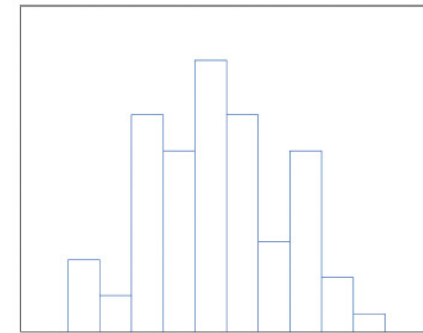
✓ When the curve is actually just a very fine relative frequency histogram in which the exceedingly narrow vertical bars have disappeared. Because the area of each such vertical bar is the proportion of the data that lies in the interval of numbers over which that bar stands, this means that for any two numbers

a and b , the proportion of the data that lies between the two numbers a and b is the area under the curve that is above the interval (a,b) in the horizontal axis. In particular the total area under the curve is 1, or 100%.

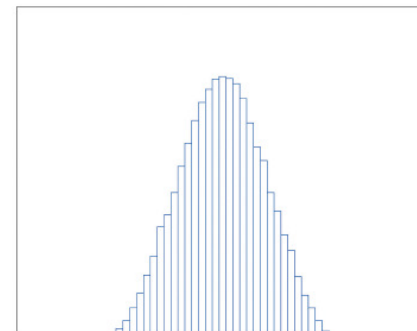
Shaded Area = Proportion of Data between a and b



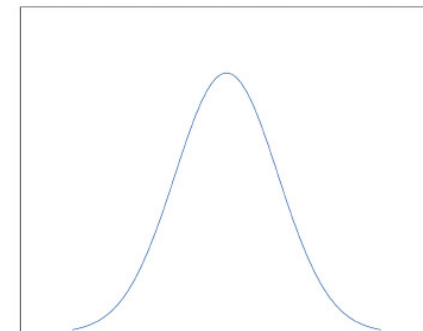
(a) Small Sample



(b) Medium Sample



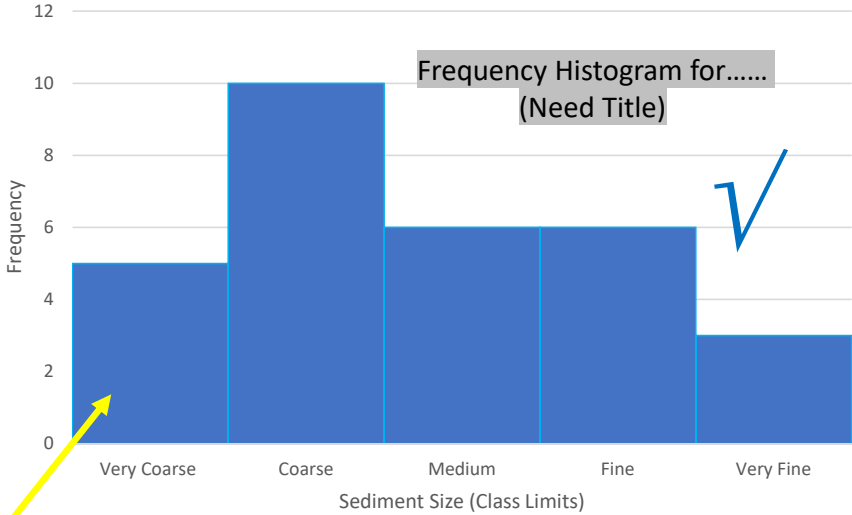
(c) Large Sample



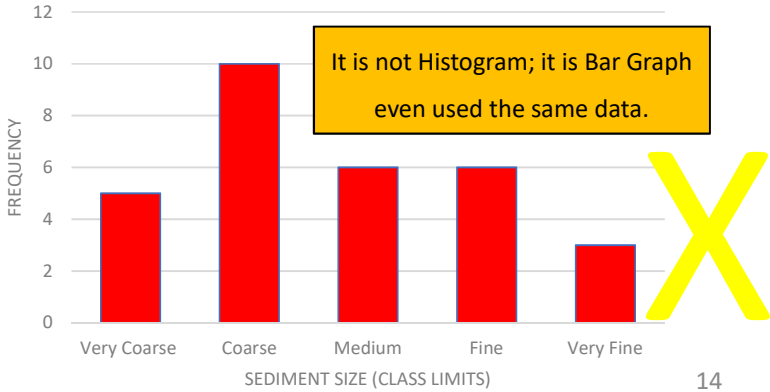
(d) Very Large Sample

Can be adopted the frequency table from Case (2) to draw the histogram and frequency histogram:

Class / Class Limit	Frequency	R. Frequency (%)
Very Coarse	5	0.166 ≈ 17%
Coarse	10	0.333 ≈ 33%
Medium	6	0.200 ≈ 20%
Fine	6	0.200 ≈ 20%
Very Fine	3	0.100 ≈ 10%

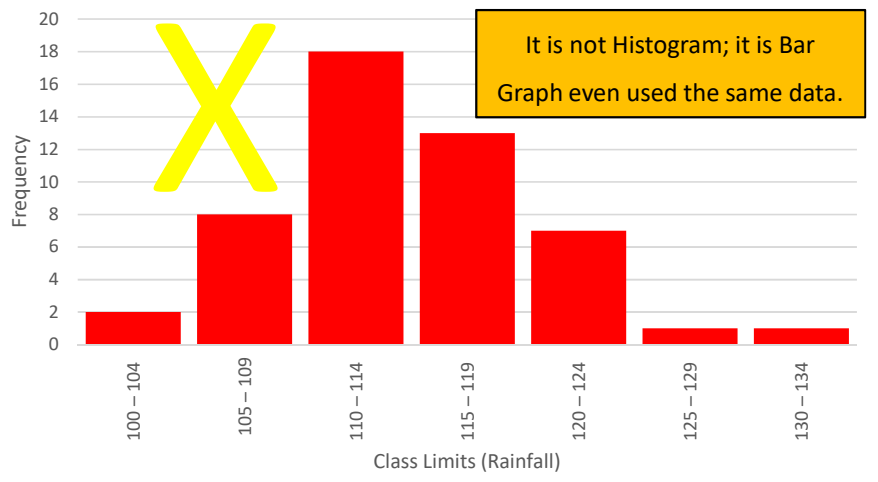
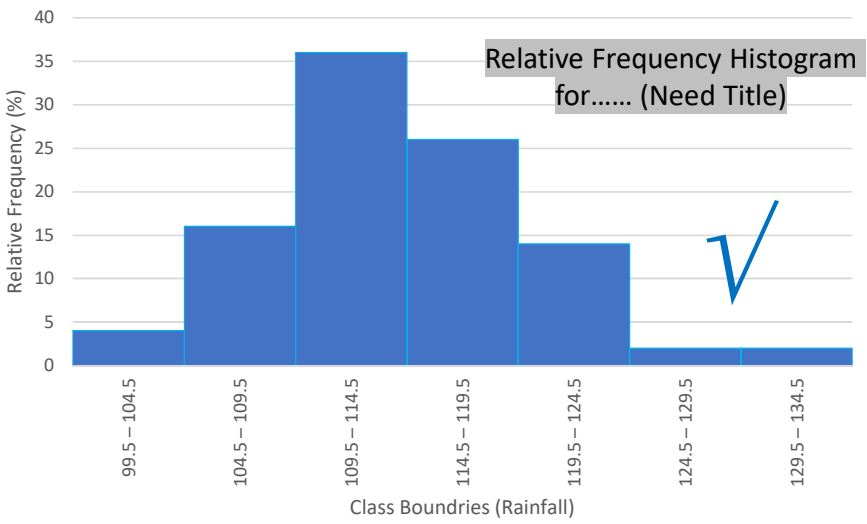
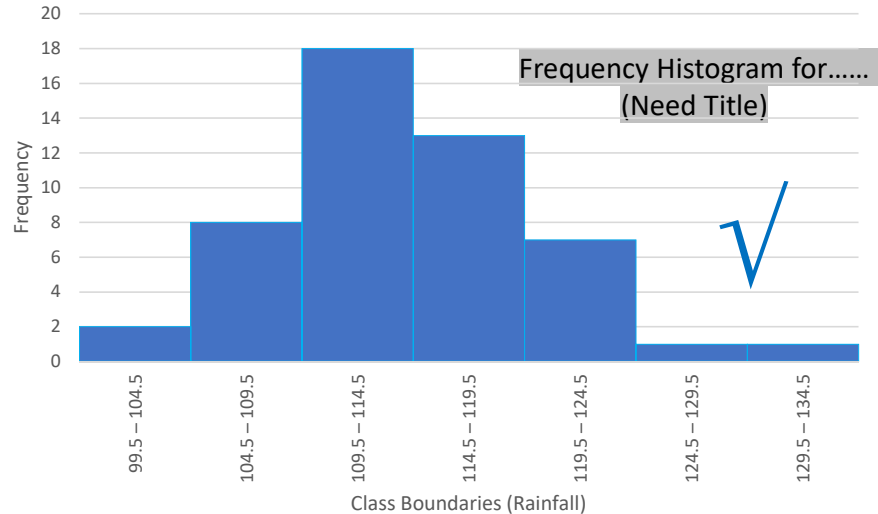


Even the data above is discrete/qualitative, the histogram is presented continually bars. It such as a vertical bar graph, but with out gaps between the bars.



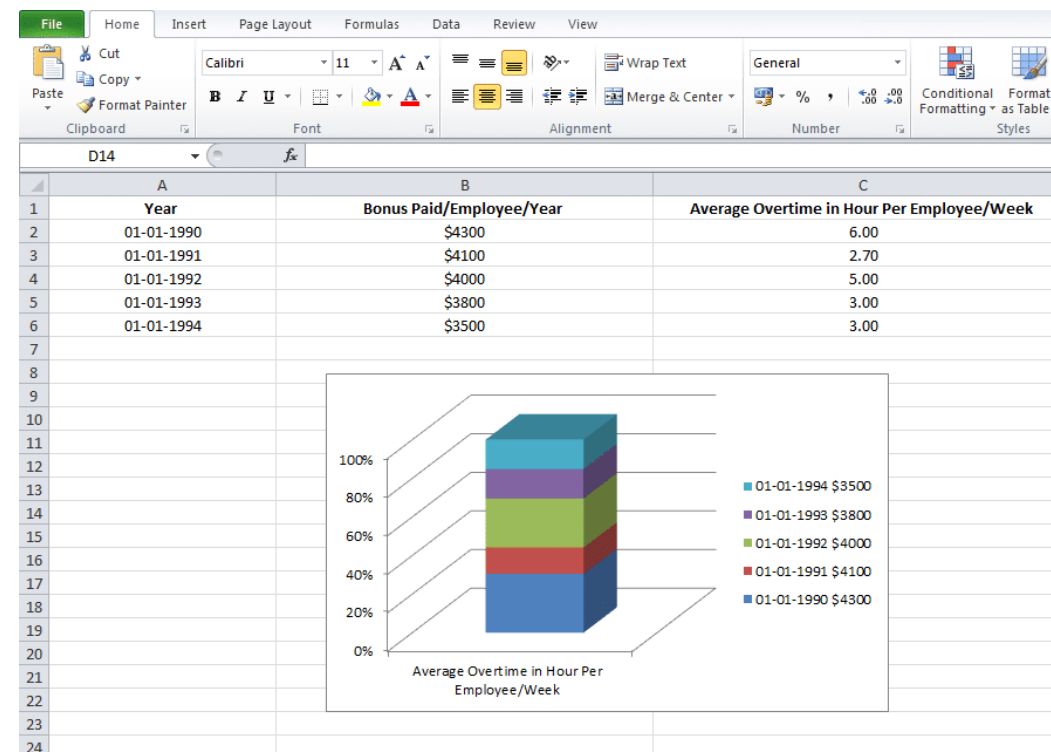
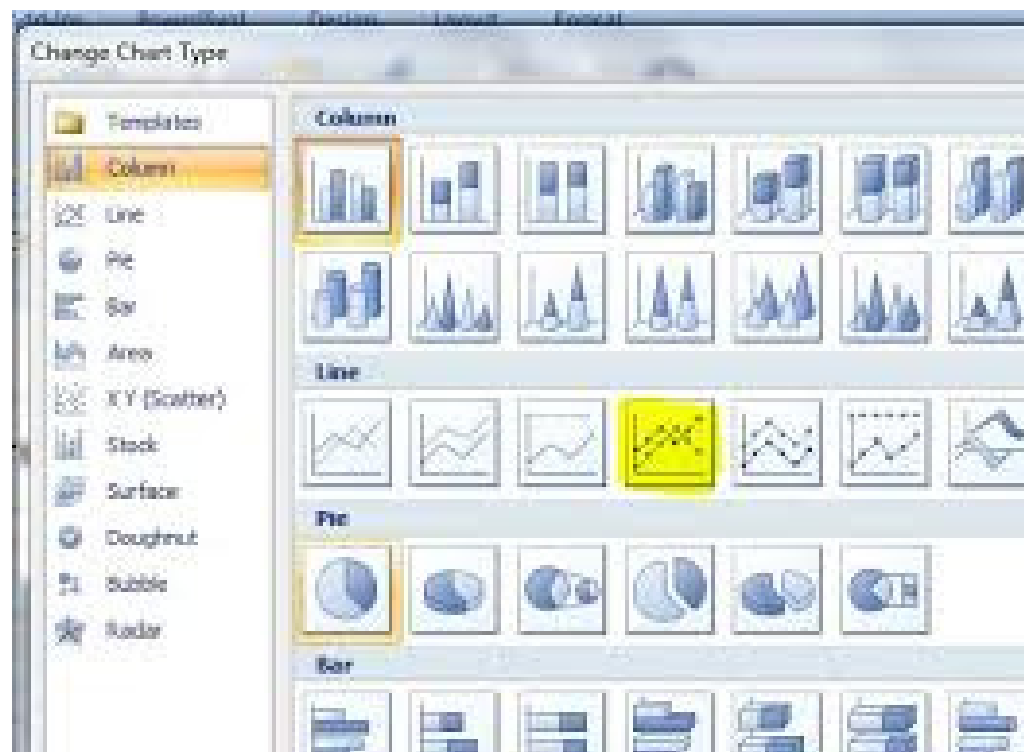
Also can be used the frequency table from Case (3) to draw the frequency histogram:

Class / Class Limit	Class Boundaries	Frequency	R. Frequency (%)
100 – 104	99.5 – 104.5	2	0.040 ≈ 4%
105 – 109	104.5 – 109.5	8	0.160 ≈ 16%
110 – 114	109.5 – 114.5	18	0.360 ≈ 36%
115 – 119	114.5 – 119.5	13	0.260 ≈ 26%
120 – 124	119.5 – 124.5	7	0.140 ≈ 14%
125 – 129	124.5 – 129.5	1	0.020 ≈ 2%
130 – 134	129.5 – 134.5	1	0.020 ≈ 2%



Break

Excel offers three varieties of graphs: Line Graphs: Both 2 dimensional and three dimensional line graphs are available in all the versions of Microsoft Excel.



Frequency Polygon

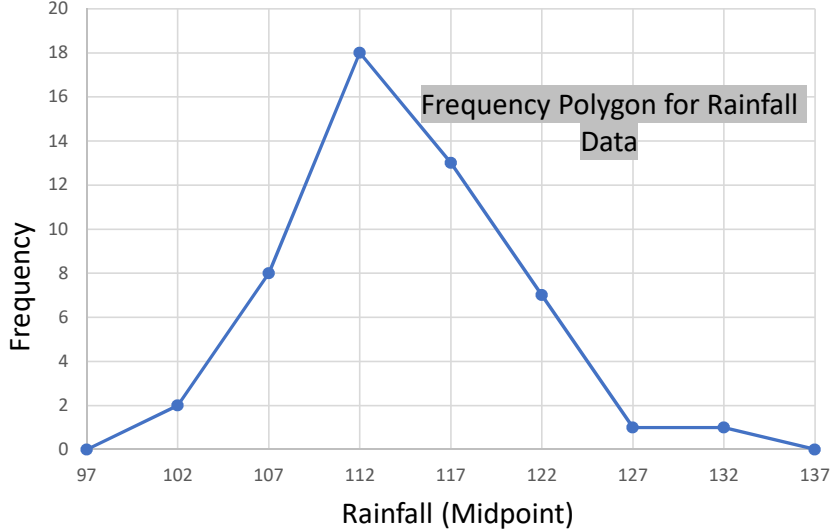
A line graph. The frequency is placed along the vertical axis and the class midpoints are placed along the horizontal axis. These points are connected with lines. It is used to understand the shapes of distributions and serve the same purpose as histograms, but it is helpful for comparing the data by overlaying the frequency polygons drawn for different data sets.

Once again can be used the data from Case (3) above to draw the Frequency Polygon...

Lower Limit Class	Upper Limit Class	Midpoint (Class Mark)	Frequency	Lower Class Boundary	Upper Class Boundary	Midpoint (Class Mark)
100	104	102	2	99.5	104.5	102
105	109	107	8	104.5	109.5	107
110	114	112	18	109.5	114.5	112
115	119	117	13	114.5	119.5	117
120	124	122	7	119.5	124.5	122
125	129	127	1	124.5	129.5	127
130	134	132	1	129.5	134.5	132

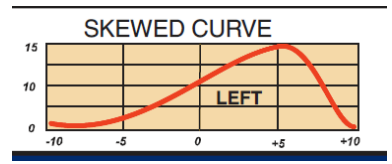
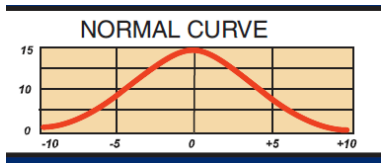
$100 + 104 / 2 = 102$

$99.5 + 104.5 / 2 = 102$



➤ Midpoint can be found by adding the upper and lower class limits and dividing by two. It can also be found by adding the upper and lower boundaries and dividing by two.

✓ Good to mention the Normal Frequency Curve and Skewed Curve



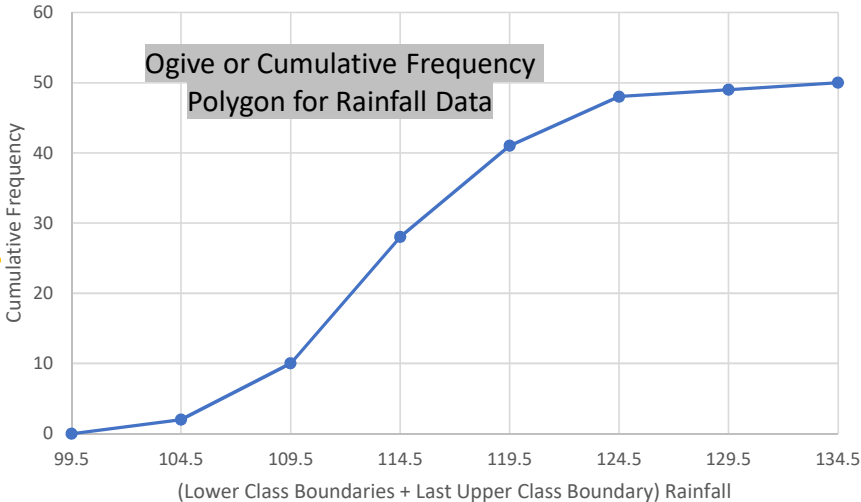
Ogive

A frequency polygon of the cumulative frequency or the relative cumulative frequency. The vertical axis the cumulative frequency or relative cumulative frequency. The horizontal axis is the class boundaries. The graph always starts at zero at the lowest class boundary and will end up at the total frequency (for a cumulative frequency) or 1.00 (for a relative cumulative frequency). The same data used in the Frequency Polygon will be adopted to draw the Ogive...

Lower Class Boundary	Upper Class Boundary	Frequency	Cumulative	Relative Cumulative	Cumulative Percentage (%)
99.5	104.5	2	2	0.04	4
104.5	109.5	8	10	0.2	20
109.5	114.5	18	28	0.56	56
114.5	119.5	13	41	0.82	82
119.5	124.5	7	48	0.96	96
124.5	129.5	1	49	0.98	98
129.5	134.5	1	50	1	100

Lower Class Boundaries + Last Upper Class Boundary	Cumulative	Relative Cumulative	Cumulative Percentage (%)
99.5	0	0	0
104.5	2	0.04	4
109.5	10	0.2	20
114.5	28	0.56	56
119.5	41	0.82	82
124.5	48	0.96	96
129.5	49	0.98	98
134.5	50	1	100

Ogives are useful for comparing distributions. This is achieved by overlaying the frequency polygons of cumulative frequencies drawn for different data sets.



Distributions of a Histogram

- A normal distribution: In a normal distribution, points on one side of the average are as likely to occur as on other side of the average.
- A bimodal distribution: In a bimodal distribution, there are two peaks. In a bimodal distribution, the data should be separated and analyzed as separate normal distributions.
- A right-skewed distribution: Called also a positively skewed distribution. In a right-skewed distribution, a large number of the data values occur on the left side with a fewer number of data values on the right side. A right-skewed distribution usually occurs when the data has a range boundary on the left-hand side of the histogram. For example, a boundary of 0.
- A left-skewed distribution: Called also a negatively skewed distribution. In a left-skewed distribution, a large number of the data values occur on the right side with a fewer number of data values on the left side. A right-skewed distribution usually occurs when the data has a range boundary on the right-hand side of the histogram. For example, a boundary such as 100.
- A random distribution: A random distribution lacks an apparent pattern and has several peaks. In a random distribution histogram, it can be the case that different data properties were combined. So, the data should be separated/analyzed separately



20 October 2019

Lecture # 2 - Dr. Mazin A.

19

Keys:

- **Raw Data:** Data collected in original form.
- **Frequency:** The number of times a certain value or class of values occurs.
- **Frequency Distribution:** The organization of raw data in table form with classes and frequencies.
- **Categorical Frequency Distribution:** A frequency distribution in which the data is only nominal or ordinal.
- **Ungrouped Frequency Distribution:** A frequency distribution of numerical data. The raw data is not grouped.
- **Grouped Frequency Distribution:** A frequency distribution where several numbers are grouped into one class.
- **Class Limits:** Separate one class in a grouped frequency distribution from another. The limits could actually appear in the data and have gaps between the upper limit of one class and the lower limit of the next.
- **Class Limits** are values halfway **between** the upper **class boundary** of one **class** and the lower **class boundary** of the next. ...
- **Class Boundaries:** Separate one class in a grouped frequency distribution from another. The boundaries have one more decimal place than the raw data and therefore do not appear in the data. There is no gap between the upper boundary of one class and the lower boundary of the next class. The lower class boundary is found by subtracting 0.5 units from the lower class limit and the upper class boundary is found by adding 0.5 units to the upper class limit.
- ✓ **Class boundaries** are values halfway **between** the upper **class limit** of one **class** and the lower **class limit** of the next. **Class boundaries** are possible data values.

- **Class Width:** Difference between upper and lower boundaries of any class. Class width is also the difference between lower limits of two consecutive classes or upper limits of two consecutive classes. It is not difference between upper and lower limits of the same class.
- **Class Mark (Midpoint):** The number in the middle of the class. It is found by adding the upper and lower limits and dividing by two. It can also be found by adding the upper and lower boundaries and dividing by two.

There are other graph techniques like:

- Pareto Chart: A bar graph for qualitative data with the bars arranged according to frequency.
- Pie Chart: Graphical depiction of data as slices of a pie. The frequency determines the size of the slice. The number of degrees in any slice is the relative frequency times 360 degrees.
- Pictograph: A graph that uses pictures to represent data.
- Interval Scale: A quantitative scale that permits the use of arithmetic operations. The zero point in the scale is arbitrary.
- Ratio Scale: Same as interval scale except that there is a true zero point.
- ✓ Graphical representations of large data sets provide a quick overview of the nature of the data.
- ✓ A population or a very large data set may be represented by a smooth curve. This curve is a very fine relative frequency histogram in which the exceedingly narrow vertical bars have been omitted.
- ✓ When a curve derived from a relative frequency histogram is used to describe a data set, the proportion of data with values between two numbers a and b is the area under the curve between a and b .

EXERCISES

BASIC

- Describe one difference between a frequency histogram and a relative frequency histogram.
- Describe one advantage of a stem and leaf diagram over a frequency histogram.
- Construct a stem and leaf diagram, a frequency histogram, and a relative frequency histogram for the following data set. For the histograms use classes 51–60, 61–70, and so on.

69 92 68 77 80
70 85 88 85 96

93 75 76 82 100
53 70 70 82 85

- Construct a stem and leaf diagram, a frequency histogram, and a relative frequency histogram for the following data set. For the histograms use classes 6.0–6.9, 7.0–7.9, and so on.

8.5 8.2 7.0 7.0 4.9
6.5 8.2 7.6 1.5 9.3

9.6 8.5 8.8 8.5 8.7
8.0 7.7 2.9 9.2 6.9

- A data set contains $n = 10$ observations. The values x and their frequencies f are summarized in the following data frequency table.

x	-1	0	1	2
f	3	4	2	1

Construct a frequency histogram and a relative frequency histogram for the data set.

- A data set contains the $n = 20$ observations. The values x and their frequencies f are summarized in the following data frequency table.

x	-1	0	1	2
f	3	a	2	1

The frequency of the value 0 is missing. Find a and then sketch a frequency histogram and a relative frequency histogram for the data set.

- A data set has the following frequency distribution table:

x	1	2	3	4
f	3	a	2	1

The number a is unknown. Can you construct a frequency histogram? If so, construct it. If not, say why not.

- A table of some of the relative frequencies computed from a data set is

x	1	2	3	4
f/n	0.3	p	0.2	0.1

The number p is yet to be computed. Finish the table and construct the relative frequency histogram for the data set.

APPLICATIONS

- The IQ scores of ten students randomly selected from an elementary school are given.

108 100 99 125 87
105 107 105 119 118

Grouping the measures in the 80s, the 90s, and so on, construct a stem and leaf diagram, a frequency histogram, and a relative frequency histogram.

- The IQ scores of ten students randomly selected from an elementary school for academically gifted students are given.

133 140 152 142 137
145 160 138 139 138

Grouping the measures by their common hundreds and tens digits, construct a stem and leaf diagram, a frequency histogram, and a relative frequency histogram.

11. During a one-day blood drive 300 people donated blood at a mobile donation center. The blood types of these 300 donors are summarized in the table.

Blood Type	<i>O</i>	<i>A</i>	<i>B</i>	<i>AB</i>
Frequency	136	120	32	12

Construct a relative frequency histogram for the data set.

12. In a particular kitchen appliance store an electric automatic rice cooker is a popular item. The weekly sales for the last 20 weeks are shown.

20 15 14 14 18

15 17 16 16 18

15 19 12 13 9

19 15 15 16 15

Construct a relative frequency histogram with classes 6–10, 11–15, and 16–20.

ADDITIONAL EXERCISES

13. Random samples, each of size $n = 10$, were taken of the lengths in centimeters of three kinds of commercial fish, with the following results:

Sample 1: 108 100 99 125 87

105 107 105 119 118

Sample 2: 133 140 152 142 137

145 160 138 139 138

Sample 3: 82 60 83 82 82

74 79 82 80 80

Grouping the measures by their common hundreds and tens digits, construct a stem and leaf diagram, a frequency histogram, and a relative frequency histogram for each of the samples. Compare the histograms and describe any patterns they exhibit.

14. During a one-day blood drive 300 people donated blood at a mobile donation center. The blood types of these 300 donors are summarized below.

Blood Type	<i>O</i>	<i>A</i>	<i>B</i>	<i>AB</i>
Frequency	136	120	32	12

Identify the blood type that has the highest relative frequency for these 300 people. Can you conclude that the blood type you identified is also most common for all people in the population at large? Explain.

15. In a particular kitchen appliance store, the weekly sales of an electric automatic rice cooker for the last 20 weeks are as follows.

20 15 14 14 18

15 17 16 16 18

15 19 12 13 9

19 15 15 16 15

In retail sales, too large an inventory ties up capital, while too small an inventory costs lost sales and customer satisfaction. Using the relative frequency histogram for these data, find approximately how many rice cookers must be in stock at the beginning of each week if

- the store is not to run out of stock by the end of a week for more than 15% of the weeks; and
- the store is not to run out of stock by the end of a week for more than 5% of the weeks.

ANSWERS

1. The vertical scale on one is the frequencies and on the other is the relative frequencies.

3.

5	3
6	8 9
7	0 0 0 5 6 7
8	0 2 3 5 5 5 8
9	2 3 6
10	0

Frequency and relative frequency histograms are similarly generated.

5. Noting that $n = 10$ the relative frequency table is:

x	-1	0	1	2
f/n	0.3	0.4	0.2	0.1

7. Since n is unknown, a is unknown, so the histogram cannot be constructed.

9.

8	7
9	9
10	0 5 5 7 8
11	8 9
12	5

Frequency and relative frequency histograms are similarly generated.

11. Noting $n = 300$, the relative frequency table is therefore:

Blood Type	O	A	B	AB
f/n	0.4533	0.4	0.1067	0.04

A relative frequency histogram is then generated.

13. The stem and leaf diagrams listed for Samples 1, 2, and 3 in that order.

6	
7	
8	7
9	9
10	0 5 5 7 8
11	8 9
12	5
13	
14	
15	
16	
6	
7	
8	
9	
10	
11	
12	
13	3 7 8 8 9
14	0 2 5
15	2
16	0

6	0
7	4 9
8	0 0 2 2 2 2 3
9	
10	
11	
12	
13	
14	
15	
16	

The frequency tables are given below in the same order.

Length	80 ~ 89	90 ~ 99	100 ~ 109
<i>f</i>	1	1	5
Length	110 ~ 119	120 ~ 129	
<i>f</i>	2	1	
Length	130 ~ 139	140 ~ 149	150 ~ 159
<i>f</i>	5	3	1
Length	160 ~ 169		
<i>f</i>	1		
Length	60 ~ 69	70 ~ 79	80 ~ 89
<i>f</i>	1	2	7

The relative frequency tables are given below in the same order.

Length	80 ~ 89	90 ~ 99	100 ~ 109
<i>f / n</i>	0.1	0.1	0.5
Length	110 ~ 119	120 ~ 129	
<i>f / n</i>	0.2	0.1	

Length	130 ~ 139	140 ~ 149	150 ~ 159
<i>f / n</i>	0.5	0.3	0.1
Length	160 ~ 169		
<i>f / n</i>	0.1		
Length	60 ~ 69	70 ~ 79	80 ~ 89
<i>f / n</i>	0.1	0.2	0.7

- a. 19.
- b. 20.