# LAB 1 INSTRUCTIONS

## DISPLAYING AND DESCRIBING DISTRIBUTIONS

This lab will assist you in learning how to display and describe the quantitative data in Excel. In particular, you will learn how to obtain line charts (time plots) and histograms. Moreover, you will study how to summarize quantitative data with the **Descriptive Statistics** and **Insert Function** tools.

### 1. Charts in Excel: The Basics

In Excel, graphs are made from ranges of numbers in a workbook. Basically, we need to specify (a) the cells that contain the values to be graphed, (b) the type of graph we want drawn, (c) the location where the graph is to be placed. The information in (a) and (b) is entered into Excel through appropriate dialog boxes.

A chart in Excel consists of several components. Some of these components are displayed by default, others can be added as needed. You can change the display of the chart components by moving them to other locations in the chart, resizing them, or by changing their format. You can also remove those chart components that are not needed to display your data.


The components of the above chart are numbered 1-7. We will discuss briefly each of them.

**1.** The chart area (the entire chart with all its components).

**2.** The plot area (the area bounded by the horizontal and vertical axes).

**3.** The data points (data points: Individual values or summaries plotted in a chart and represented by bars, columns, lines, pie or doughnut slices, dots, and various other shapes)

4. The horizontal and vertical axes (lines bordering the chart plot area used as a frame of reference for measurement). The y axis is usually the vertical axis and contains data. The x-axis is usually the horizontal axis and contains categories.)

5. The legend (box that identifies the patterns or colors that are assigned to the data values or categories in a chart.).

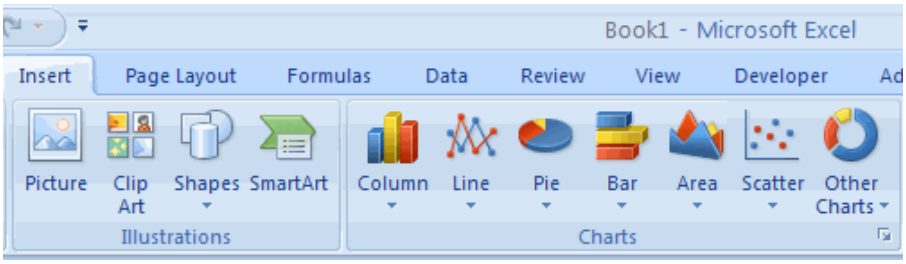
6. The chart title (text centered at the top and/or at the bottom of the chart).

7. Data label (a label that provides additional information about a single data point in the chart).

## 2. Creating a Chart.

To **create** a chart:

- Select the cells that contain the data that are to be used in the chart,
- Click the **Insert** tab on the Ribbon,
- In the **Charts** panel click the type of chart that you wish to create.



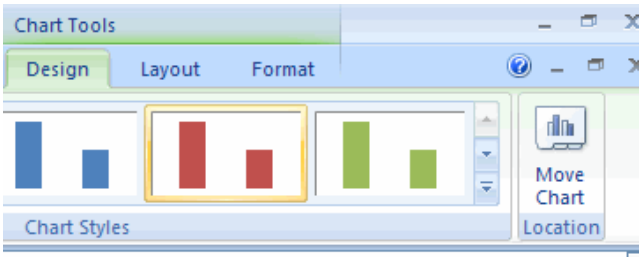

When you click anywhere in the chart area, and then hover the mouse pointer over one of these elements, Excel displays information about that element in a ScreenTip.

## 3. Modifying Charts.

After your chart is created, you can modify it. For example, you may want to change the way tit is displayed, add a chart title, move the chart to another location, move or hide the legend, or display additional chart elements.

To **move** the chart:

- Click the chart and drag it another location on the same worksheet, or
- Click the **Move Chart** button on the Design tab,
- Choose the desired location (either a new sheet or a current sheet in the workbook).

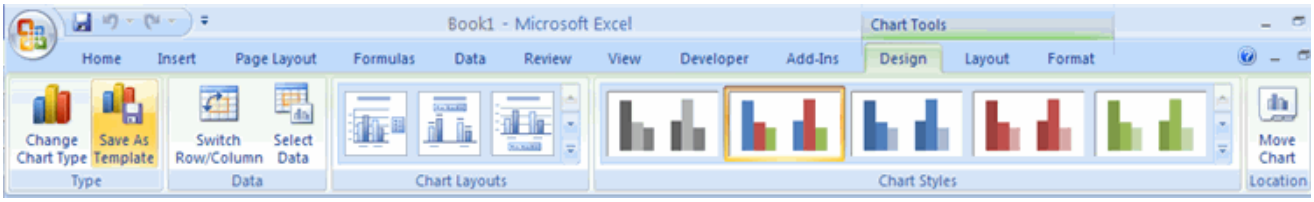



To **modify** the labels and titles:

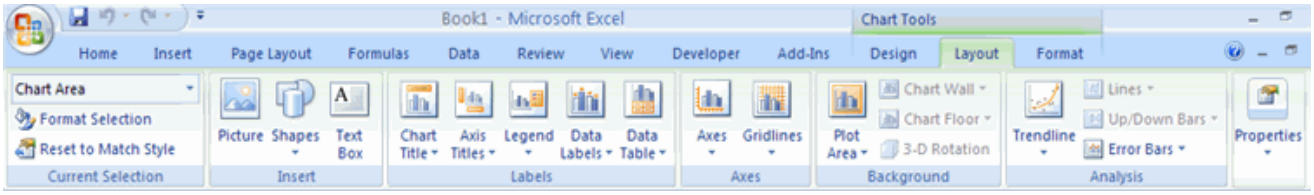
- Click the chart**,**
- On the **Layout** tab, click the **Chart Title** or the **Data Labels** button
- Change the **Title** and click **Enter**


The **Chart Tools** appear on the Ribbon when you click on your chart. The tools are located on three tabs: Design, Layout, and Format.

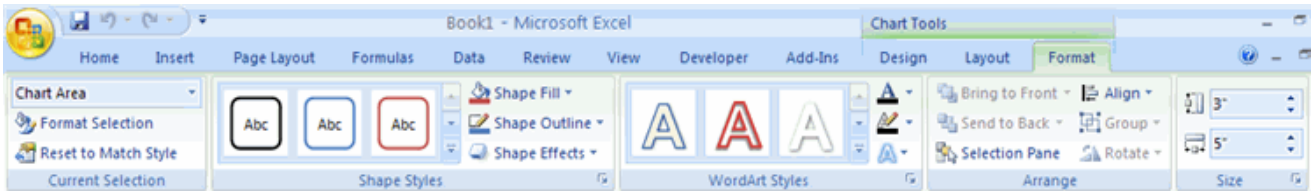
Within the **Design** tab you can control the chart type, layout, styles, and location.




Within the **Layout** tab you can control inserting pictures, shapes and text boxes, labels, axes, background, and analysis.

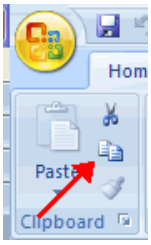



Within the **Format** tab you can modify shape styles, word styles and size of the chart.



To **Copy** a Chart to Word:

- Select the chart,
- Click **Copy** on the **Home** tab,




- Go to the **Word** document where you want the chart located,
- Click **Paste** on the **Home** tab.

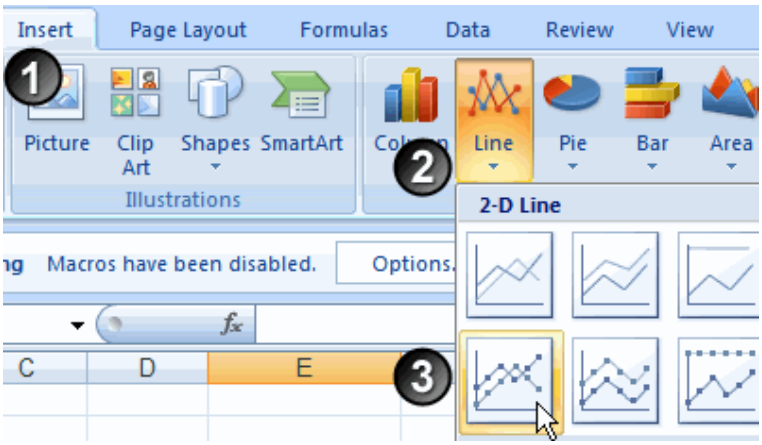
To **resize** a plot, select it, and then drag one of the eight handles located on the edges of the plot in the desired direction indicated by the pointer. To **delete** it, select the plot and then press *Delete.*

### 4. Example: Line Charts

Line chart (time plot) shows trends or changes in data over a period of time. Excel requires that the points on the horizontal axis are equally spaced. We will demonstrate how to construct a line chart using sales data for two stores.

| | A | B | C |
|---|---|---|---|
| 1 | **Sales (in thousands of dollars)** | | |
| 2 | | **Store 1** | **Store 2** |
| 3 | **January** | 10 | 69 |
| 4 | **April** | 5 | 53 |
| 5 | **July** | 208 | 76 |
| 6 | **October** | 145 | 74 |
| 7 | | | |

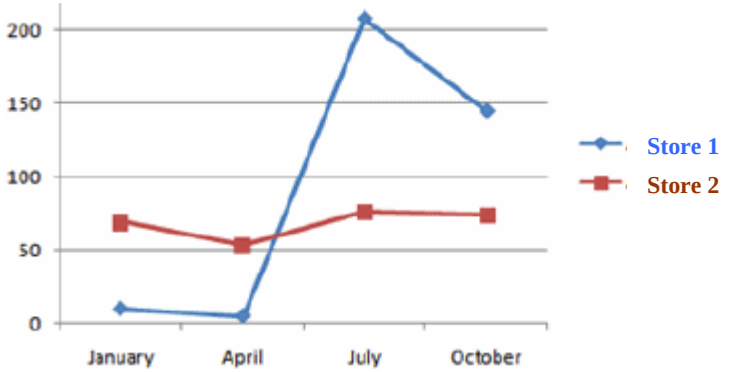
Drag select with the mouse button to highlight the cells containing the data to be included in the line graph. Highlight the block of cells from A2 to C6, which includes the column titles and the row headings.



**Selecting a Line chart Type:**

1. Click on the **Insert** ribbon tab.
2. Click on a chart category to open the drop down list of available chart types (Hovering your mouse pointer over a chart type will bring up a description of the chart).
3. Click on a chart type to select it.

In particular, if you choose **Insert > Line > Line with Markers**, you will obtain the following line chart:




**Formatting the Line Chart**

When you click on a chart, three tabs - the Design, Layout, and Format tabs are added to the ribbon under the title of **Chart Tools.**

**Choosing a style for the line chart:**

1. Click on the line graph.
2. Click on the **Design** tab.
3. Choose **Style 4** of the Chart Styles.

**Adding a title to the line chart:**

1. Click on the **Layout** tab.
2. Click on **Chart Title** under the **Labels** section.
3. Select the third option - **Above Chart**.
4. Type in the title "**Sales for Stores 1 and 2**"

**Changing the font color of the chart title**

1. Click once on **Chart Title** to select it.
2. Click on the **Home** tab on the Ribbon menu.
3. Click on the down arrow of the **Font Color** option to open the drop down menu.
4. Choose **Dark Red** from under the **Standard Colors** section of the menu.

**Changing the font color of the chart legend**

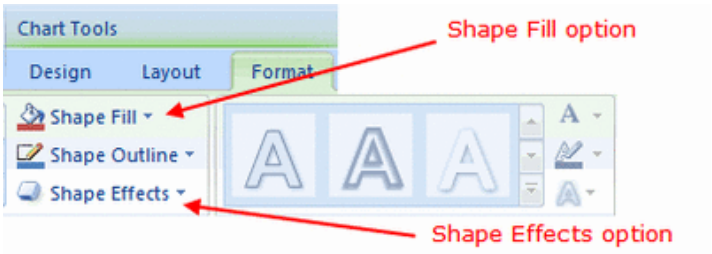
1. Click once on **Chart Legend** to select it.
2. Repeat steps 2 - 4 in the **Changing the font color of the chart title**.

**Changing the font color of the axis labels**

1. Click once on the months labels below the horizontal **X axis** to select them.
2. Repeat steps 2 - 4 in the **Changing the font color of the chart title**.
3. Click once on numbers beside the vertical **Y axis** to select them.
4. Repeat steps 2 - 4 above.

**Coloring the chart background**

1. Click on the graph background.
2. Click on the **Shape Fill** option to open the drop down menu.
3. Choose **Red, Accent 2, Lighter 80 %** from the **Theme Colors** section of the menu.




**Coloring the plot area background**

1. Click on one of the horizontal grid lines to select the *plot area* of the graph.
2. Choose the **Shape Fill > Gradient > From Center** option from the menu.

**Beveling the graph edge**

1. Click on the graph to select it.
2. Click on the *Shape Fill* option to open the drop down menu.
3. Choose **Bevel > Cross** from the menu.

At this point, your graph should be similar to the line graph shown on page 5.

## 5. Histograms

The data given below represent the examination scores of 50 students:

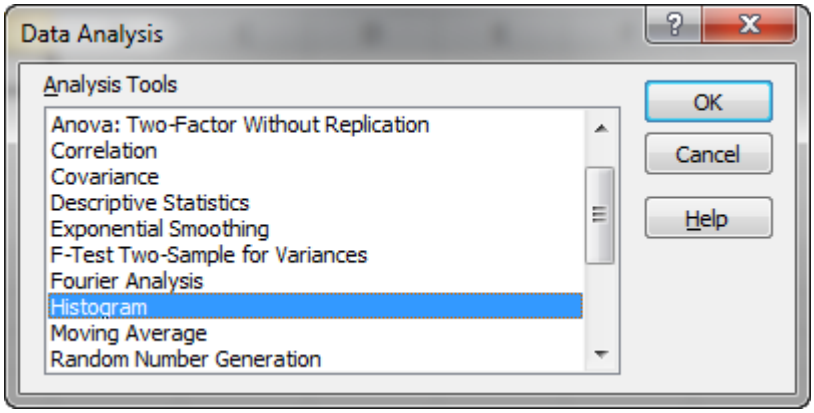
| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 75 | 43 | 42 | 75 | 84 | 36 | 65 | 59 | 63 | 34 |
| 78 | 37 | 99 | 66 | 90 | 79 | 80 | 89 | 68 | 57 |
| 28 | 55 | 79 | 88 | 76 | 60 | 77 | 49 | 92 | 83 |
| 71 | 78 | 53 | 81 | 77 | 58 | 93 | 85 | 70 | 62 |
| 80 | 74 | 69 | 90 | 62 | 84 | 64 | 73 | 48 | 72 |

Enter the data into a column called **SCORES** on your worksheet. We will use Excel to create a frequency distribution table for the data and a frequency histogram.

Before you use an appropriate feature in Excel, you have to determine the number of class intervals in the distribution. Otherwise, Excel will use a number of intervals approximately equal to the square root of the number of observations in your data set, with equal-width intervals starting and ending at the minimum and maximum values of your data set. Excel refers to the maximum value for each interval as a **bin**. If you define as the first bin 10, the corresponding class interval will contain all observations that are 10 or less. The bin 20 defines a class interval containing all observations that are above 10 and below or equal to 20. In other words, the left endpoint of a class interval is excluded, the right endpoint is included.

It's much quicker and easier to let Excel choose the class intervals (or equivalently the bins) than to determine them yourself, but Excel's choices are often not ideal. Let's choose in our data set class intervals of width 10 beginning at 0 and going to 100. Thus the corresponding bins are 10, 20, 30,...,100. Enter **Bin** as a label in cell C1, enter 10 in cell C2, and enter 20 in cell C3. select C2:C3. Drag the fill handle in the lower right of the selected range down to cell C11. The bins will be entered into C2:C11.

Choose the **Data** tab and the **Analysis** group and click on **Data Analysis.** If the **Data Analysis** group is not there, then see **Activating the Data Analysis Add-in** in Section 10. If the **Data Analysis** is available, you will see the **Data Analysis** dialog box as shown below:




Finally, choose option **Histogram** and OK this operation. Excel opens the **Histogram** dialog box.

Notice that the **Histogram** dialog box has two sections. The first is **Input** which is where you specify the location of data being graphed on the worksheet **(Input Range*)*** and the class intervals **(Bin Range).** With the cursor blinking in the **Input Range** box, bring the mouse pointer to cell A1 and drag the mouse down until range A1:A51 is selected. You can also simply type this range into the input box. As you can see Excel automatically converts this to absolute address.

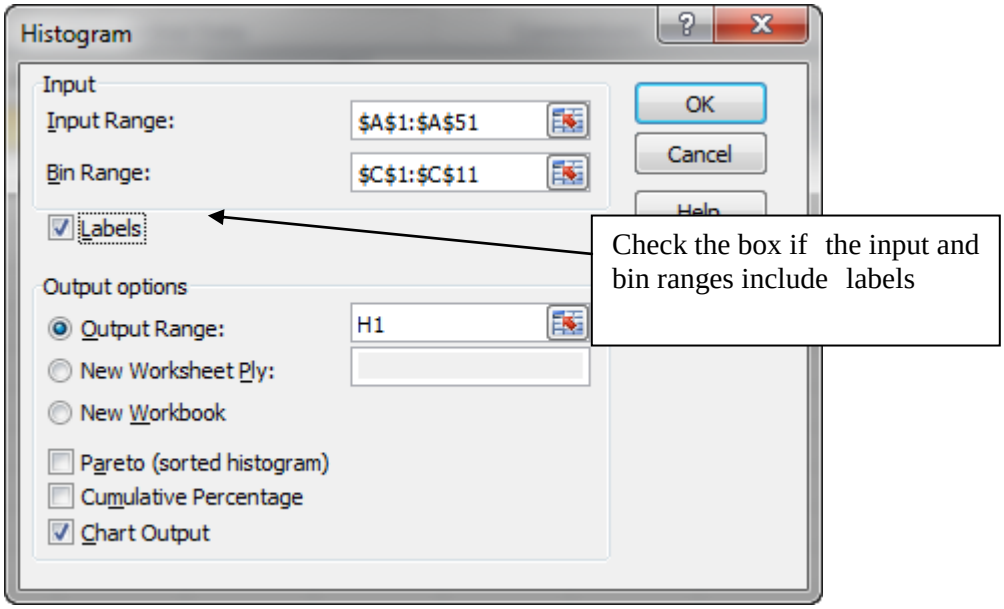
Press the **Tab key** to move to the **Bin Range** field. Do not press **Enter** or click OK until all the boxes are filled in the dialog box. Fill in C1:C11 for the **Bin Range** ( the label is included).

Check the box *Labels* to indicate that labels have been included in the addresses for the **Input Range** and **Bin Range**.

If you want the output to appear on the same worksheet as data, click the **Output Range** radio button and enter the address of the upper-left cell of the range where you want the output table to appear. Enter here H1. If you want the output to appear on a separate worksheet, check the option **New Worksheet.**

To obtain a standard frequency distribution table and chart, clear the **Pareto** check box. If this box is checked, the intervals are sorted according to frequencies before preparing the chart. Check the **Cumulative Percentage** box to obtain the cumulative frequency distribution table. In our example, we want the frequency distribution table only, so we leave the **Cumulative Percentage** box unchecked.

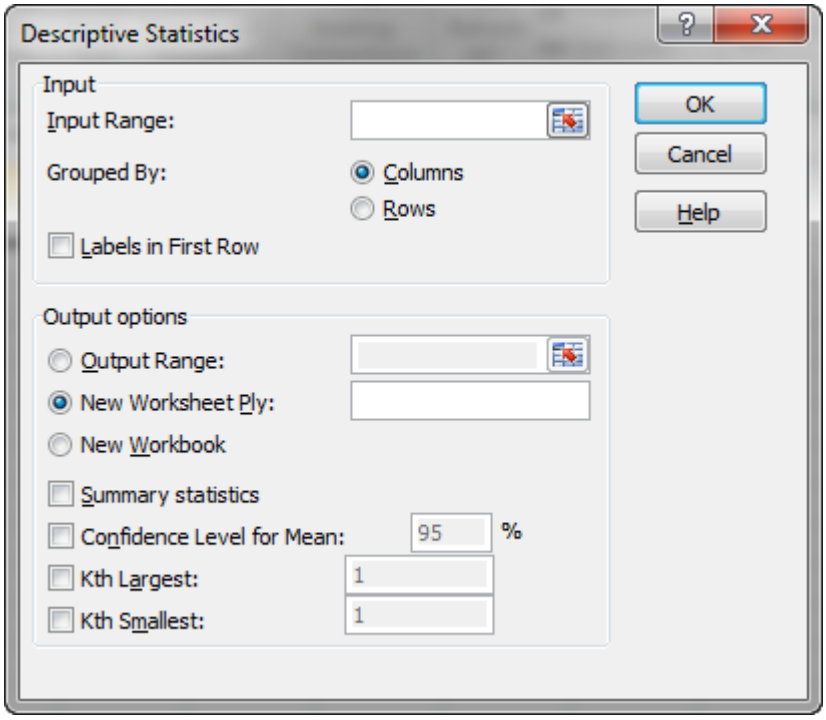
Now click the **Chart Output** check box to obtain a histogram in addition to the frequency distribution table. Without activating the option, Excel would only generate the frequency distribution table. The dialog box should look in the following way:




Finally, click OK. Excel puts the frequency distribution table and histogram on the worksheet. When you click the histogram, you can use the **Design**, **Layout**, and **Format** options of the **Chart Tools** to change the display of the histogram.

## 6. Descriptive Statistics Tool

The **Descriptive Statistic**s analysis tool provides measures of central tendency, variability, and shape. In order to access the tool, choose the **Data** tab, the **Analysis** group and click on **Data Analysis.** Excel opens the **Data Analysis** dialog box. Double-click on **Descriptive Statistics**. The **Descriptive Statistics** dialog box will appear.



In the **Input Range** enter the address of cells containing the data, including the labels for the data. In our example, either type A1:A51 or click on cell A1 and drag to cell A51. Excel automatically converts this to absolute address.

Press the **Tab** key to move to the next field of the dialog box. Do not press **Enter** or click *OK* until all the boxes are filled.

In **Grouped By** click **Columns** because your data are arranged in a column.

Select the check box Labels *in* **First Row** because your **Input Range** includes a label. If your data were arranged in rows with the labels in the first column, you would also check the box. When the analysis tool is told the first row contains a label, that label will be used in the output.

Leave the box **Confidence Level for Mean** unchecked. We will be discussing the concept later.

Select the check box **Kth Largest** if you want to know the kth largest value in the data set and type a number for k in the box. Request the fourth highest score in our data set. Analogously, select the **Kth Smallest** check box to get the kth smallest value in the data set and type a number for k in the box. Request the fourth lowest score.

In **Output Range** field enter the address of the upper-left cell of the range where the Descriptive Statistics output should appear. For our example with scores, select C1. Select the **Summary statistics** check box. This feature enables you to obtain very important numerical measures of center, variability, and shape. Now click OK. Excel computes the descriptive statistics and puts the results in the specified output range.

The **Descriptive Statistics** output contains three measures of central tendency: the *Mean* (69.0, the arithmetical average), the *Median* (72.5, the middle-ranked value), and the *Mode* (75, the most frequently occuring value).

The output table contains also several measures of variation: the *Range* (71, the maximum minus the minimum), the sample *Standard Deviation* (16.87287, the typical deviation of a data value from the mean), the sample *Variance* (284.6939, the square of the standard deviation). The *Standard Error* of the mean (2.386185) equals the sample standard deviation divided by the square root of the sample size. The standard error is a measure of uncertainty about the mean, and it is used for statistical inference.

There are two measures of shape reported in the output: the skewness and kurtosis. The *Skewness* coefficient is a measure of the lack of symmetry of the distribution. *Kurtosis* measures the heaviness of the tails of the data distribution. As these two measures of shape are not covered in the course material, we do not discuss them here in detail.

Click OK to obtain the **Descriptive Statistics** output. To adjust the width of a column to fit the longest entry, double-click the column heading border between the column and the next column.

## 7. Insert Function

Two other important measures of center and variation: the *First Quartile*, and the *Third Quartile* are not included in the Excel's **Descriptive Statistics** output. We will calculate the values using the **Insert Function** feature $\mathbf{f_x}$ and we add them to the output. We can access the **Insert Function** either by clicking on **Formulas** in the Ribbon.


or by clicking on the **fx in the formula bar.**


Switch to your **Descriptive Statistics** output. Select and move the part of the output with the first cell *Median* by one row down and the part of the output with the first cell *Mode* by one row down to make the room for two extra entries. In the cell below *Standard Error* type in *First Quartile* and in the the cell just below *Median*, type in *Third Quartile*. Make the cell that is supposed to include the numerical value of the First Quartile the active cell.

Press the **Insert Function** button and choose *QUARTILE.INC* in **Statistical functions** menu (in older versions of Excel this faction is named *QUARTILE*). The **QUARTILE.INC** dialog box is displayed with two input fields: *an array* field to enter the range of data, and *quart* field to enter the quartile number: 0-minimum, 1-First Quartile, 2-Second Quartile (Median), 3-Third Quartile, 4-Maximum. To calculate the first quartile enter in the first field: A2:A51 and 1 in the second field. The first quartile value of 59.25 is displayed in our output table. In the same way you can calculate the third quartile value equal to 80. Note that *QUARTILE.EXC* calculates the quartiles in a slightly different way.

## 8. Scatterplots

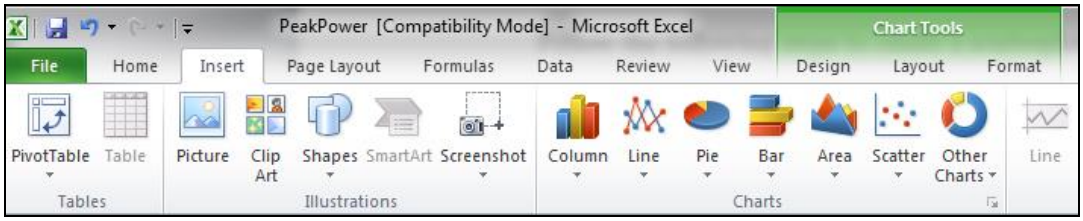
Scatterplots have already been discussed in *Lab 1 Instructions*. For your convenience we will cover here the basic commands necessary to obtain and edit scatterplots in Excel. In Excel a scatterplot is called an XY(Scatter) chart and it can be created and edited using *ChartWizard*.

Follow the following steps to create a scatterplot:

1. Arrange the data in adjacent columns on a worksheet with the x variable (for the horizontal axis) on the left and the y variable (for the vertical axis) on the right.

2. Highlight the columns that contain the data you want to represent in the scatter plot.

3. Open the *Insert* tab on the Excel ribbon. Click on *Scatter* in the *Charts* section to expand the chart options box. Select the appropriate chart, from this box.




4. After making this selection, the initial scatter plot will be created in the same worksheet. You can resize this chart window and drag it to any other part of the worksheet.

5. Make any formatting or design changes you wish in the **Design**, **Layout**, and **Format** tabs located under **Chart Tools** on the Excel ribbon.

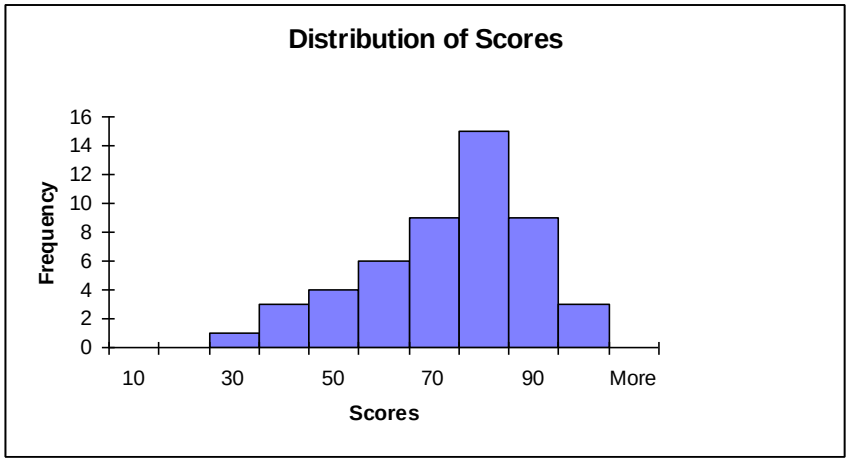
For example, to label the horizontal axis, select the **Layout** tab under **Chart Tools**, click on **Axis Titles** in the **Labels** section, then choose **Primary Horizontal Axis** and finally select **Title Below Axis**. A text box with the default wording **Axis Title** will appear on the chart. Click anywhere in that text box and edit the information. In order to change the chart title, click on the title to open the text box that contains it and edit it with your new description.

## 9. Distribution of Scores: Summary

The frequency distribution table, histogram and the **Descriptive Statistics** output for the 50 scores are now ready to print. Your histogram and the **Descriptive Statistics** output should look like this.

| ***Scores*** | |
|---|---|
| Mean | 69 |
| Standard Error | 2.386185 |
| First Quartile | 59.25 |
| Median | 72.5 |
| Third Quartile | 80 |
| Mode | 75 |
| Standard Deviation | 16.87287 |
| Sample Variance | 284.6939 |
| Kurtosis | -0.21292 |
| Skewness | -0.61043 |
| Range | 71 |
| Minimum | 28 |
| Maximum | 99 |
| Sum | 3450 |
| Count | 50 |
| Largest(4) | 90 |
| Smallest(4) | 37 |
| Confidence Level (95.0%) | 4.795215 |

The above summaries can be used to obtain the **five-number-summary** defined as: minimum, first quartile, second quartile (median), third quartile and maximum. Those five numbers can be used to obtain a boxplot for the data.




**Conclusions**

The histogram of scores is one-peaked and left-skewed. It indicates that the students did very well in this exam, most of them got pretty high scores.

**Remark:** Notice that in the above histogram, Excel puts the x-axis values at the center of each interval, although they are in fact the values that separate the intervals (bins).

## 10. Inserting the Excel Output into a Word Document

In your lab assignments you will be required to answer some questions about various data sets using Excel. You can paste the results obtained with Excel into a word processing document. Excel has a feature making it possible to export the whole spreadsheet or its part, charts and tables to a word processing document.

Start up **MS Word** and click on the **New** command in the **File** menu to create a Word document that is supposed to contain the answers to the questions in your lab assignment. Students should submit assignments with a cover sheet that contains only their name and course section. The student ID number should not appear on the cover sheet (FOIPP act). The assignments must be **typewritten**.

In order to paste the part of the Excel worksheet containing the results into the word processing document, copy it in Excel into the clipboard and then paste in into the word document using **Paste** in the **Home** menu (or Ctrl-C to copy the chart to Clipboard and Ctrl-V to paste the chart into the document).

## 11. Activating the Data Analysis Add-In

In order to use the **Data Analysis** feature in Excel, you have to load the **Analysis ToolPak** first. The **Analysis ToolPack** is a Microsoft Office Excel add-in (a supplemental program that adds custom features to Excel) program that is available when you install **Microsoft Excel**.

If the **Data Analysis** add-in does not appear at right-end of **Data** menu, then follow the instructions below:

1. Click the **Microsoft Office Button**  , and then click **Excel Options**.
2. Click **Add-Ins**, and then in the **Manage** box, select **Excel Add-ins**.
3. Click **Go**.
4. In the **Add-Ins available** box, select the **Analysis ToolPak** check box, and then click **OK**. If **Analysis ToolPak** is not listed in the **Add-Ins available** box, click **Browse** to locate it.
5. After you load the Analysis ToolPak, the **Data Analysis** command is available in the **Analysis** group on the **Data** tab.

In Excel 2003 the **Data Analysis** add-in should appear in the **Tools** menu. If not then, click **Add-Ins** on the **Tools** menu. In the **Add-Ins available** box, select the check box next to **Analysis Toolpak**, and then click **OK**. The **Data Analysis** command will be added to the **Tools** menu.