

Chapter 2

- I. Organizing Data in Tables
- II. Describing Data by Graphs

I. Tables:

1. Frequency Distribution (Nominal or Ordinal)
2. Grouped Frequency Distribution (Interval or Ratio data)
3. Joint Frequency Distribution for Two Variables (Qualitative or Quantitative)

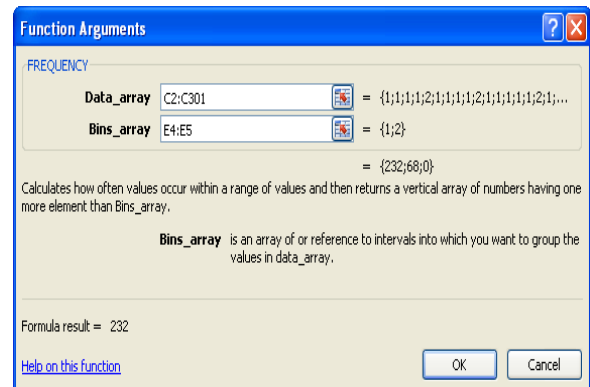
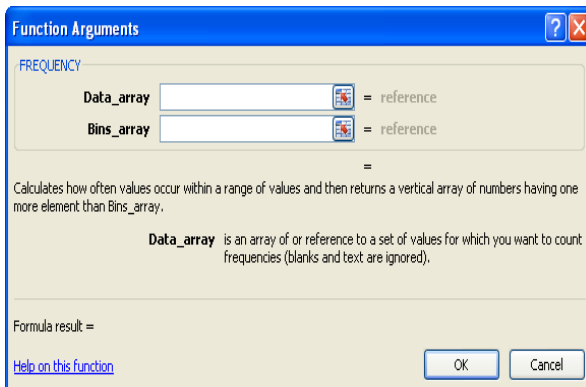
The Excel functions, Excel Analysis ToolPak Add-ins or Excel PHStat2 Add-ins needed to create frequency distributions are:

1. Frequency Distribution (Nominal or Ordinal)

- A. ***FREQUENCY***: To construct a Frequency Distribution for numeric discrete values such as the number of magazines sold, number of absent students, number of children in a family, or numeric code such as (1=Male, 2=Female)

- a. *List the possible values of the variable under study (the bins).*
- b. *Select the cells to contain the frequency values.*
- c. *Click Formulas Tab → Functions → Statistical → FREQUENCY*
- d. *Select the Data array*
- e. *Select the Bins array*
- f. *Press (Control +Shift) + Enter*

Note: You have to hold the (Control +Shift) then press ENTER

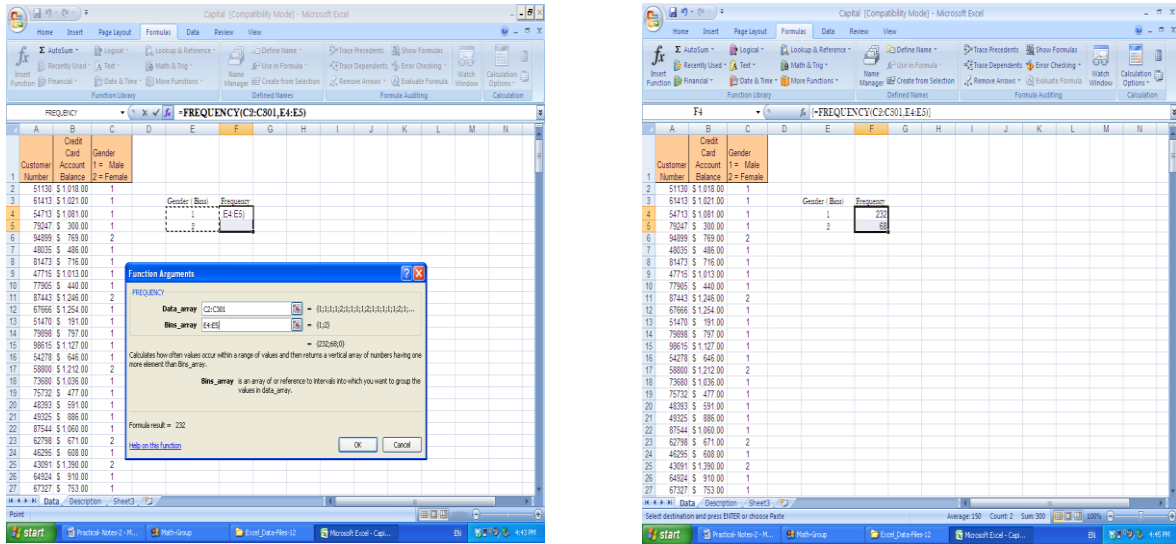


Example: Construct a frequency distribution for the variable gender, for the data in the file *Capital.xls*.

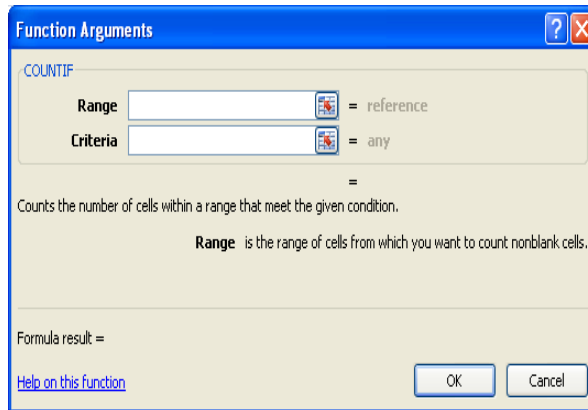
Open file: *Capital.xls*

- a. *Enter the possible values for gender 1 and 2 in cell E4:E5*
- b. *Select the cells to contain the frequency F4:F5*
- c. *Select the data array cells C2:C301*
- d. *Select the bins array cells E4:E5*
- e. *Press (Control + Shift) +Enter.*

The results will be displayed in cells F4:F5

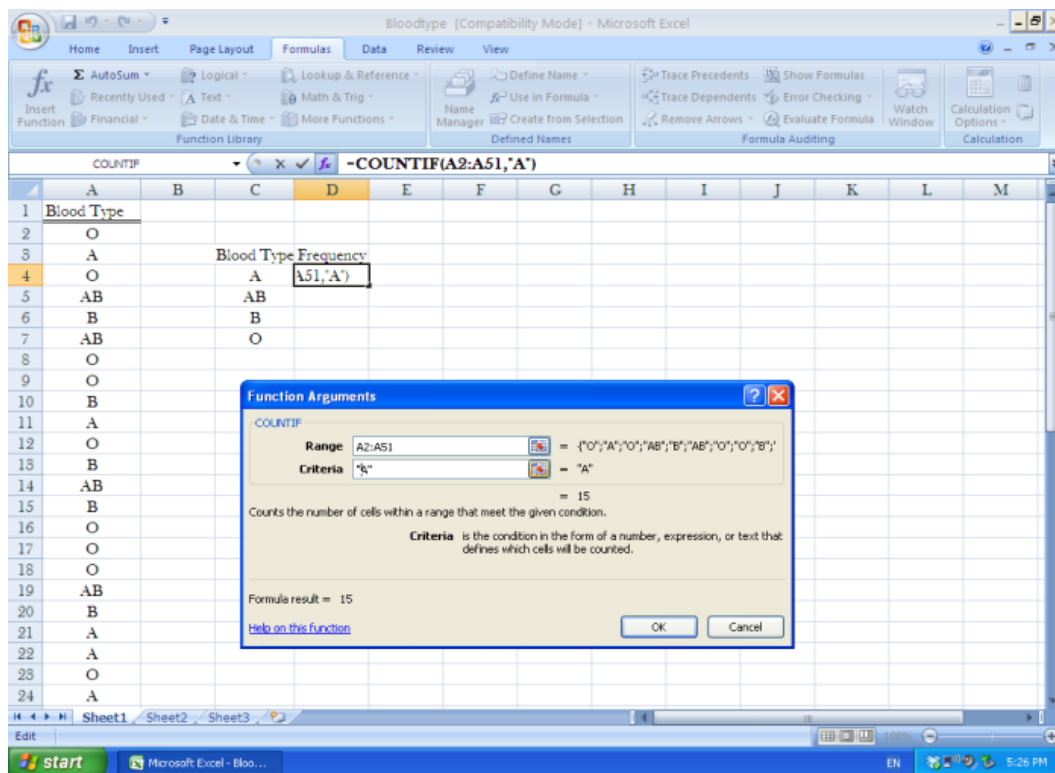


- B. **COUNTIF**: To construct a Frequency Distribution for numeric or non-numeric values such as the blood type, major, drink preference, marital status, color preference.
- List the possible values of the variable (the bins).
 - Select the cell to contain the frequency.
 - Click **Formulas Tab** → **Functions** → **Statistical** → **COUNTIF**
 - Select the input range
 - Select the criteria (Number, text, a condition)

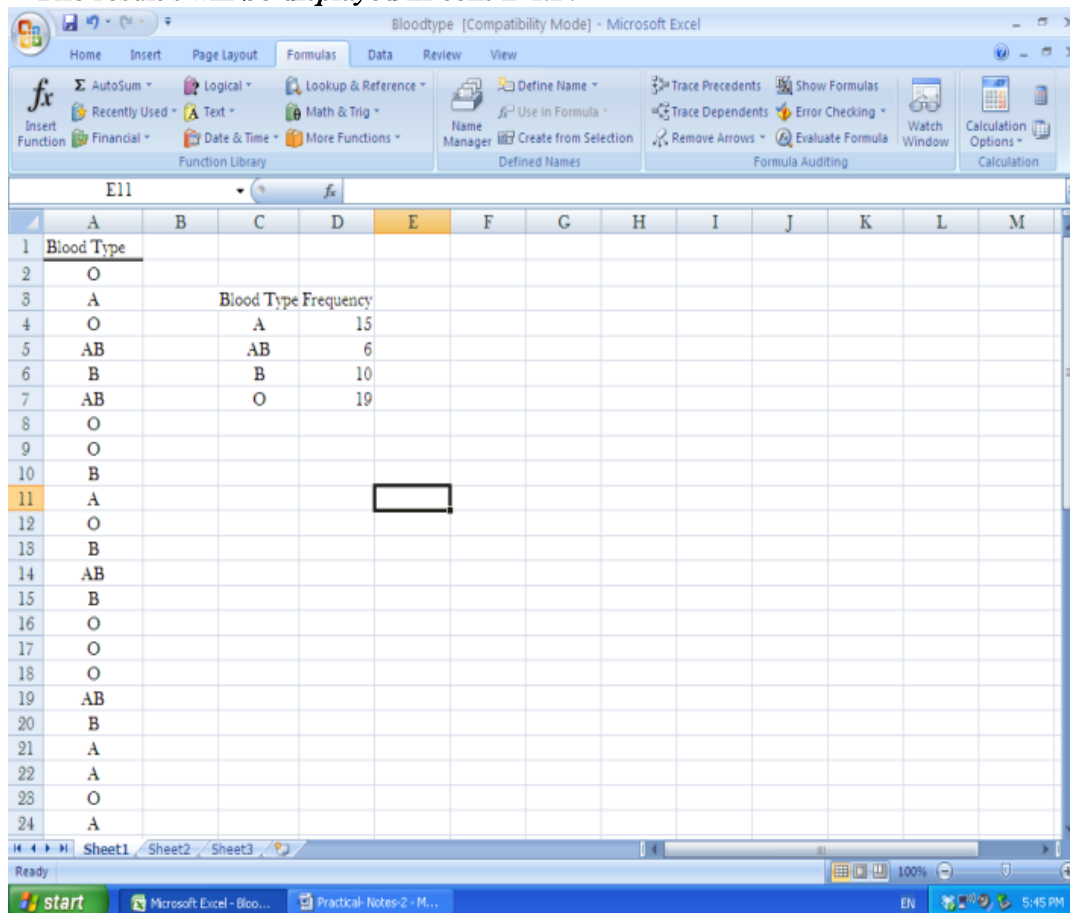


Example: Construct a frequency table for the blood type, for the data in the file **Blood.type.xls**
 Open file: **Blood.type.xls**

- Enter the possible values for blood type A, AB, B and O in cell C4:E7
- Select the cells to contain the frequency D4:D7
- Select the data array cells Range A2:A51
- Criteria "A" and repeat the same steps for the blood type AB, B, and O
- Or type the cell reference A4 which contains the letter "A", then use the filler for the other blood types. (you need to change the Range A2:A51 that you selected to A\$2:A\$51)



The results will be displayed in cells D4:F7



2. Grouped Frequency Distribution (Interval or Ratio data)

Determine the number of classes using the $2^k \geq n$ guideline rule.

Find the Maximum and Minimum Values

Find the class width $w = \frac{Max - Min}{K} = \frac{84 - 18}{7} = 9.4 \approx 10$; it should contain the number of decimal values as the data.

Use the lower limit a nice number such as 5, 10 or 20, must be less than or equal the minimum value.

Use upper class limits (Excel bins) for discrete classes.

Example: Given a set of data with $n=100$, $min=18$, $max=84$

$$k = \frac{\log n}{\log 2} = \frac{\log 100}{\log 2} = 6.64 \approx 7$$

$$w = \frac{Max - Min}{K} = \frac{84 - 18}{7} = 9.4 \approx 10$$

Classes can be constructed by choosing the lower limit =minimum=18 as in Table-1 or the lower limit =15 as in Table-2 or lower limit=10 as in Table-3

Table -1

Classes
18-27
28-37
38-47
48-57
58-67
68-77
78-87

Table - 2

Classes
15-24
25-34
35-44
45-54
55-64
65-74
75-84

Table - 3

Classes
10-19
20-29
30-39
40-49
50-59
60-69
70-79
80-89

Table - 4 (continuous classes)

Classes
10 --< 20
20 --< 30
30 --< 40
40 --< 50
50 --< 60
60 --< 70
70 --< 80
80 --< 90

You can see that Table-3 is easier than table-1 even if it has one more class, because multiples of 10=class width is used for lower limits.

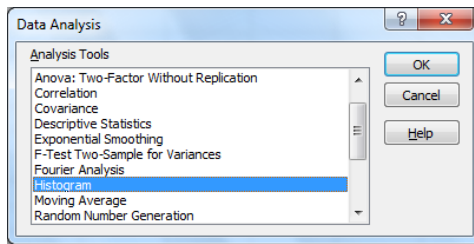
Grouped frequency distributions can be constructed in Excel by:

- A. Data → Data Analysis → Histogram (with or without chart)
- B. Formulas → Functions → FREQUENCY
- C. Formulas → Functions → COUNTIF

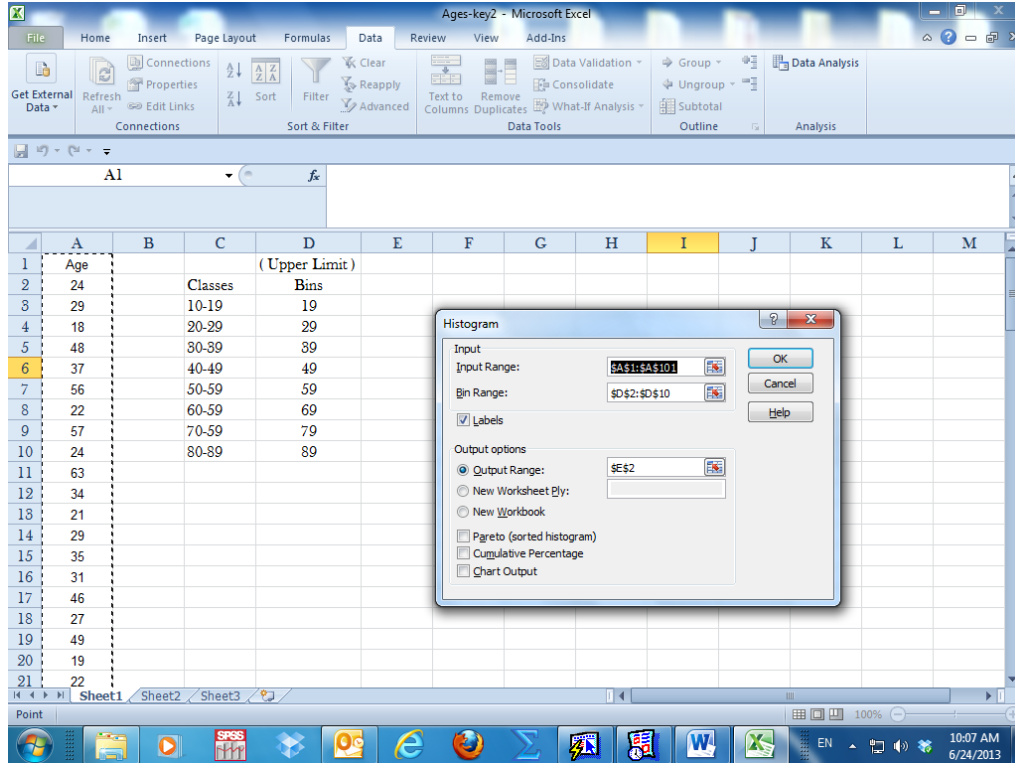
Example: Use the data in the file Age.xls to construct a grouped frequency distribution using the classes in Table-3 above.

A. Histogram

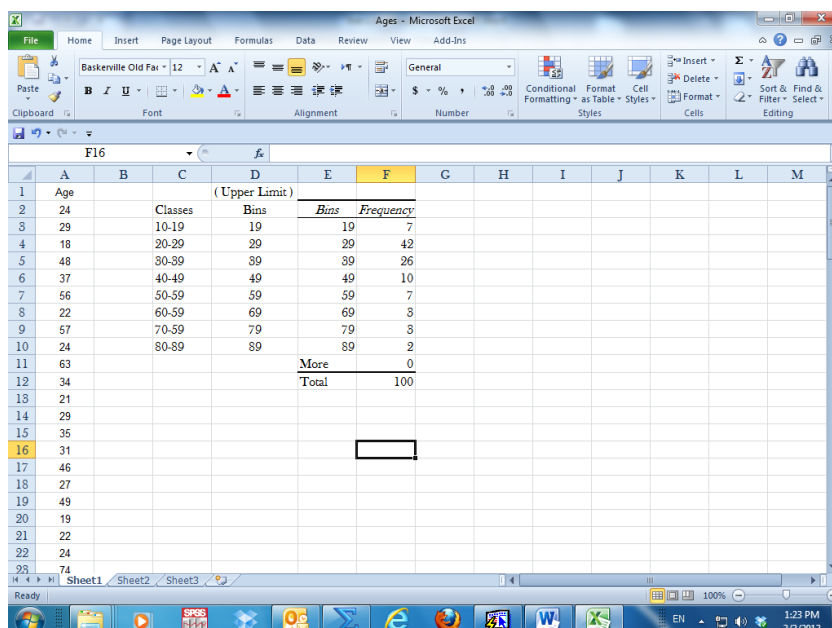
- a. **Input the classes in column C3:C10**
- b. **Input the “Bins =Upper class limits” in column D3:D10**
- c. **Click Data Tab → Data Analysis → Histogram → OK**



d. Input data cell A1:A101 → Bins Range Cell D2:D10 → Click Label box → for Output Range click cell E2 → Ok

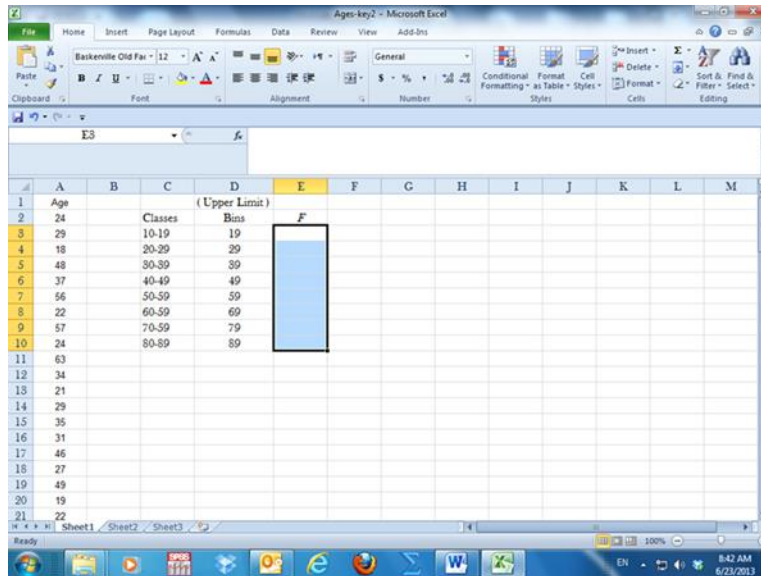


e. The output will be in cell E2:E11, Excel will give a “more” count of zero if the classes are inclusive.

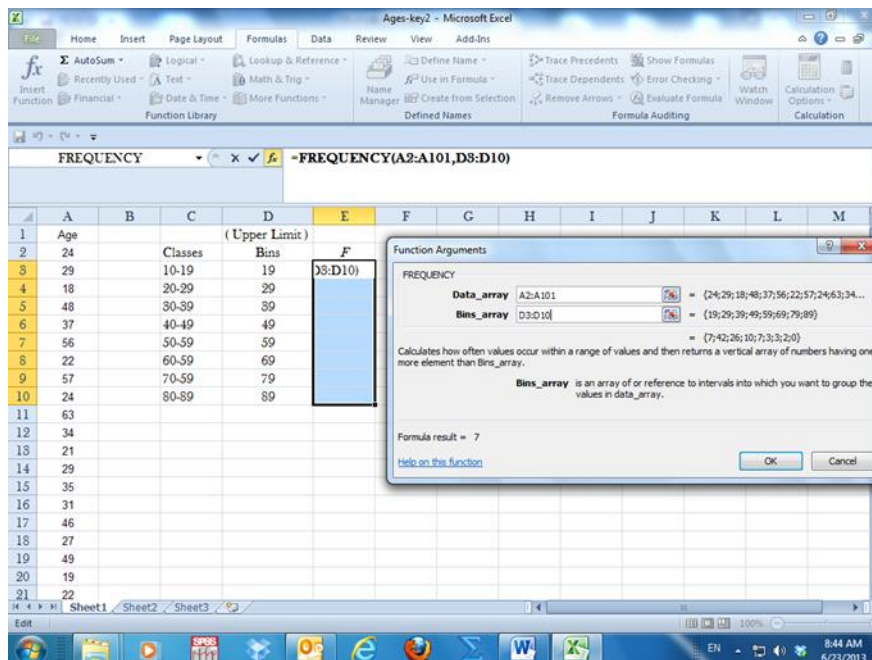


B. FREQUENCY

- Input the classes in column C3:C10
- Input the Bins =Upper class limit in column D3:D10
- Select the cells to contain the frequency values E3:E10

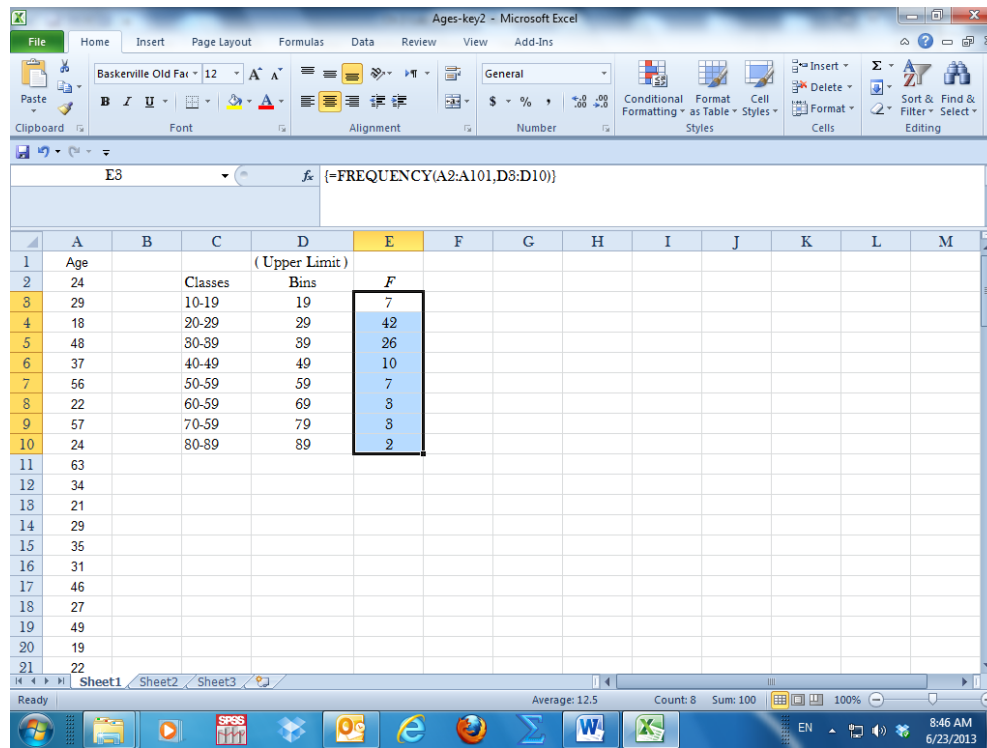


- Click Formulas Tab → Functions → Statistical → FREQUENCY
- Select the Data array A2:A101
- Select the Bins array D3:D10



- Press (Control +Shift) + Enter
Note: You have to hold the (Control +Shift) then press ENTER

h. The frequency will be in cell E3:E10



Note 1: The advantage of the frequency function is that when the raw data changes the result will be updated automatically, while the histogram command if the raw data changes the result will not change, you have to repeat the steps every time the raw data changes.

The advantage of the histogram command is that it gives an extra class “more” which will include the count of any values which were not included by the last class; (if classes are not inclusive).

Also histogram command you can have the graph of the data as a “Histogram” by choosing the chart output which will be explained later.

Note 2: The bins for continuous classes are upper limit minus a small number based on the number of decimal places of the classes.

For example Bins=Upper limit-0.1 as in Table -1

Bins=Upper limit - 0.01 as in Table-2

Bins=Upper limit - 0.001 as in Table-3

Table -1

Classes	Bins
10-- < 20	19.9
20-- < 30	29.9
30-- < 40	39.9

Table -2

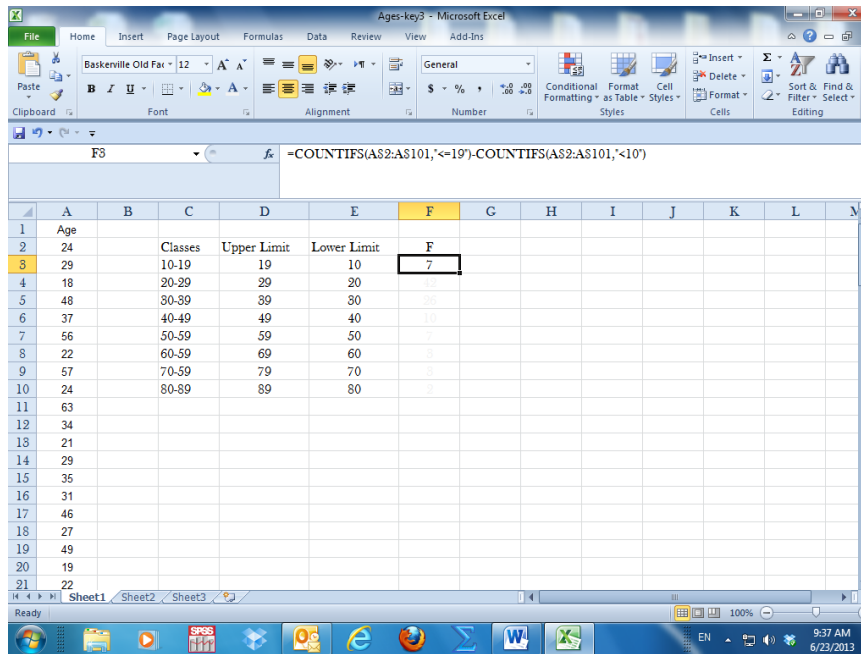
Classes	Bins
1.5-- < 2.0	1.99
2.0-- < 2.5	2.49
2.5-- < 3.0	2.99

Table -3

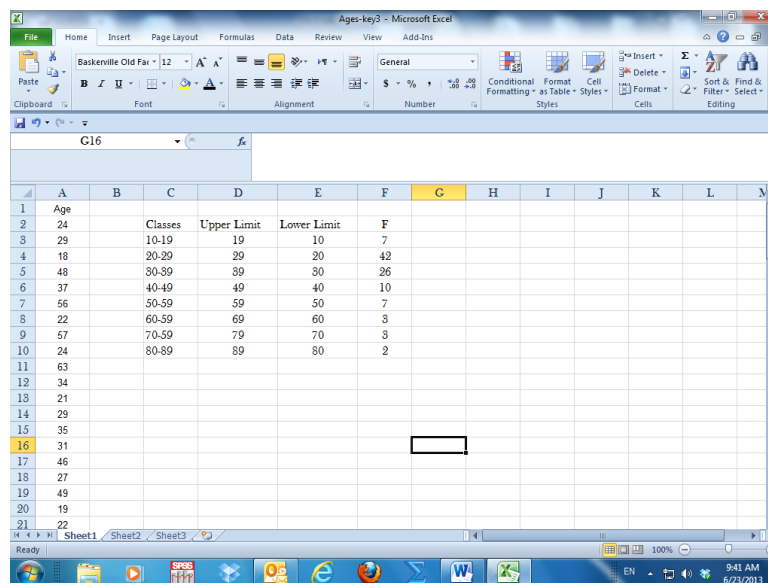
Classes	Bins
1.50 -- < 1.54	1.539
1.54 -- < 1.58	1.579
1.58 -- < 1.62	1.619

C. COUNTIF (Optional)

- Input the classes in column C3:C10
- Input the Upper class limits 19, 29,.....89 in column D3:D10
- Input the Lower class limits 10, 20,...80 in column E3:E10



- Type in cell F3 the formula $=COUNTIF(A2:A101, "<=19") - COUNTIF(A2:A101, "<10")$
The result frequency count will be in cell F3
Continue typing the formulas in cells F4... F10
The formula in F10 will be $=COUNTIF(A2:A101, "<=89") - COUNTIF(A2:A101, "<90")$
The result as shown below



Note: For continuous classes replace " \leq " by " $<$ " in the COUNTIF formula
 $=COUNTIF(A2:A101, "<20") - COUNTIF(A2:A101, "<10")$

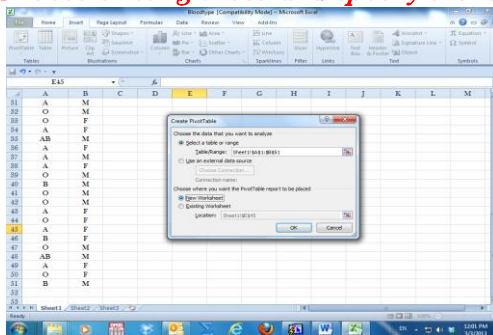
3. Joint Frequency Distribution for Two Variables (Qualitative or Quantitative)

A. **Pivot Table:** To construct a Joint Frequency Distribution for two variables.

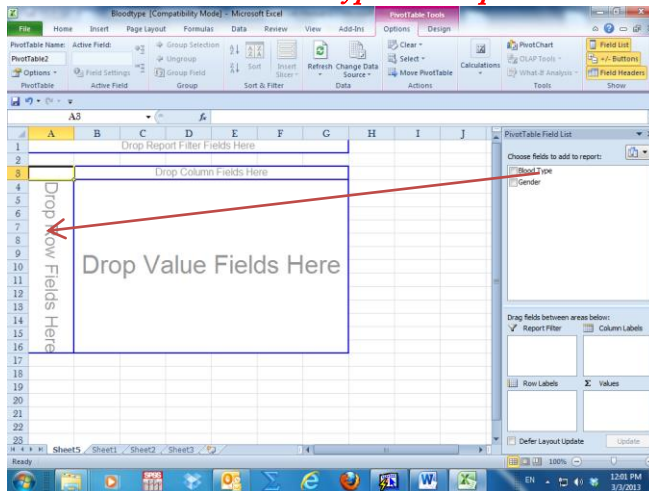
- a. Click **Insert** → **Pivot** → **Pivot Table** → **Select the variables or all the table**
- b. **Choose existing sheet and specify the location or choose a new work sheet**
- c. **Move the variables for the column and rows**
- d. **Move value to the value field**
- e. **Right click and summarize data by count**
- f. **Format the table as needed**

Example 1: To construct a joint frequency table for two variables in the file Blood Type:

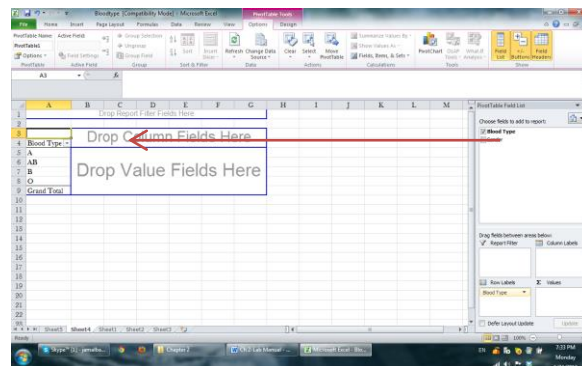
- a. Click **Insert** → **Pivot Table** → **Select the variables** Blood type and Gender
- b. **Choose existing sheet and specify the location or choose a new work sheet** → **click OK**



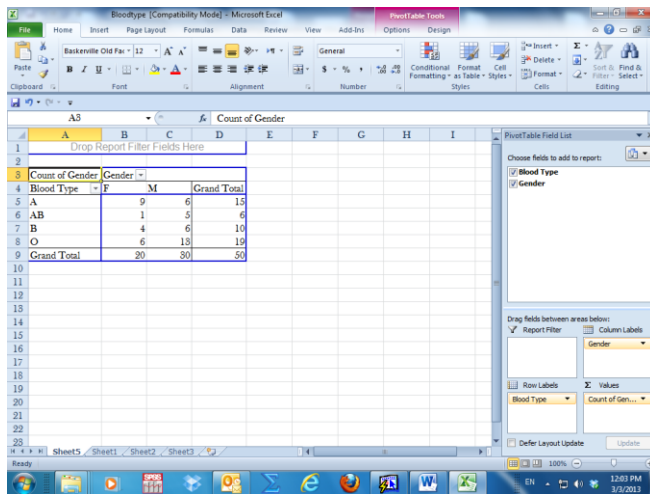
c. **Move the variable Blood type to Drop Row Field Here**



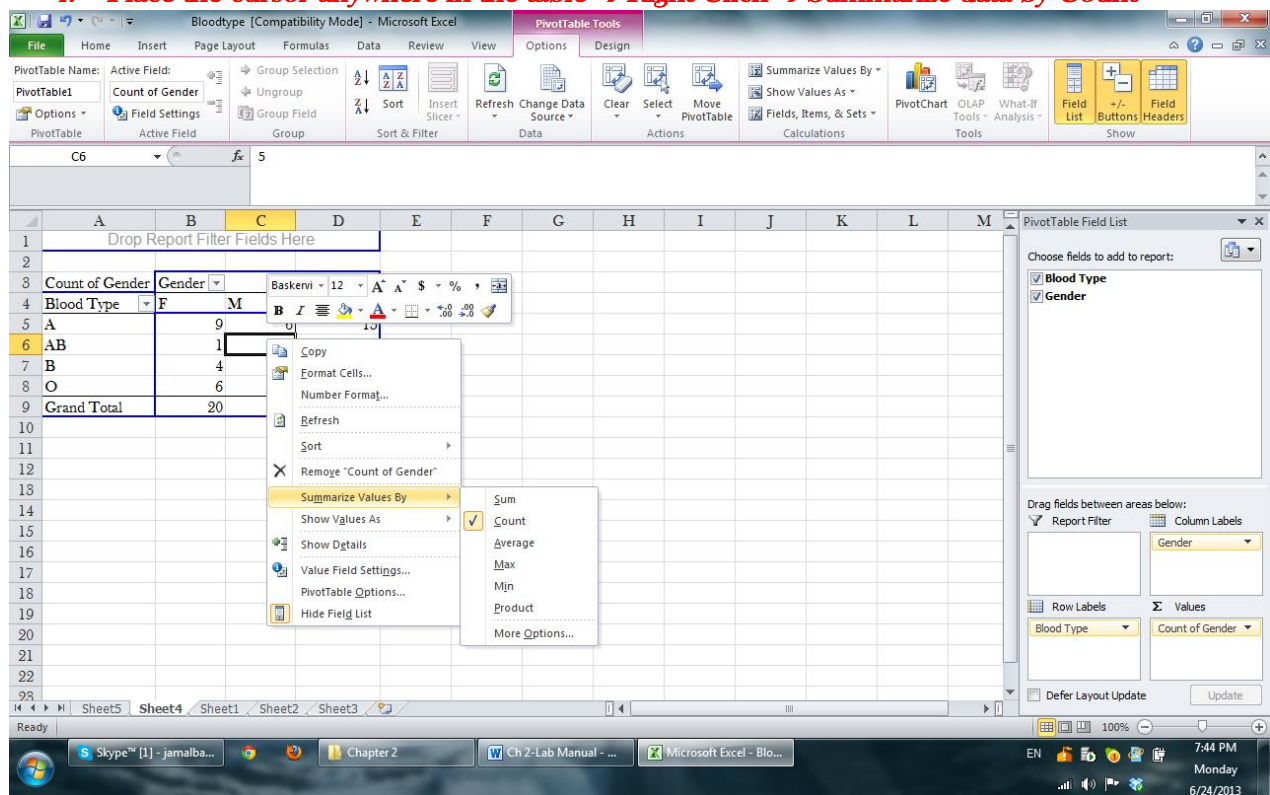
d. **Move the variable Gender to Drop Column Fields Here**



e. Move the variable Gender to Drop Value Fields Here



f. Place the cursor anywhere in the table → Right Click → Summarize data by Count



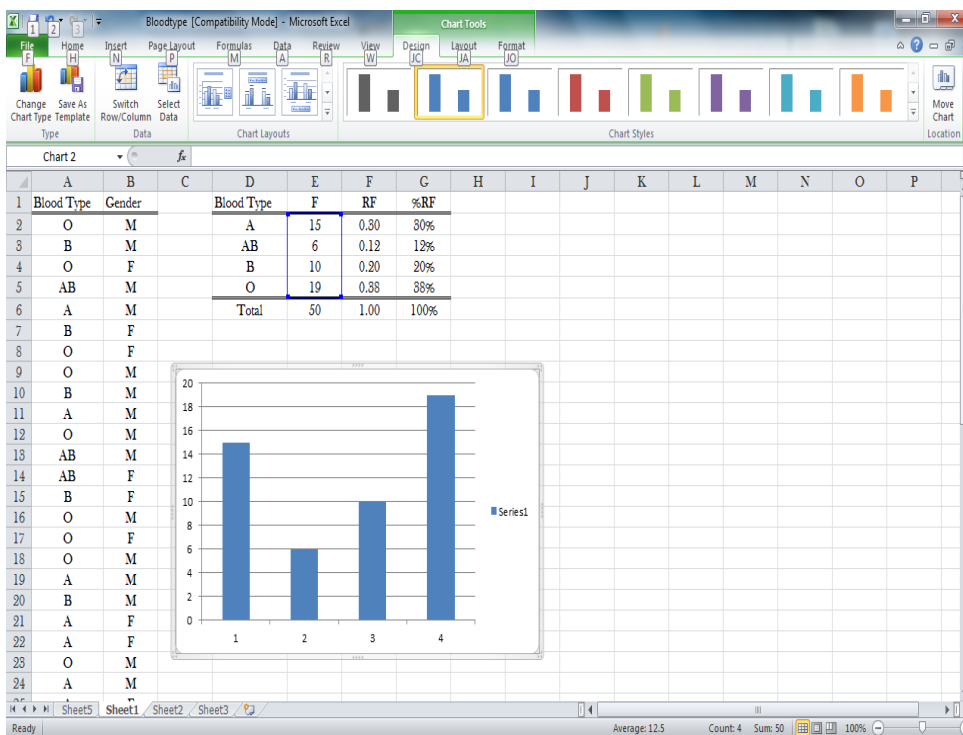
Joint Frequency Table of Blood Type by Gender

Count of Gender	Gender		Grand Total
	F	M	
Blood Type			
A	9	6	15
AB	1	5	6
B	4	6	10
O	6	13	19
Grand Total	20	30	50

II. Describing Data by Graphs

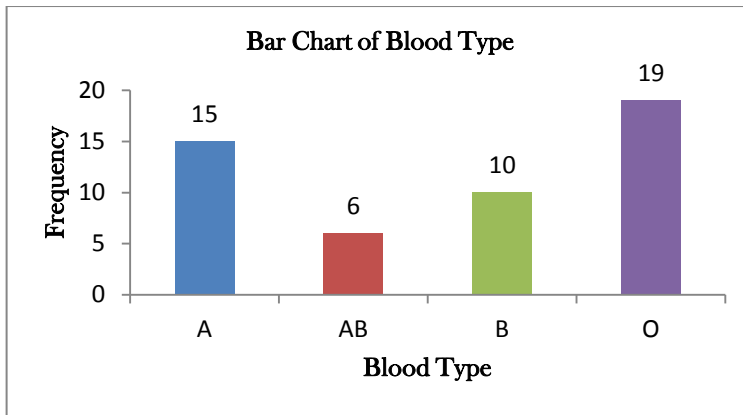
1. Bar Chart (Vertical or Horizontal)
2. Pie Chart
3. Cluster or Stack Bar Chart
4. Histogram
5. Ogive
6. Stem and Leaf
7. Line Chart
8. Scatter Diagram

1. Bar Chart (Vertical or Horizontal): The following steps describe how to use Excel's Chart Wizard to construct a bar graph for the blood type data using the frequency distribution appearing in cells E2:E5
 - a. Select cells E2:E5
 - b. Click the Chart Wizard button on the Standard toolbar (or select the Insert menu and choose the Chart option)
 - c. Choose Column in the Chart type list

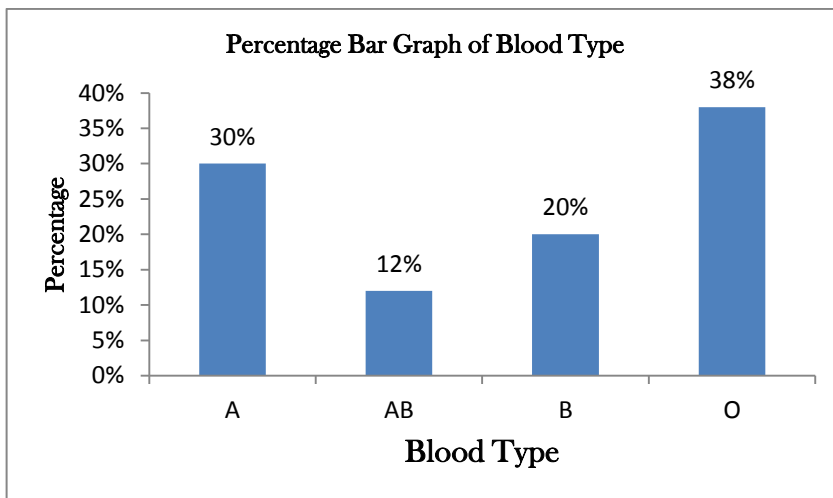
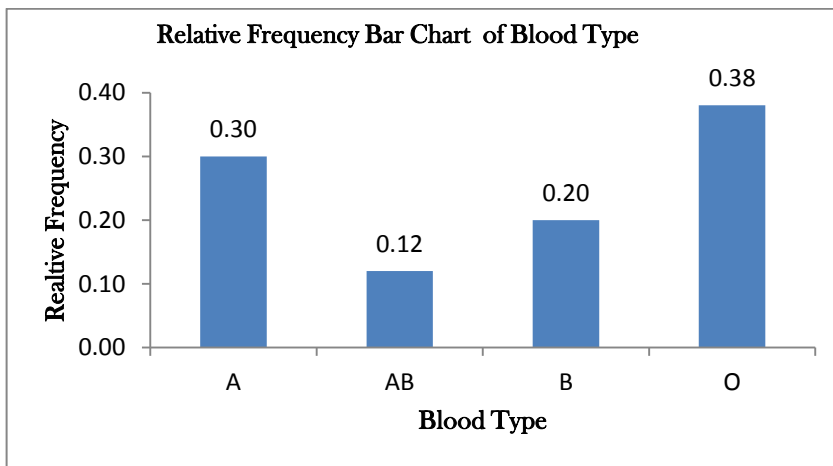


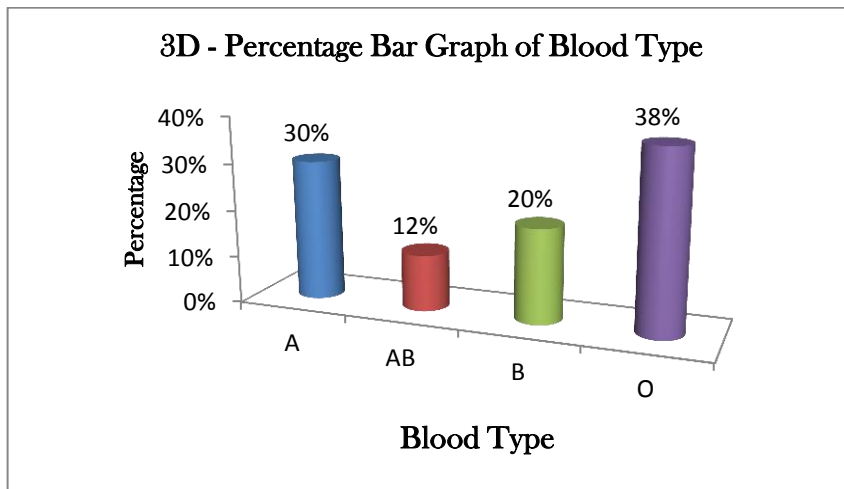
- d. Select the Layout Tab → Gridlines → Horizontal > none
- e. Select the Titles tab → Chart Title → Choose Above Chart (where you want the title)
Type "Bar Chart of Blood Type" in the Chart Title Box
- f. Select the Titles Tab → Axis Title → Horizontal → Choose below axis
Type "Blood Type" in the Category (X) axis box
- g. Select the Titles Tab → Axis Title → Vertical → Choose rotated or (anything you prefer)
Type "Frequency" in the Value (Y) axis box
- h. Select the Legend Tab and then Remove the check in the Show legend box
- i. Click the Design Tab → Move chart to the location you want or use the default sheet.

The resulting bar graph (chart)



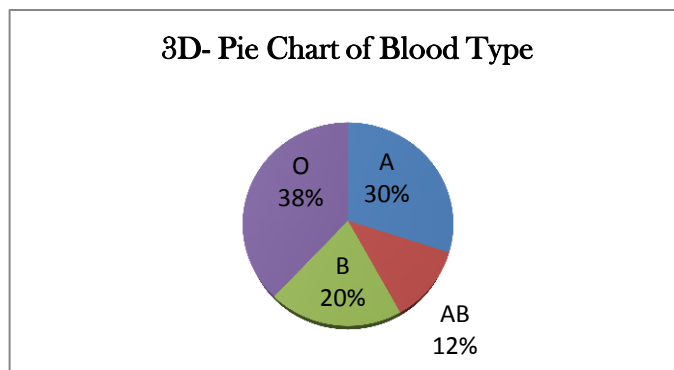
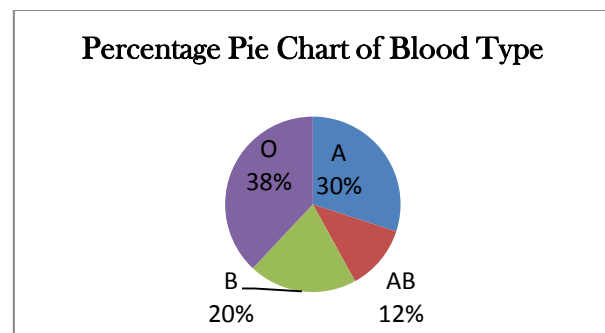
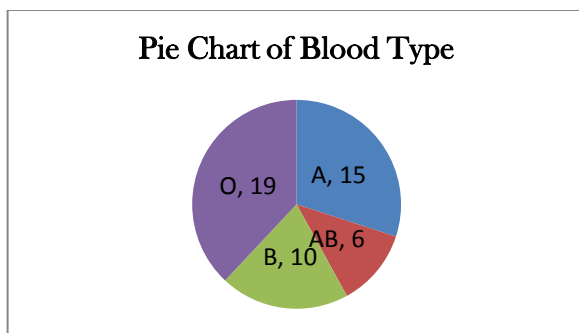
Note: Relative frequency and percentage bar charts can be constructed using the same steps. The only changes are in the cells selected. Select F2:F5 for relative frequency and G2:G5 for percentage. The chart title and y-axis title will change accordingly.





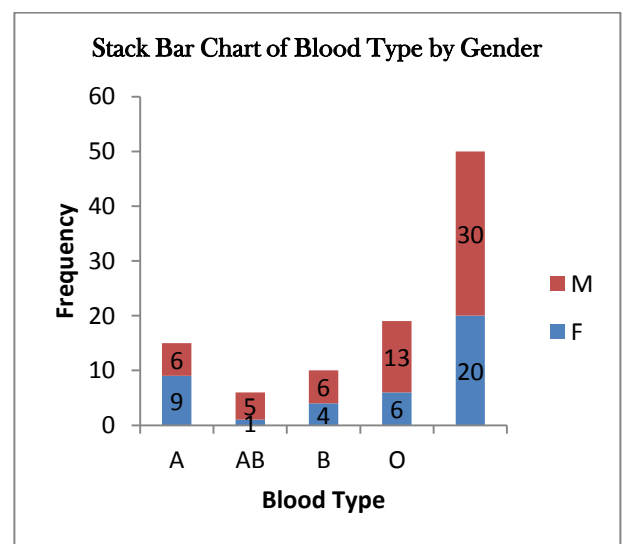
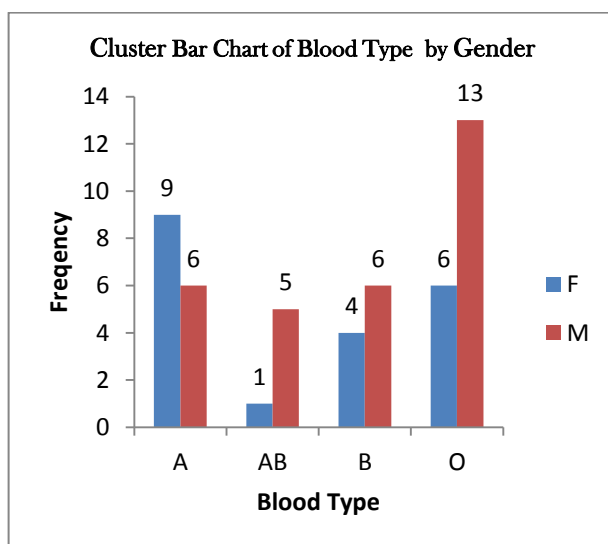
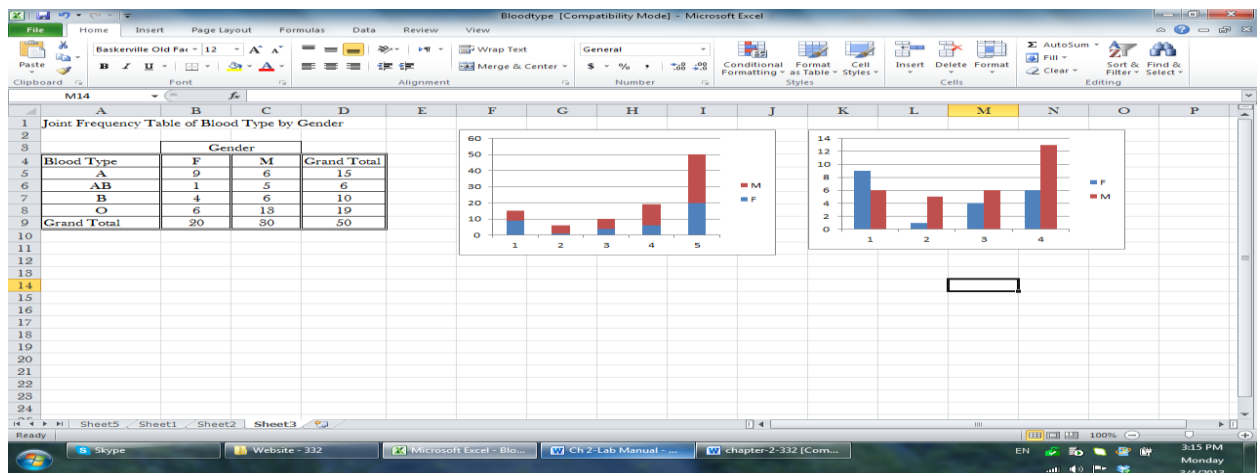
2. **Pie Chart:** The following steps describe how to use Excel's Chart Wizard to construct a Pie chart for the blood type data in the file *Boodtype.xls* using the frequency distribution appearing in cells E2:E5
 - a. Select cells E2:E5
 - b. Click the Chart Wizard button on the Standard toolbar (or select the Insert menu and choose the Chart option)
 - c. Choose pie in the Chart type list
 - d. Select the Titles tab → Chart Title → Choose Above (where you want the title)
Type "Pie Chart of Blood Type" in the Chart title box
 - e. Select the Legend tab and then Remove the check in the Show legend box
 - f. Click the design tab → Move chart to the location you want or use the default sheet.

The resulting Pie graphs (charts)



3. **Cluster Bar and Stack Bar:** A bar graph is used to represent discrete values for more than one item that share the same category. It can be side by side (cluster) or the bar divided into subparts that represent the discrete value for items that represent a portion of a whole group. They can be constructed the same way as a bar chart.

- Select cells that contain the variables.
- Click the Chart Wizard button on the Standard toolbar (or select the Insert menu and choose the Chart option)
- Choose Column (Cluster) or Stack in the Chart type list
- Select the Layout Tab → Gridlines → Horizontal → None
- Select the Titles tab → Chart Title → Choose Above Chart (where you want the title)
Type "Cluster Bar Chart of Blood Type by Gender" in the Chart Title Box
- Select the Titles Tab → Axis Title → Horizontal → Choose below
Type "Blood Type" in the Category (X) axis box
- Select the Titles Tab → Axis Title → Vertical → Choose rotated or (anything you prefer)
Type "Frequency" in the Value (Y) axis box
- Select the Legend Tab and then Remove the check in the Show legend box
- Click the Design Tab → Move chart to the location you want or use the default sheet.



4. **Histogram:** The following steps describe how to use Excel to construct a Histogram for quantitative data.

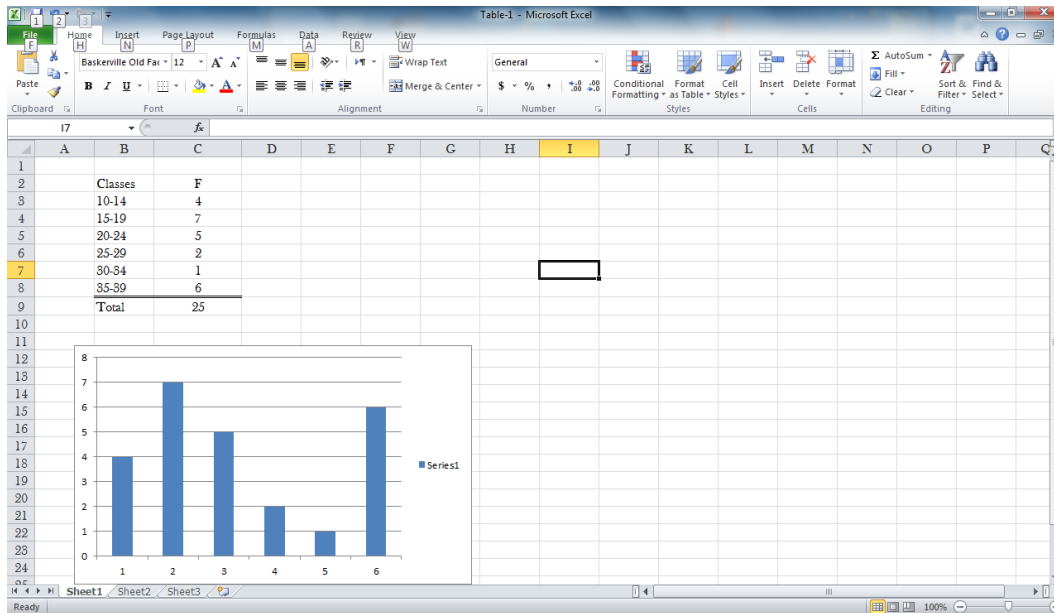
A histogram is a bar graph with no gaps between the bars. The x-axis represents the classes.

A. Histogram from Grouped Frequency Distribution.

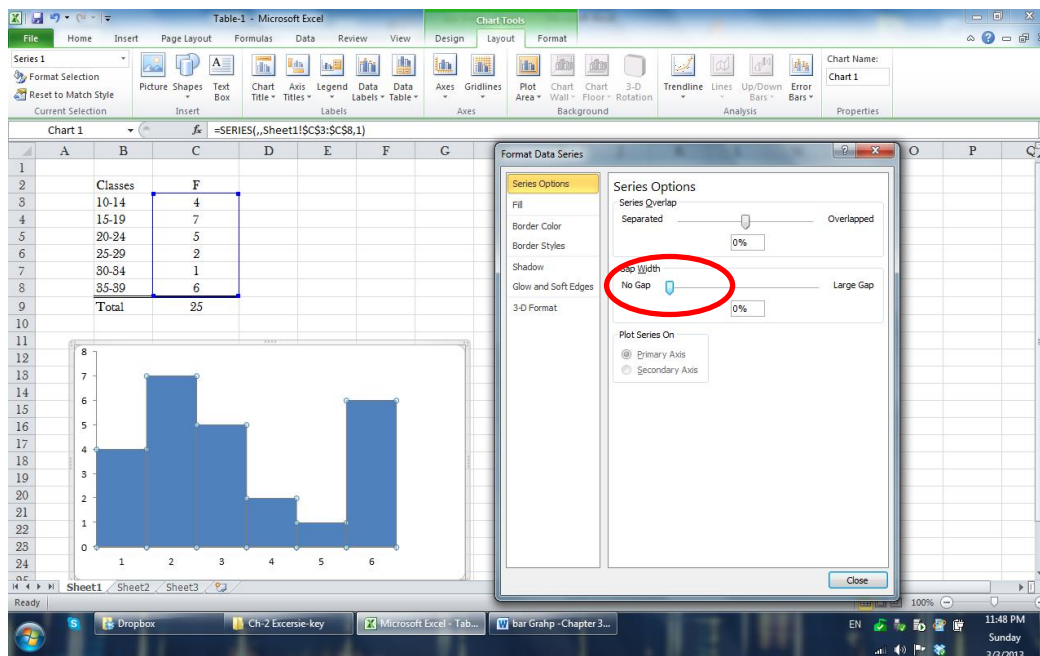
B. Histogram from Raw Data.

A. **Histogram from Grouped Frequency Distribution as in the file Table-1**

a. **Select the cells that represent the frequency and construct a bar chart.**



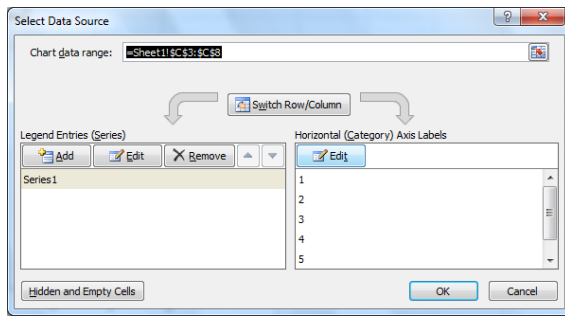
b. **Click on any one of the bars → right click → Format data series → Move the gap to No Gap**



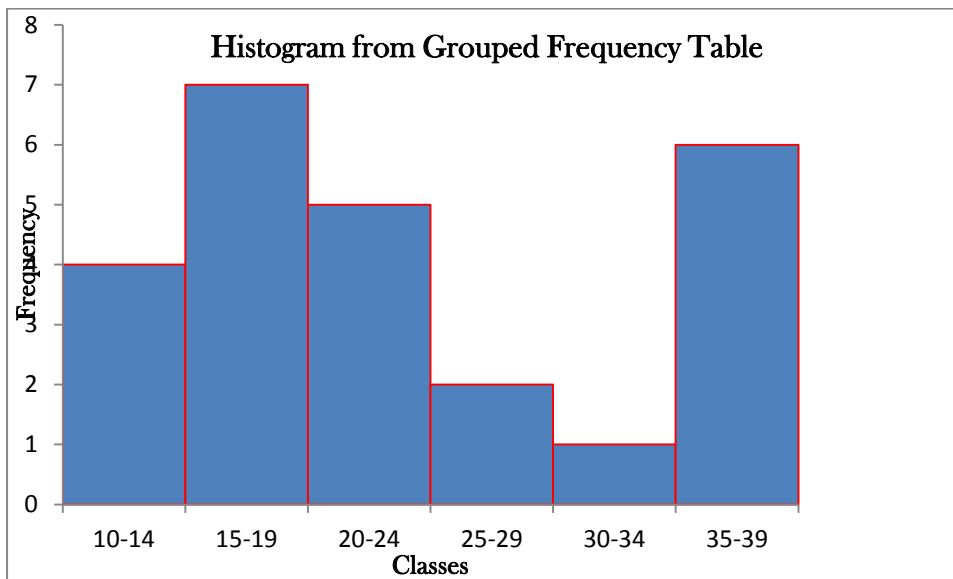
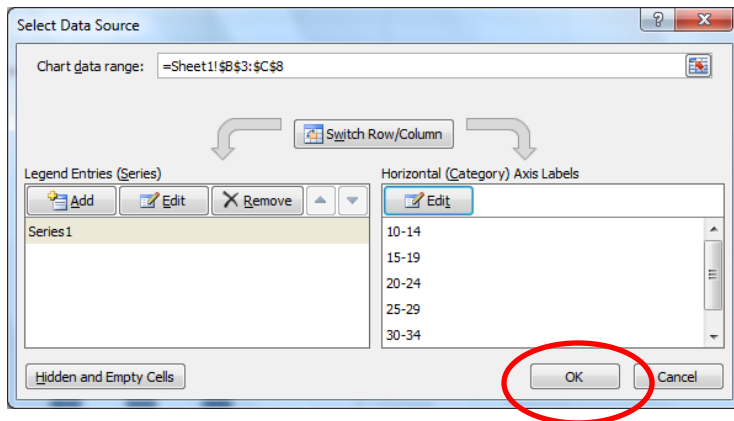
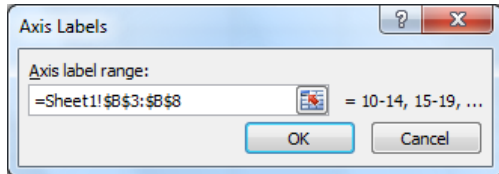
c. **You can choose border color to show the border between the bars in a different color.**

d. **The x-axis should have the classes**

Select the labels on x-axis then right click → Select Data → Edit → OK

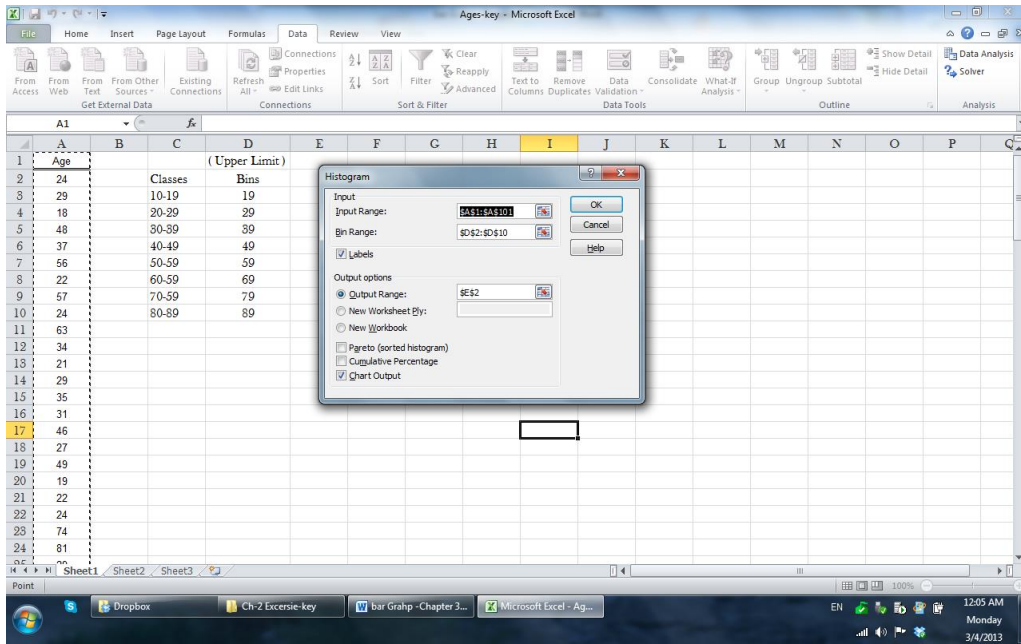


Select the classes → Ok

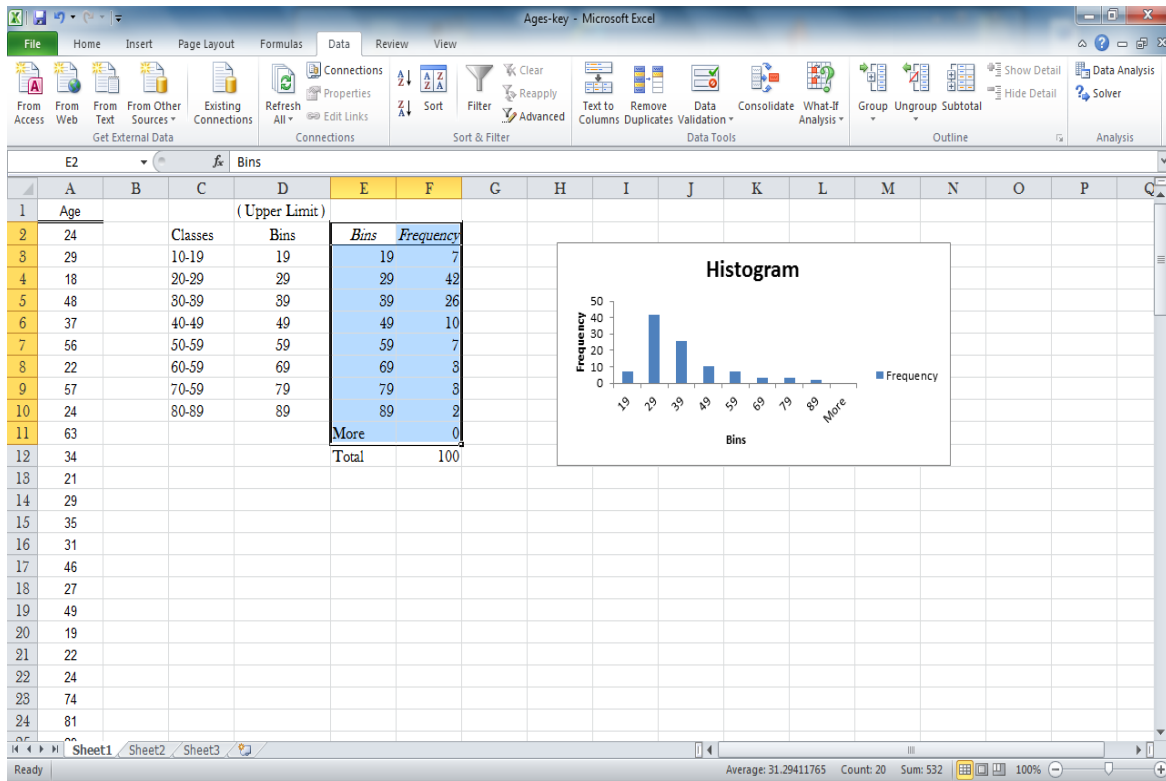


B. Histogram from Raw Data (File Age.xlsx)

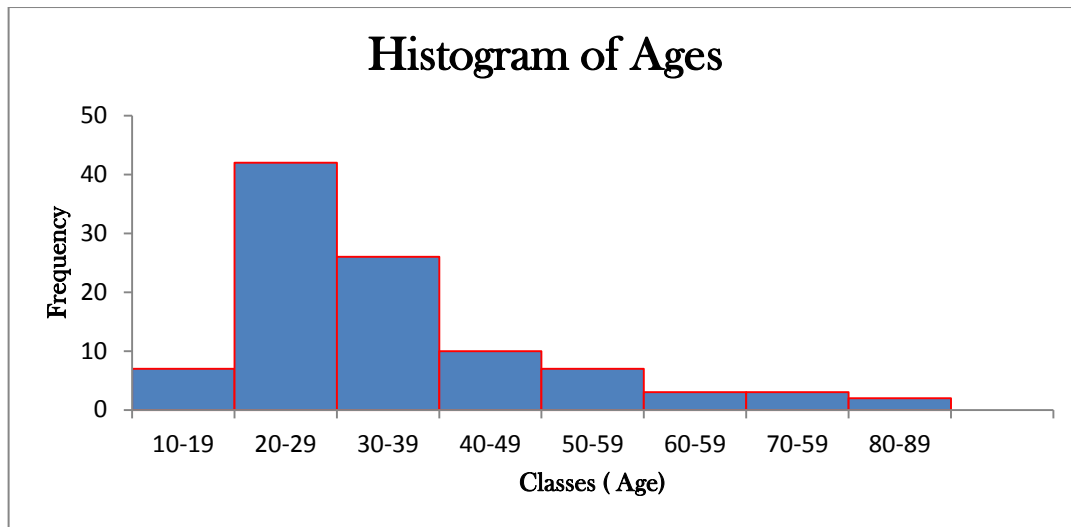
- a. Type in the classes in cell C3:C8, the bins as the upper limits of the classes in cell D3:D8
- b. Click Data Tab → Data Analysis → Histogram
- c. Input Range → Bins Range → Click label → Output range
- d. Click Chart Output → Ok



e. The result will be a frequency table and a bar chart

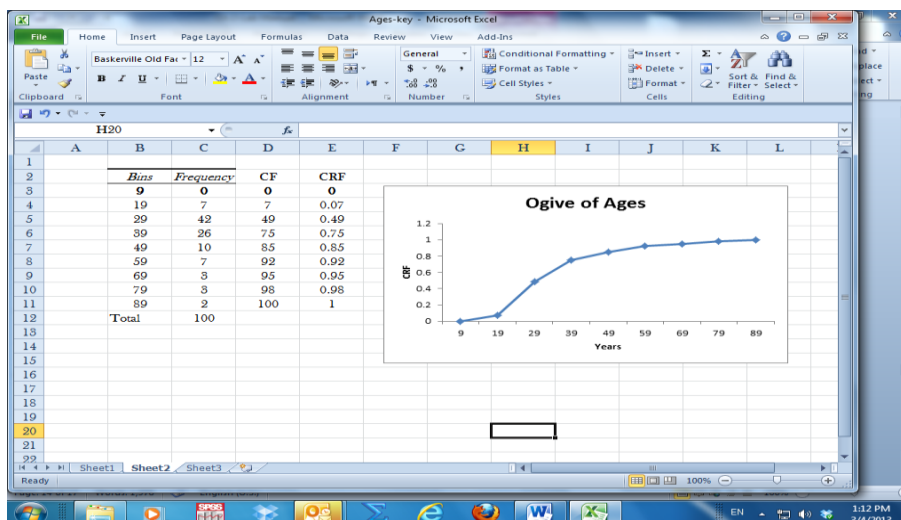


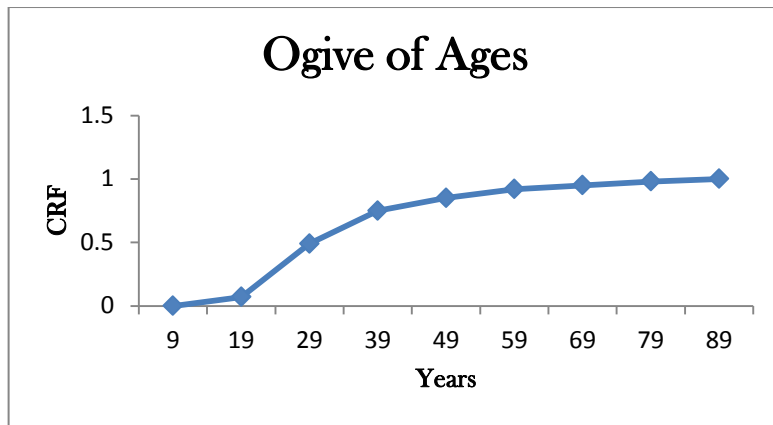
f. Remove the gaps and format the chart as discussed before.



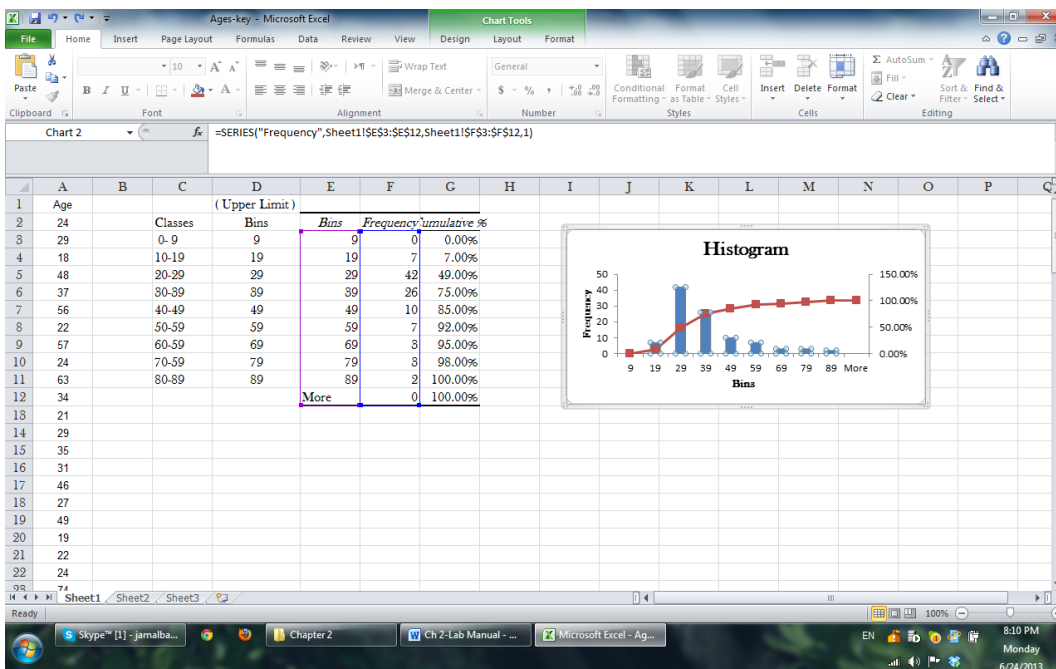
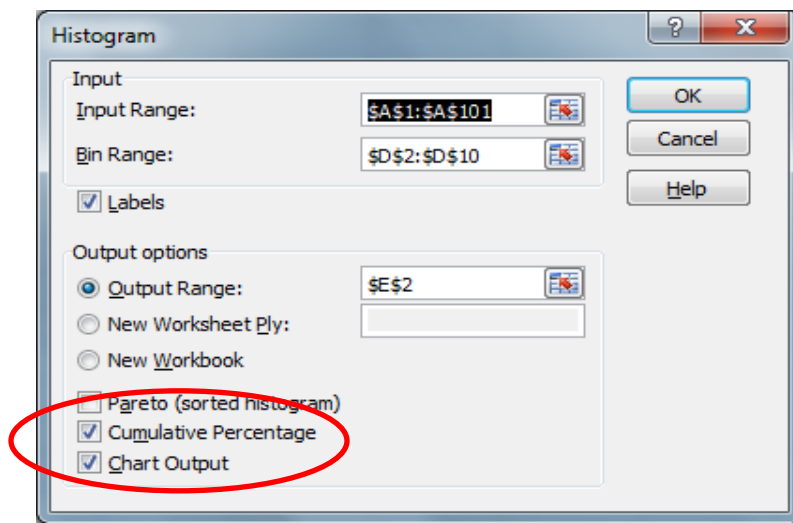
5. **Ogive:** A line chart of Cumulative Relative Frequency versus the upper limit of classes.
 Example: Develop an ogive for the data in the file “Age”.

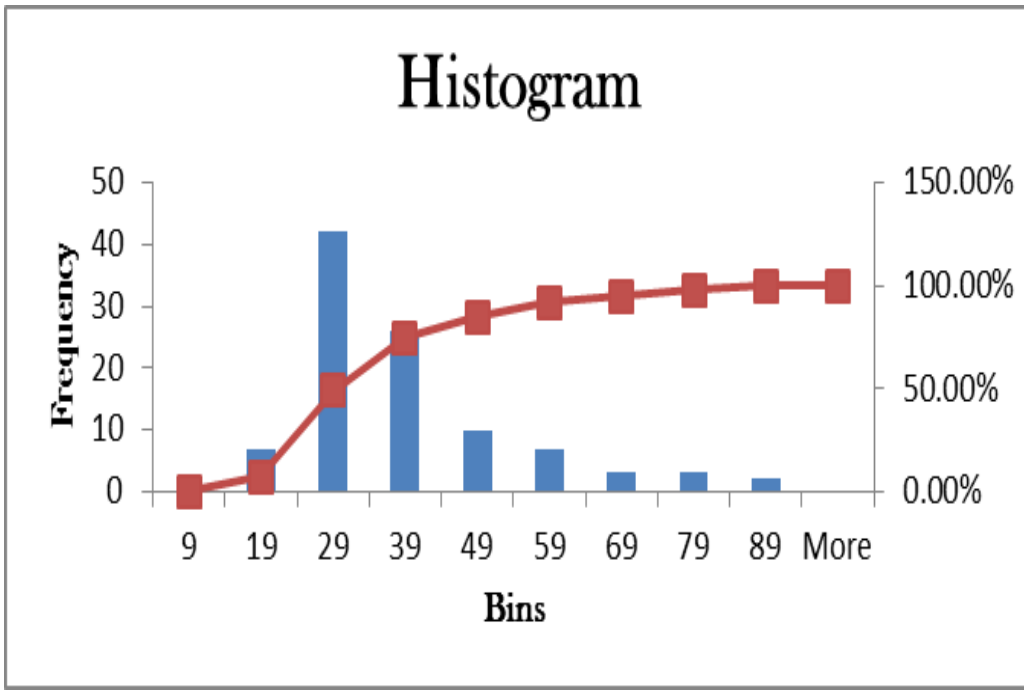
- a. Select cells E3:E11
- b. Click the Chart Wizard button on the Standard toolbar (or select the Insert menu and choose the Chart option)
- c. Choose line with marks in the Chart type list
- d. Select the Layout Tab → Gridlines → Horizontal → None
- e. Select the Titles Tab → Chart Title → Choose Above (where you want the title)
Type “Ogive of Ages” in the Chart Title Box
- f. Select the Titles Tab → Axis Title → Horizontal → Choose below Chart
Type “Years” in the Category (X) axis box
- g. Select the Titles Tab → Axis Title → Vertical → Choose rotated or (anything you prefer)
Type “CRF” in the Value (Y) axis box
- h. Select the Legend Tab and then Remove the check in the Show legend box
Click the Design Tab → Move chart to the location you want or use the default sheet.



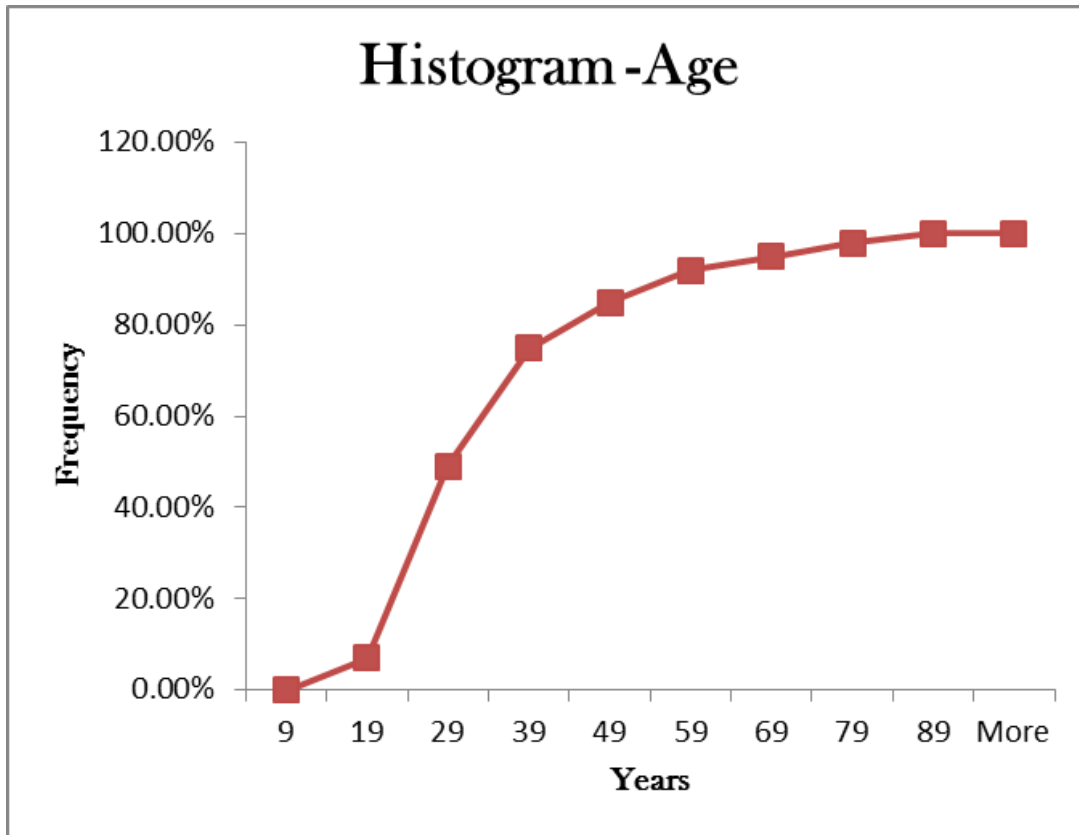


Note: Ogive can be developed along with the histogram from raw data by checking the box as shown below. You need add a class at the beginning with frequency =0 because the ogive must be connected to the x-axis.



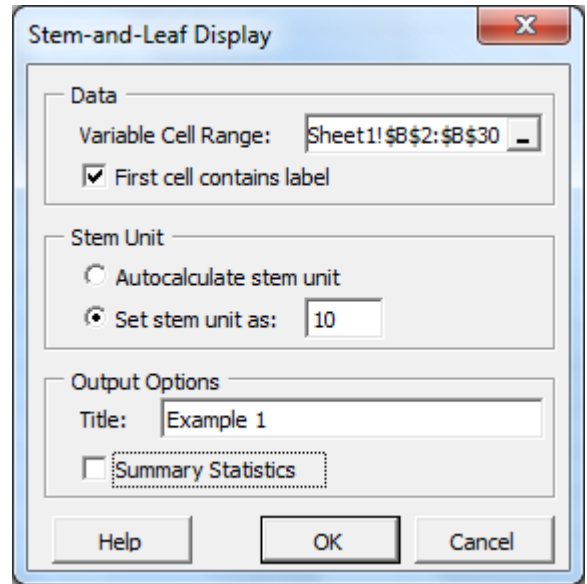
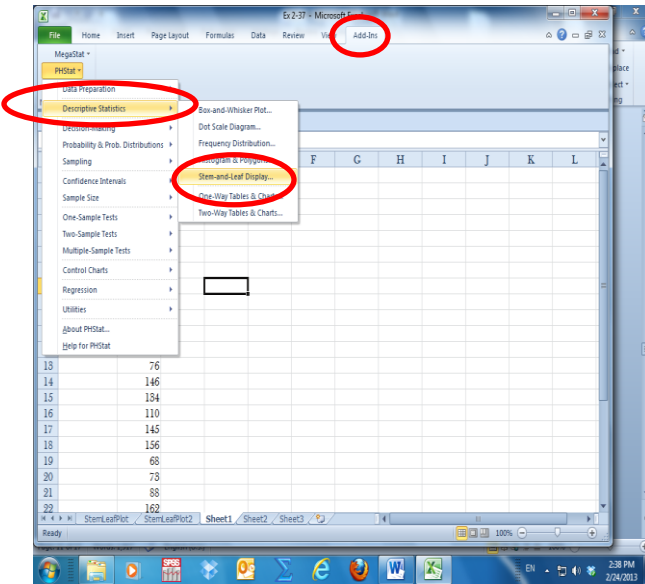


You can delete the bars by clicking on a one, right click, delete.
 Add chart title and axis title for the ogive. The vertical axis is Cumulative Relative Percentage.



6. **Stem and Leaf (The data of exercise 2-37)**

- a. **Open PHStat2 → Enable Macros → Add-In >PHStat2**
- b. **Descriptive Statistics → Stem & Leaf**
- c. **Input Range –Select the data with the label**
- d. **Click label in the first column**
- e. **The stem unit =10**
- f. **Then uncheck the summary statistic box, the OK**



Stem and leaf constructed using Excel Add-in PHStat2 for the data of exercise 2-37:

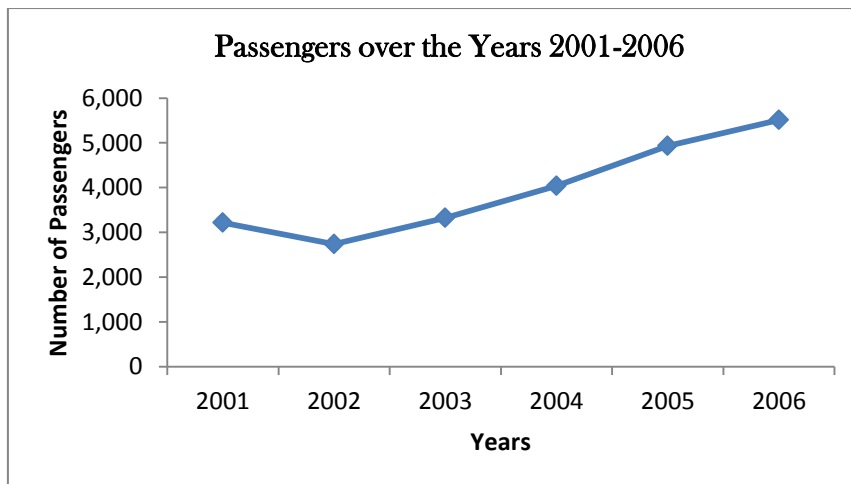
Stem unit: 10

6	8
7	1 3 4 6 9
8	3 5 8
9	0 2 3
10	3 5
11	0 6 9
12	
13	0 4 8
14	5 6 7
15	6 6
16	2
17	8
18	1

7. **Line Chart:** A two dimensional chart showing the time on the horizontal axis and the variable of interest on the vertical axis.

Example: Develop a line chart for the data in the file Line-Chart.xls

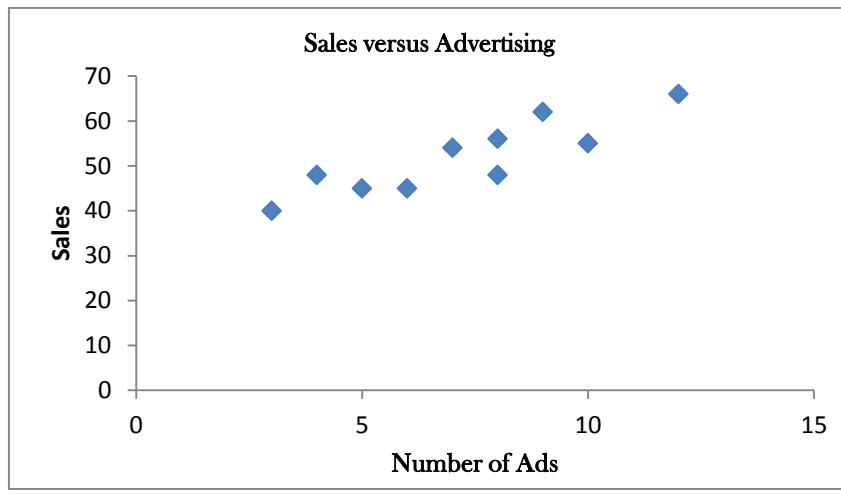
- a. Select cells B3:B8
- b. Click the Chart Wizard button on the Standard toolbar (or select the Insert menu and choose the Chart option)
- c. Choose line with marks in the Chart type list
- d. Select the Layout Tab → Gridlines → Horizontal → None
- e. Select the Titles Tab → Chart Title → Choose Above (where you want the title)
Type "Passengers over the Years 2001-2006" in the Chart Title Box
- f. Select the Titles Tab → Axis Title → Horizontal → Choose below Chart
Type "Years" in the Category (X) axis box
- g. Select the Titles Tab → Axis Title → Vertical → Choose rotated or (anything you prefer)
Type "Number of Passenger" in the Value (Y) axis box
- h. Select the Legend Tab and then Remove the check in the Show legend box
- i. Click the Design Tab → Move chart to the location you want or use the default sheet.



8. **Scatter Diagram:** A two dimensional graph of plotted points. The horizontal axis represents the values of one quantitative variable (Independent variable) and the vertical axis represents the values of the other quantitative variable (Dependent variable)

Example: Develop a scatter plot of the data in the file Scatter-plot.xls

- a. Select cells C4:C13, D4:D14
- b. Click the Chart Wizard button on the Standard toolbar (or select the Insert menu and choose the Chart option)
- c. Choose scatter
- d. Select the Layout Tab → Gridlines → Horizontal → None
- e. Select the Titles Tab → Chart Title → Choose Above (where you want the title)
Type "Sales versus Advertising" in the Chart Title Box
- f. Select the Titles Tab → Axis Title → Horizontal → Choose below chart
Type "Adds" in the Category (X) axis box
- g. Select the Titles Tab → Axis Title → Vertical → Choose rotated or (anything you prefer)
Type "Sales" in the Value (Y) axis box
- h. Select the Legend Tab and then Remove the check in the Show legend box
- i. Click the Design Tab → Move chart to the location you want or use the default sheet.



The chart shows a positive relation between the sales and advertising