

# GeoGebra 4

---

Statistics

# Raw Data

---

Descriptive Statistics and Bar Chart

NOTE: For each new exercise below I suggest you open up a new GeoGebra file.

Begin this exercise by typing the following data set directly into the Input Bar thus:

$\text{MyDataSample}=\{63,80,54,70,54,72,63,66,76\}$

Once this data set has been entered (on pressing <ENTER> or <RETURN>) as a list it can then be included into the argument of many of GeoGebra's Statistical functions, some of which are illustrated below:

The screenshot shows the GeoGebra interface with the Algebra and Graphics views. The Algebra view displays a list of objects: MyDataSample, p, MyMean, MyMedian, MyMode, MyPerc, MyQ1, MyQ3, and MySD. The Graphics view shows a slider tool for p, with a value of 62. The Input Bar contains several formulas using the data set and the slider value. Annotations explain the naming convention and the use of the slider tool.

Original data set can be entered in to the Input Bar as shown. Notice that the set has been named "MyDataSample" for later convenience and, as it has curly brackets, is recognised as a 'list' object by GeoGebra

In-built GeoGebra statistical functions

Slider Tool

Slider value, p, (its name and its value) integrated into a formula

Type each into the Input Bar. Personalised names (recommended for clarity and written **without** spaces) are written with a trailing '=' sign followed by the required GeoGebra function acting on the original data set - as previously named 'MyDataSample'

Input:  $\text{MyDataSample}=\{63, 80, 54, 70, 54, 72, 63, 66, 76\}$

Input:  $\text{MyMean}=\text{Mean}[\text{MyDataSample}]$

Input:  $\text{MyMedian}=\text{Median}[\text{MyDataSample}]$

Input:  $\text{MyMode}=\text{Mode}[\text{MyDataSample}]$

Input:  $\text{MyPerc}=\text{Percentile}[\text{MyDataSample}, p / 100]$

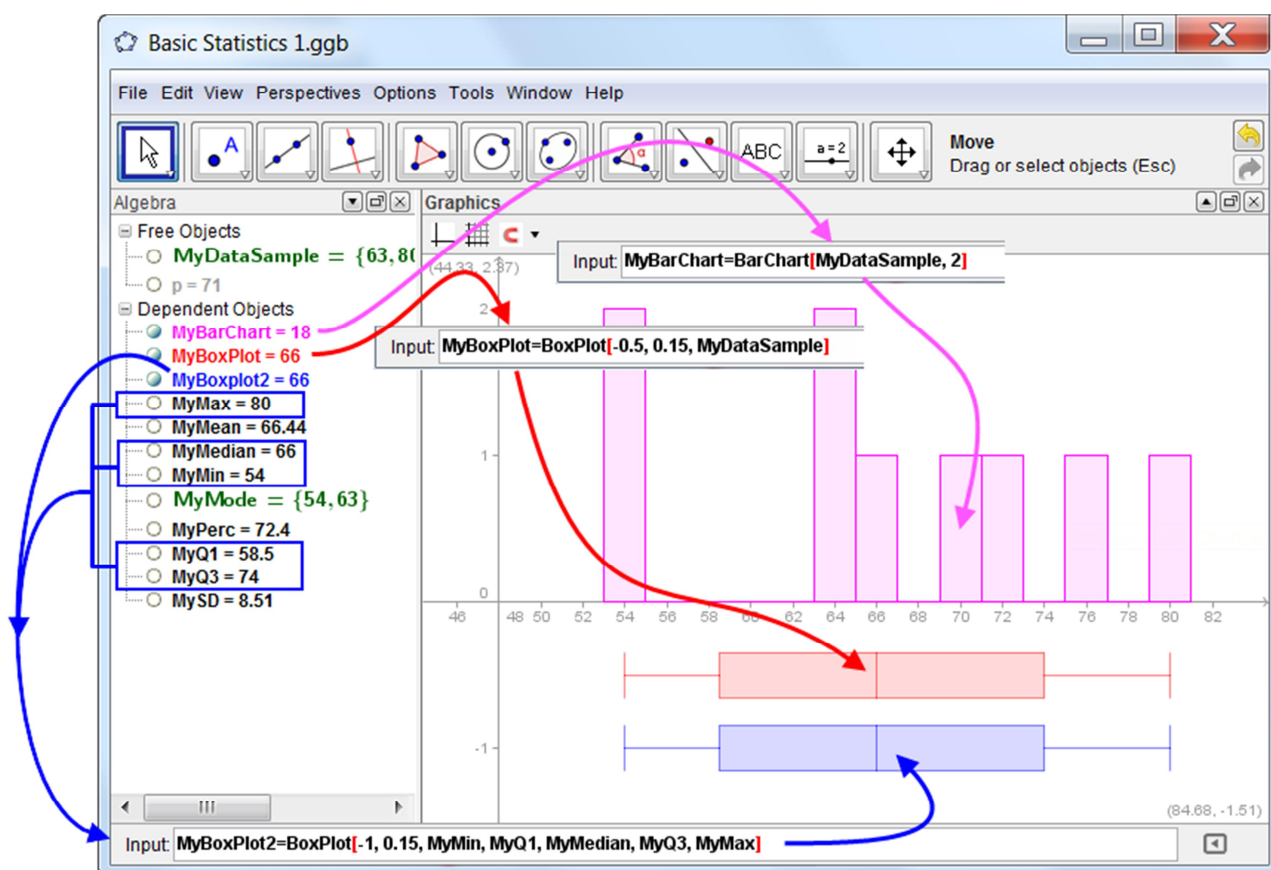
Input:  $\text{MyQ1}=\text{Q1}[\text{MyDataSample}]$

Input:  $\text{MyQ3}=\text{Q3}[\text{MyDataSample}]$

Input:  $\text{MySD}=\text{SD}[\text{MyDataSample}]$

Notice that if we had not named our dataset "MyDataSample" and instead just typed in  $\{63,80,54,70,54,72,63,66,76\}$  into the Input Bar, then GeoGebra would have automatically given this list a name – something along the lines "list1", "list2", for example. However, giving objects names in context often serves to increase clarity, especially if the file contains many functions and manipulations of them.

Follow the picture below to see how to draw a Bar Chart and also to see two methods to draw a Box and Whisker plot.



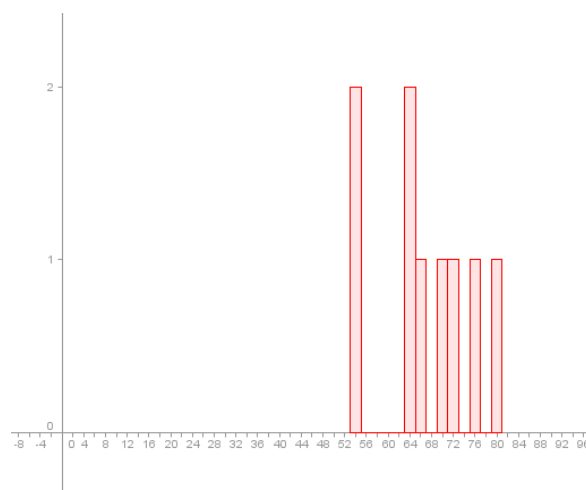
Note that we could have initially just typed into the Input Bar:

$$\text{BarChart}[\{63,80,54,70,54,72,63,66,76\},2]$$

This would have immediately given us a Bar Chart of bar-width 2 units as shown here.

However, in using this approach we would not have created and defined a list structure containing our data within GeoGebra. Consequently, we would have nothing to use as an argument for the calculation of any further statistical measures.

Thus it is of use for impromptu plots but that's about all!



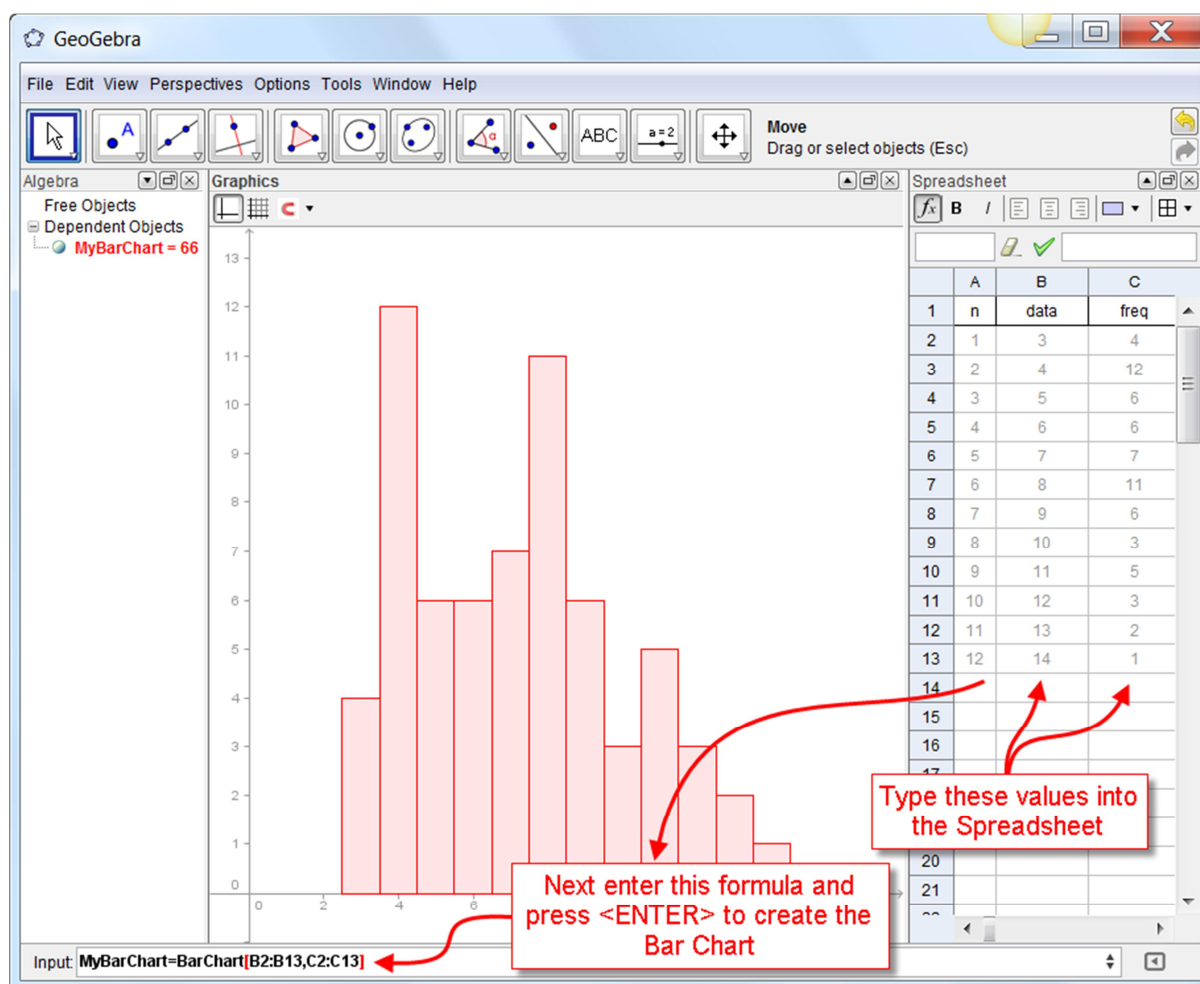
# Frequency Data

## Bar Chart

We will now consider several methods to create a Bar Chart. One based upon using GeoGebra's in-built spreadsheet, others not.

### Using the Spreadsheet:

Press <CTRL>+<SHIFT>+S (or go through the View menu) to open up the spreadsheet window and type in the data and associated frequencies in columns B and C, respectively.



When you go back to the spreadsheet and alter your data you should see your Bar Chart update accordingly

Alternatively, but starting from two spreadsheet columns for data and frequency we can create list objects from these two columns as shown below:

Repeating with the frequencies gives:

Because the 'Dependent Objects' option was chosen in the 'Create List' dialogue box, if any 'data' or 'Freq' spreadsheet column values are altered then so are the respective lists and all other objects dependent upon them.

First highlight the range of cells you want to create a list object from

Then give the list a contextualised name

Algebra

Free Objects

Dependent Objects

MyData = {3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14}

MyFreq = {4, 12, 6, 6, 7, 11, 6, 3, 5, 3, 2, 1}

GeoGebra

File Edit View Perspectives Options Tools Window Help

Algebra

Free Objects

Dependent Obj

MyData

Spreadsheet

A	B	C
n	data	Freq
1	3	4
2	4	12
3	5	6
4	6	6
6	7	7
7	8	11
8	9	6
9	10	3
	11	5
	12	3
	13	2
	14	1

Create List

Name

MyData

Options

Dependent Objects  Free Objects

Row Order

Preview

{3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14}

Cancel Create

Once lists are created then, once again, we can use them within other functions. This gives us much greater flexibility.

Using Lists:

Type the following two named lists into the Input Bar

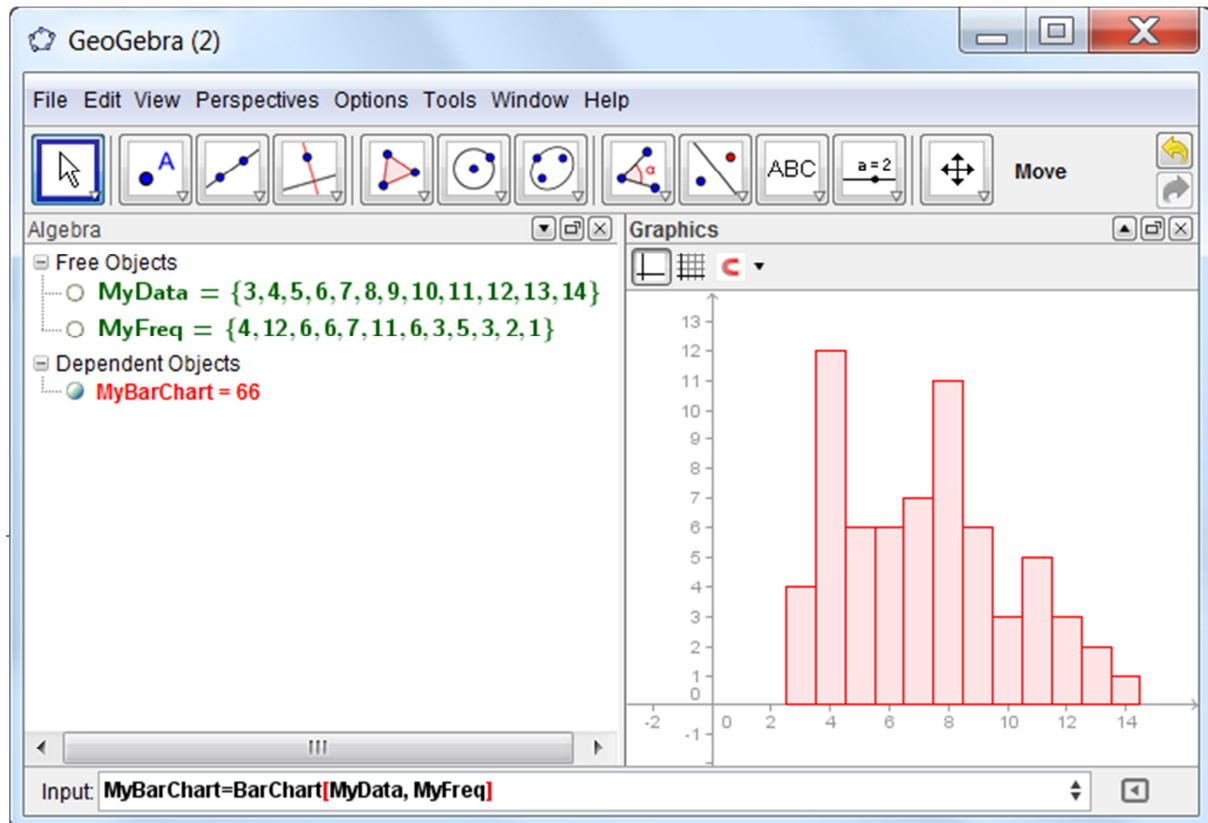
Input: **MyData**={3,4,5,6,7,8,9,10,11,12,13,14}

Input: **MyFreq**={4,12,6,6,7,11,6,3,5,3,2,1}

Then type the following into the Input Bar to create your Bar Chart

Input: **MyBarChart**=BarChart[MyData, MyFreq]

Giving:



If you now wish to copy the numbers from your data list and frequency list **into the spreadsheet** then type the following into the Input Bar:

Input: **FillColumn**[2,MyData]

and similarly:

Input: **FillColumn**[3,MyFreq]

'2' represents the 2nd column or Column 'B'

Notice now that the dynamic link that was present between the spreadsheet values and the list, as illustrated in the previous method, is no longer present when attempting updating in either direction.

# Frequency Data – Uneven group widths

## Histogram

Somewhat more work is required for Histograms but the output pictures are generally very good and GeoGebra is one of the few pieces of mathematical software in which Histograms are an in-built feature.

The syntax for the various forms of Histograms, taken from the Input Help (on GeoGebra available functions and their syntax) window are:

Histogram[ <List of Class Boundaries>, <List of Heights> ]

Histogram[ <List of Class Boundaries>, <List of Raw Data>, <Use Density>, <Density Scale Factor (optional)> ]

Histogram[ <Boolean Cumulative>, <List of Class Boundaries>, <List of Raw Data>, <Use Density>, <Density Scale Factor (optional)> ]

We show below the steps in creating the first of these above, first through the use of the spreadsheet and secondly through the manipulation of List Objects

### Using the Spreadsheet:

The screenshot shows the GeoGebra Histogram tool interface. On the left, a histogram is displayed with a y-axis from 0 to 24 and an x-axis from 0 to 40. On the right, a spreadsheet is visible with columns A, B, C, and D. Red callouts point to specific parts of the interface:

- "Area" of each bar: points to the area of a bar in the histogram.
- Width of each bar: points to the 'Group width' column in the spreadsheet.
- Height of each bar (Frequency Density): points to the 'Bar Height' column in the spreadsheet.
- Group 8 - 10 Freq = 28: points to the frequency value 28 in the spreadsheet.
- Notice that in the data values column we must have one more row than in our associated Frequencies column. This is the set-up that GeoGebra recognises: points to the extra row in the data column.

	A	B	C	D
1	Data	Freq	Group width	Bar Height
2	0	23	A3 - A2	B2 / C2
3	1	43	A4 - A3	B3 / C3
4	8	28	A5 - A4	B4 / C4
5	10	42	A6 - A5	B5 / C5
6	14	33	A7 - A6	B6 / C6
7	19	12	A8 - A7	B7 / C7
8	24	24	A9 - A8	B8 / C8
9	40			
10				
11				
12				

	A	B	C	D
1	Data	Freq	Group width	Bar Height
2	0	23	1	23
3	1	43	7	6.14
4	8	28	2	14
5	10	42		10.5
6	14	33	5	6.6
7	19	12	5	2.4
8	24	24	16	1.5
9	40			
10				



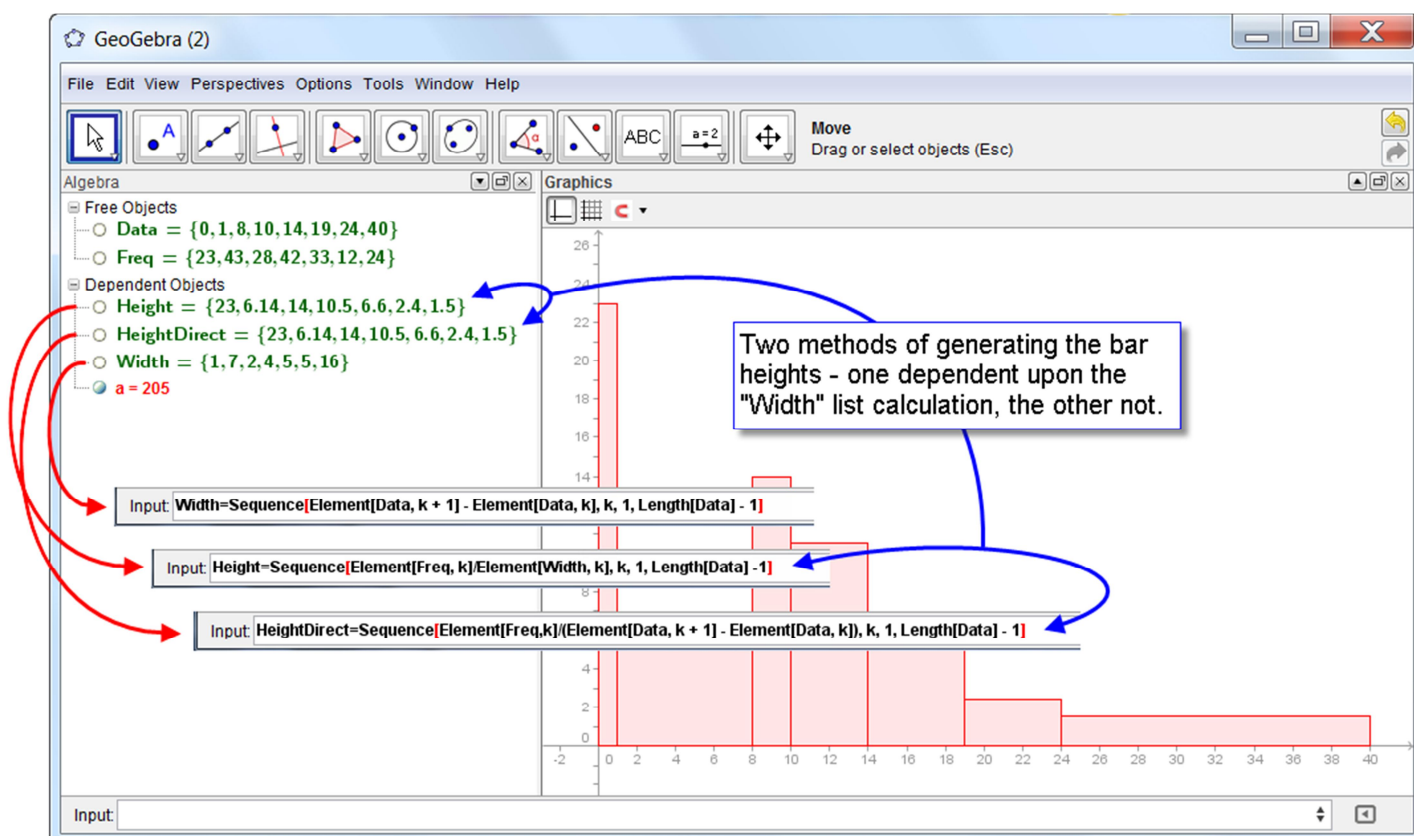
Note that, as in a method shown above, I used columns A and D indirectly in the Histogram Plot. This was only after I had first “converted” them into List Objects, namely: “Data” and “Heights” relating to columns A and D, respectively. This allows for updates in the spreadsheet cells to dynamically alter values in the List Objects and therefore in the Histogram too.

Note further:

- (a) Entering formulae into cells is identical to when doing so in Excel. Begin by typing an Equal sign and then either click on the cell you wish to include in the formula or write in the cell’s address.
- (b) Dragging a formula into other cells is again similar to when doing so in Excel. Grab the cell “Fill-handle” towards the bottom right of a cell having a formula within it and drag down as far as you need.
- (c) Alternating between the cell values and cell formulae is done via the “Options → Algebra Description” on the Toolbar and then select either the “Value” or “Definition” options.

Using Lists:

This is getting to be a little more advanced in the use of List Objects now so you may wish to ignore it. In brief it is using the “Sequence” function to put several values in turn into an expression and then the “Element” function to pick out individual components (values) from a List Object. In essence, we are doing the Height = Area/Width calculation one element at a time as the sequence progresses through each of its terms.



# What doesn't currently work (easily!)

---

## Descriptive Statistics on Non-Raw Data

We have seen that a Bar Chart can immediately be created from the frequency distribution

x	f
4	1
5	4
6	6
7	12
8	5
9	2

through the command:

```
BarChart[{4,5,6,7,8,9},{1,4,6,12,5,2}]
```

being typed into the Input Bar. Or, similarly, if 'x' and 'f' are first entered, with names of our choosing, as List Objects, say 'MyX' and 'MyF', then the command:

```
BarChart[MyX,MyF]
```

will also produce the Bar Chart.

Unfortunately, unlike with the raw data, there are currently no *in-built* Descriptive Statistics functions, such as 'Mean', Median, etc., that are able to handle frequency distributions.

However, there is a work-around using List Objects that is able to 'convert' a frequency distribution into raw data set. The familiar Descriptive Statistics functions can then be applied to this raw data set.

Overleaf is a typical example:

Can you understand it? If you are a programmer of sorts then you may find it helpful to interpret each "Sequence" function as a "For loop", for example

```
For i = 1 to 10
  For j = 1 to 5
    < Now do something in here dependent upon i and j >
  Next j
Next i
```

Basic Statistics 2.ggb

File Edit View Perspectives Options Tools Window Help

Algebra

Move  
Drag or select objects (Esc)

Graphics

Input

```

Fres Objects
MyDataSample = {12,13,14,15,16,17,18,19,20,21,22,23}
MyFreqData = {3,4,1,0,0,5,7,2,5,10,3,4}
p = 44
Dependent Objects
MyFreqBarChart = 44
MyMax = 23
MyMean = 18.61
MyMedian = 19.5
MyMin = 12
MyMode = {21}
MyOrderedData = {{12,12,12}, {13,13,13}, {14}, {}, {17,17,17,17}, {18,18,18,18,18,18}, {19,19}, {20,20,20,20,20}, {21,21,21,21,21,21,21,21}, {22,22,22}, {23,23,23,23}}
MyPerc = 18
MyQ1 = 17
MyQ3 = 21
MyRawData = {12,12,12,13,13,13,14,17,17,17,18,18,18,18,18,19,19,20,20,20,21,21,21,21,21,21,21,22,22,22,23,23,23}
MySD = 3.28
List(MyOrderedData: Sequence(Sequence(Element(MyFreqData, dj), d, 1, Length(MyDataSample)))
List(MyRawData: Join(MyOrderedData))

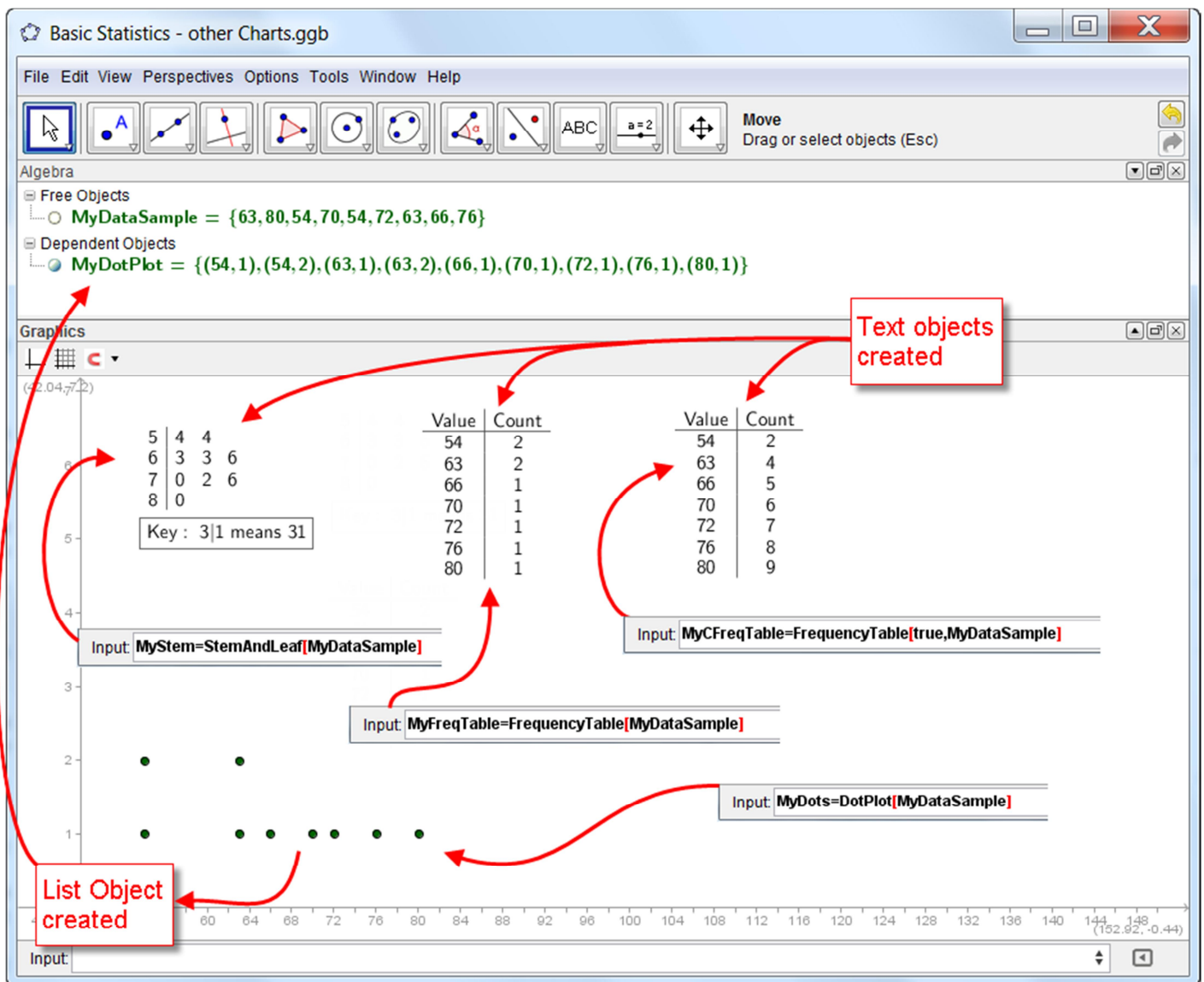
```

Data Value	Frequency
12	3
13	4
14	1
15	0
16	0
17	5
18	7
19	2
20	5
21	10
22	3
23	4

# Raw Data - again

## Other charts

Once we have our raw data list defined somehow then we can, with a single function call in the Input Bar create:




Other statistics topics for you to investigate may be found at:

[http://wiki.geogebra.org/en/Statistics\\_Commands](http://wiki.geogebra.org/en/Statistics_Commands)

# Multivariable Statistics

## Creating Lists from previously drawn points

This is a useful facility if you wish to create your own set of points that have particular correlations, for example: weak-positive correlation, etc. It is a simple matter to put any number of free points on to an axis pair in the Graphics Window by first choosing the point Tool  from the 2<sup>nd</sup> Tool Group Icon set. Now click on the screen about 10 times or so, creating a random assortment of points having a weak-positive correlation, as the diagram below illustrates:

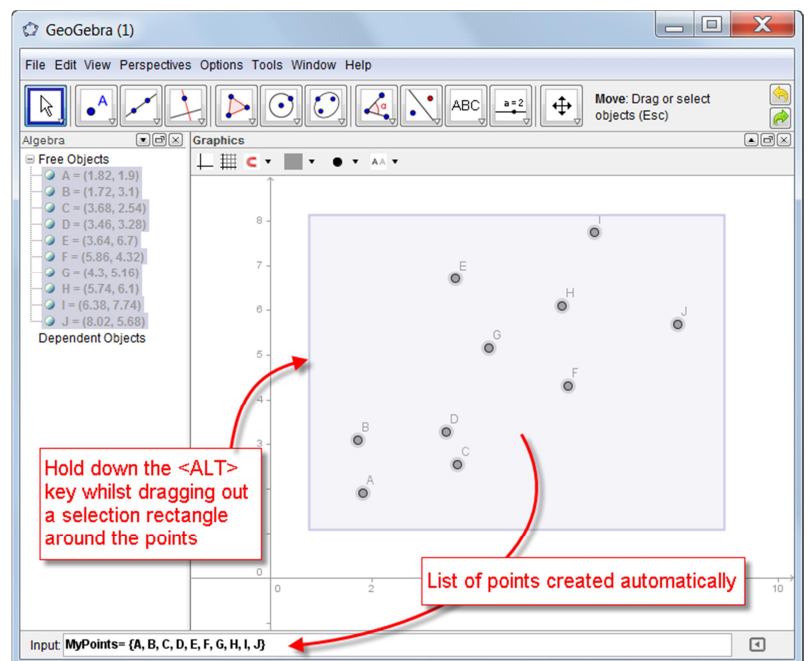
Now make a list by selecting your points with the mouse whilst holding down the <ALT> key. When you let go of the mouse button you should see in the Input Bar the list {A, B, C,...}.

Click in front of this list and enter a name for it.

For example:

`MyPoints= {A, B, C, D, E, F, G, H, I, J}`

Press enter to create this list named 'MyPoints'



Once again you can elect to do any resulting statistics by using either the Spreadsheet or List Objects to manipulate your data.

### Using the Spreadsheet:

#### Method 1

In the Input Bar write: `FillColumn[1,MyPoints]`

This will send the coordinates of the points into the first column, Column A, in the spreadsheet.

#### Method 2:

Select the points in the Algebra View (Click on the first point and hold down <SHIFT>, then click on the last point to select them all) then physically drag them from there into the spreadsheet.

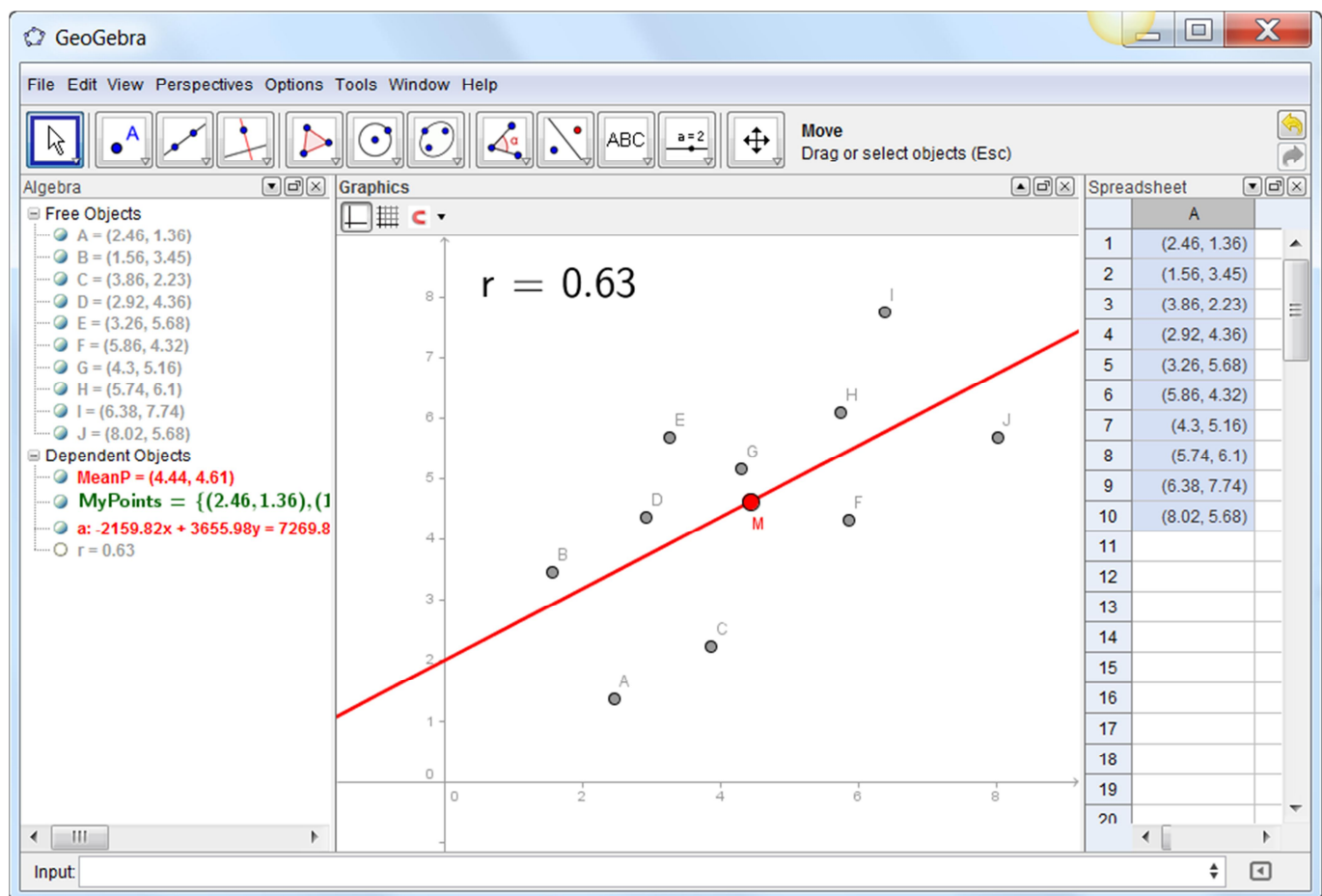
*\*Note that, if you hold down <CTRL> whilst dragging then further options are available to you\**

## These extra options are VERY IMPORTANT! Experiment with this operation What do you notice?

Once you have found out how to create a set of dependent points in a column in your spreadsheet (that is, they update when points are moved) you can experiment with the following functions:

- CorrelationCoefficient[A1:A10]
- FitLine[A1:A10]
- (MeanX[A1:A10], MeanY[A1:A10])

Note that the range 'A1:A10' is used directly from the column of points within the GeoGebra file and its spreadsheet as illustrated below:



# Probability Distributions and Random Numbers

## Creating a data set approximating a known law

Say we have a data set which, once plotted, suggests an underlying connection between  $x$  and  $y$ . Then

- How can we investigate possible relationships between  $x$  and  $y$
- How can we create such a data set in the first instance?

## Recording to the Spreadsheet

Set up a slider  $k$  between  $-2$  and  $5$  with increments of  $0.5$ . Once on the screen set it to its lowest value.

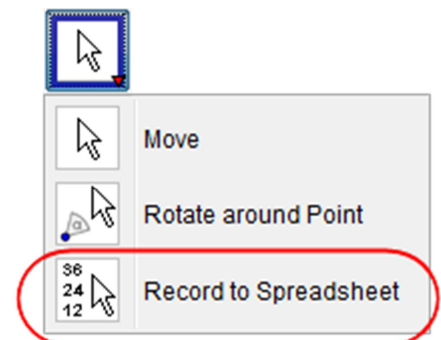
In the Input Bar type:  and press <ENTER>

Next, in the Input Bar type

What happens now when the slider  $k$  is altered?

Follow these instructions to record to the spreadsheet:

- Reset the slider  $k$  to its lowest value.
- From the 1<sup>st</sup> Toolbar Icon set drop-down list select 'Record to Spreadsheet'.
- Select the Point  $P$  (newly created) on the curve
- Slowly move the slider  $k$  increasingly through its values.
- Press the escape button <ESC> to stop recording



You should now see a set of points created automatically in columns A and B of your spreadsheet (with headers  $x(P)$  and  $y(P)$ )!

In Column C, row 2, create a new point by typing the following as shown to the right.

Then copy this formula down accordingly.

A screenshot of a spreadsheet window titled 'Spreadsheet'. The formula bar shows the formula  $=(A2, B2 + \text{RandomNormal}[0, 1])$ . Below it is a table with columns A, B, and C. Row 1 has headers  $x(P)$  and  $y(P)$ . Row 2 has values  $-2$  and  $4$  in columns A and B respectively, and the formula  $=(A2, B2 + \text{RandomNormal}[0, 1])$  in column C. Rows 3 through 8 contain numerical values for columns A and B, and blank cells for column C.

	A	B	C
1	$x(P)$	$y(P)$	
2	$-2$	$4$	$=(A2, B2 + \text{RandomNormal}[0, 1])$
3	$-1.5$	$1.75$	
4	$-1$	$0$	
5	$-0.5$	$-1.25$	
6	$0$	$-2$	
7	$0.5$	$-2.25$	
8	$1$	$-2$	

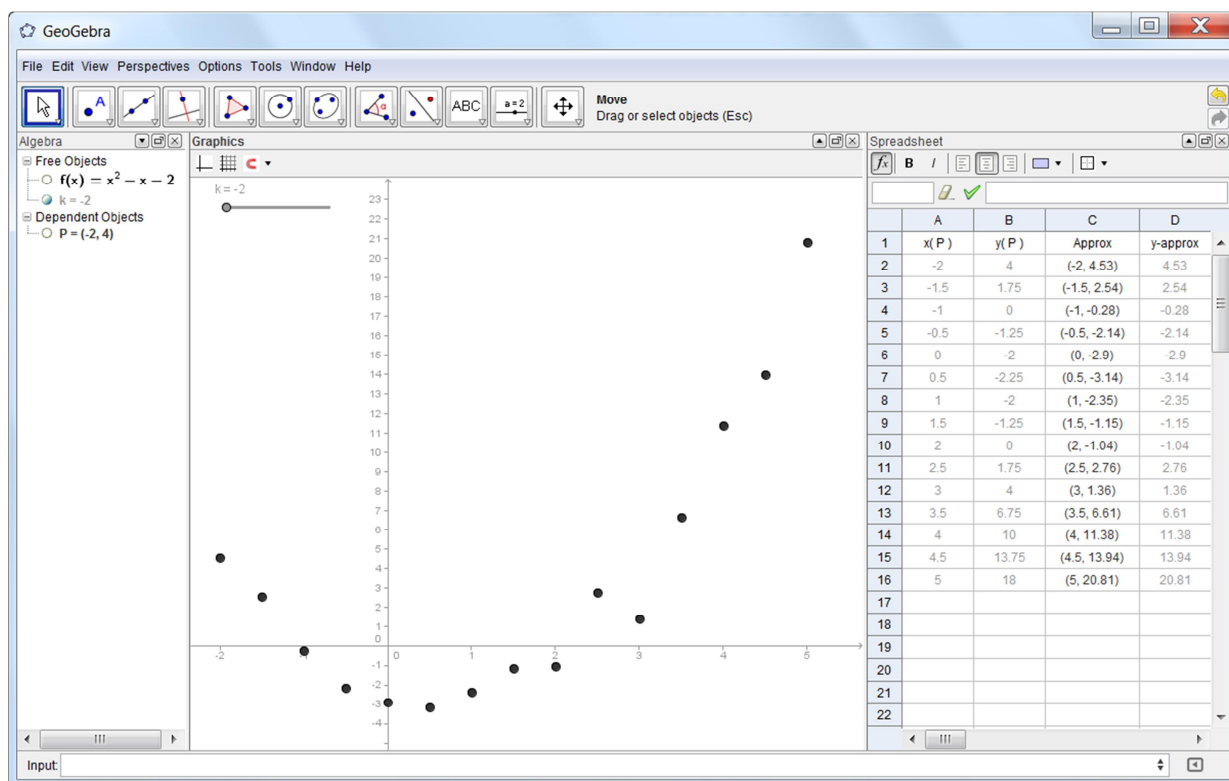
It is convenient to click into cell C2 first and to then type the formula into the Spreadsheets own Formula Bar



What you should have created is a set of points in the Graphics Window approximating a quadratic function. In fact the deviation from the actual law behaves randomly according to the Normal Distribution:  $N(0,1)$ . This is a data set on which you can begin your analysis – such as investigating repeated differences.

To consider just the y-values in a new column you will find the formula '= y(C2)' helpful once entered into the cell D2 and then copied down the column.

One such picture (in which the original graph and point P are hidden) is:



Pressing the F9 special function key on your keyboard should re-calculate the normally distributed random variable in each cell and update the points in the Graphics window.

### Using Lists:

The above work could have been generated using the following steps:

- First define the function, as above, in the Input Bar:
- Next, type the following directly into the Input Bar:

The red brackets serving to illustrate that it is a sequence of (x, y) coordinate points being created by the Sequence function.





