

MATHEMATICS 201-510-LW

Business Statistics

Martin Huard

Fall 2008

Organizing Data with Excel

Bar Graphs - Example 1

Suppose that an ecologist desires to know the composition of a forest in a particular region. To do this, a random sample of 50 trees is taken from the forest, and the species of each tree is noted. What we will now do is compile the information obtained by the ecologist and graphically represent it.

Go to Excel.

In cell A1 write in Bold 16 “Business Statistics”. In cell A2 write “Organizing Data with Excel”, in cell A3 write “By: your name” and in cell A4 write the date, all in the default font. Label Sheet 1 “Forest Composition”. Save your work under LAB 4. Don’t forget to periodically save your work!

In cell A6 write “Sample” in bold. Open the worksheet “Data for Organizing Data with Excel” from my web site and copy the sample results from the worksheet “Forest Composition” in cells A7:A56.

Frequency Distribution Table

The first thing to do is to compile the result in a table. In cells C6:E13, do the following:

| Composition of Forest | | |
|------------------------------|------------------|---------------------------|
| Tree | Frequency | Relative Frequency |
| Pine | | |
| Maple | | |
| Spruce | | |
| Oak | | |
| Elm | | |
| <i>Total</i> | | |

where the width of the columns are adjusted so that the words fit properly.

To find the frequency of each tree, we will get EXCEL to do the counting for us. Go to cell D8. In the functions (the f_x button), go to STATISTICAL (in category) and COUNTIF (in function). Click in the space for the RANGE, and with your mouse, blacken the list of trees A7:A56. Put \$ signs before the 7 and the 56, so that this block will stay fix. Next, in CRITERIA, click on the cell C7 where Pine is written. Click OK. Copy this formula, with your mouse, for the other trees. For the cell D12, use the Σ button to find the sum. For the relative frequency column, we divide the number of trees of one kind by the total number, and express the answer as a percentage. Use a formula making references to cells for this, and use your mouse to copy the formula down the column. For the last cell, you should obtain 100% when adding the percentages.

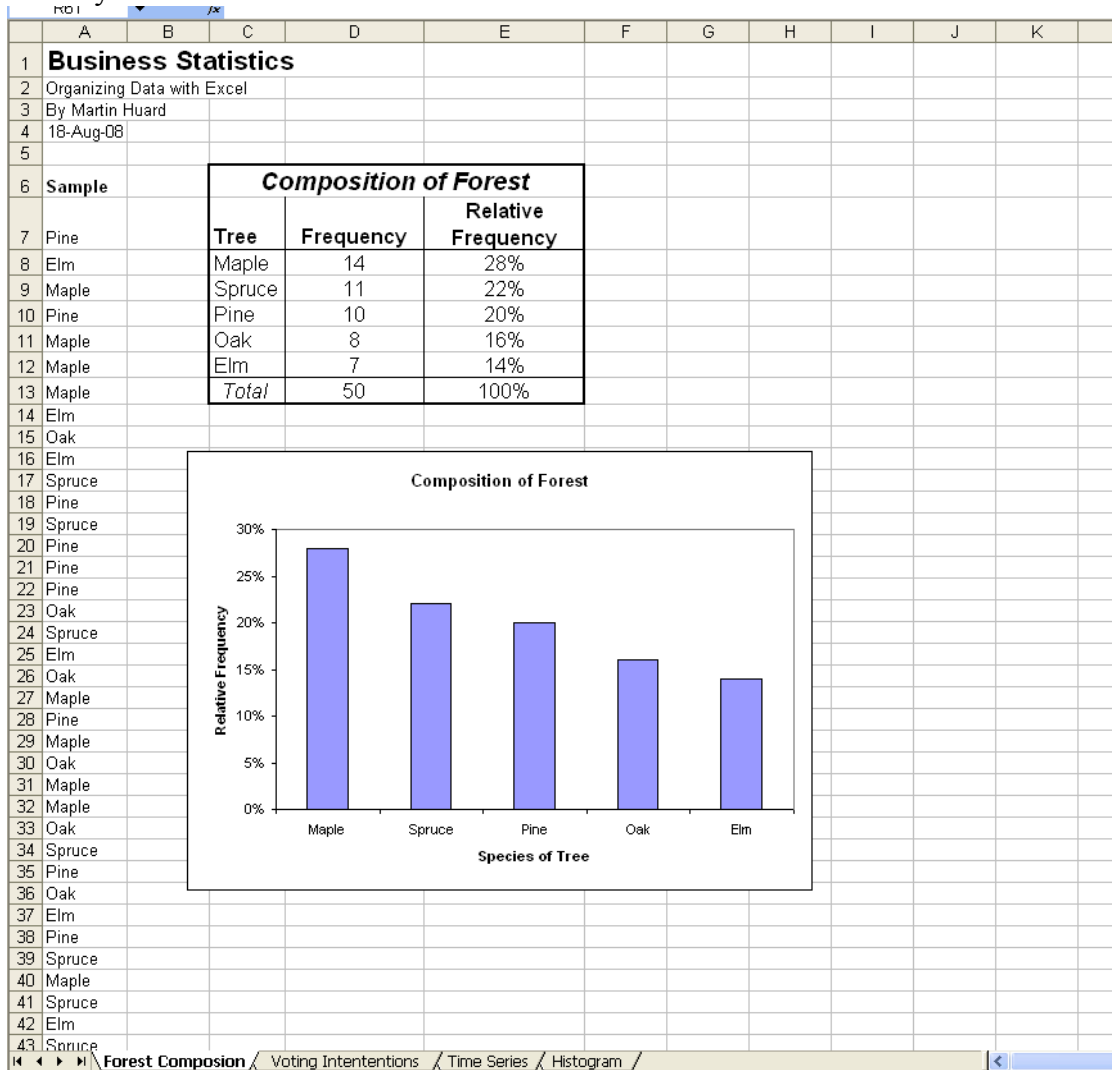
Graphing

For graphing, start by blackening the values in the relative frequency column, the cells E8:E12. Go to INSERT and choose a COLUMN graph (the 2-D one). To have the name of trees appearing below each column, click on SELECT DATA and on EDIT in the HORIZONTAL (CATEGORY) AXIS LABELS. Blacken the list of trees, that is, the cells C8:C12. To insert titles, either choose the appropriate CHART LAYOUT and give appropriate titles, or add each title using the appropriate tab. Get rid of the Legend and the Gridlines.

Pareto Diagram

To make a Pareto Diagram, we need to put the rectangles in decreasing order of height. To do this, blacken the cells C7:E12 and go to DATA – SORT – SORT BY - FREQUENCY (the option MY DATA HAS HEADERS must be checked). Also, the order must be LARGEST TO SMALLEST. The graph and table should automatically me updated.

Here is what your worksheet should look like.



Pie Diagrams and Bar Graphs - Example 2

Suppose that Gallup Polls conducted a survey on the voting intentions of Quebecers at the next Provincial Elections. In their random sample, they obtained the following responses.

| Party | Frequency |
|-------|-----------|
| PLQ | 355 |
| PQ | 348 |
| ADQ | 336 |

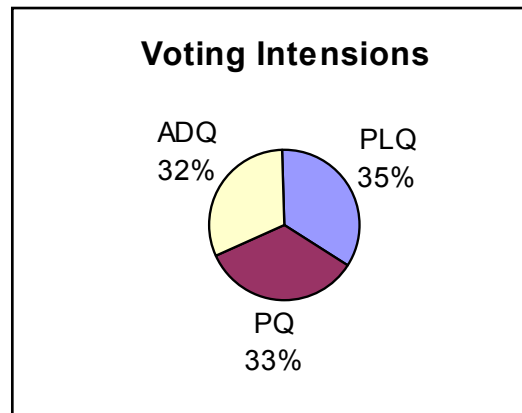
Go to Sheet 2 and rename it “Voting Intentions”. In cells A1:A4 do a heading similar to that of Sheet 1. From the data above, construct a frequency distribution table in cells A6:C11 that looks like:

| Voting Intentions | | |
|-------------------|-----------|--------------------|
| Party | Frequency | Relative Frequency |
| PLQ | 355 | |
| PQ | 348 | |
| ADQ | 336 | |
| <i>Total</i> | | |

Fill in the column Relative Frequency and find the two totals. All this should be done using formulas with reference to other cells.

Pie Diagrams

Make a Pie Diagram in the same way you made the bar graph in example 1. It should look like this.

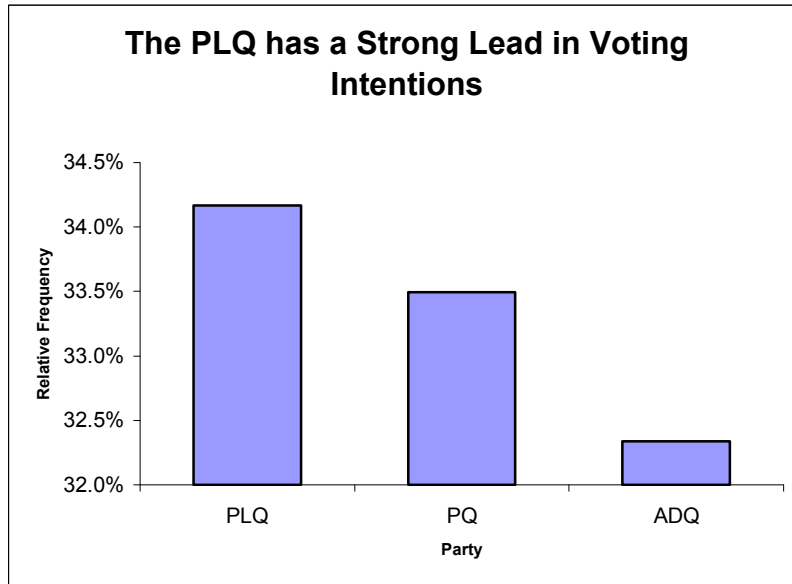


To see the percentages and the labels, choose the appropriate CHART LAYOUT.

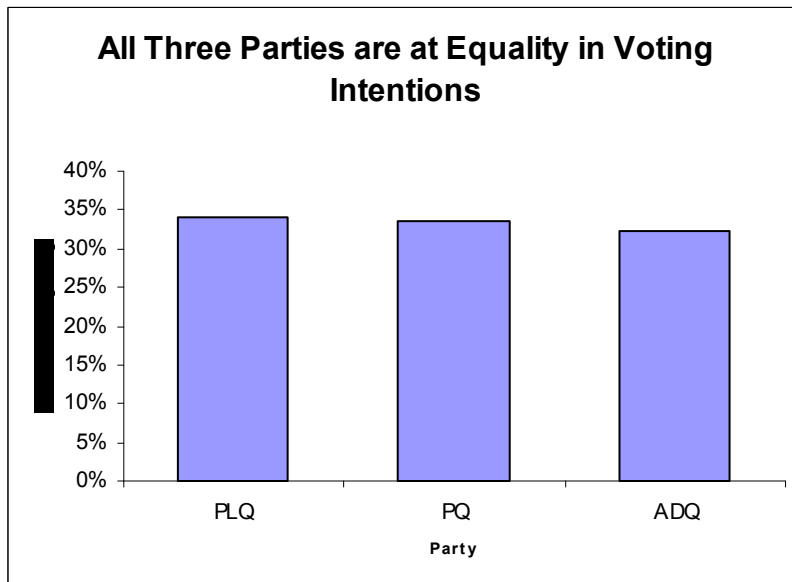
How to Cheat with Bar Diagrams

Make a Bar Diagram in the same way as in Example 1.

Suppose we wish to publish the diagram in the paper with the headline “The PLQ has a Strong Lead in Voting Intentions”. To support this statement, we *decrease* the range of values for the frequencies by changing the minimum and maximum values on the vertical axis. Go to AXES – PRIMARY VERTICAL – AXIS OPTIONS and change the minimum and maximum to 0.32 and 0.345 respectively. Your results should look like this.



Suppose that we now want the headline to be “The Three Parties are at Equality in Voting Intentions”. To observe this, we make similar modifications as before, except that we *increase* the range of values for the frequencies by changing the minimum and maximum values to 0 and 0.4 respectively. Note that there are no decimals here on the vertical axis, since the range of values given on the axis renders them meaningless.



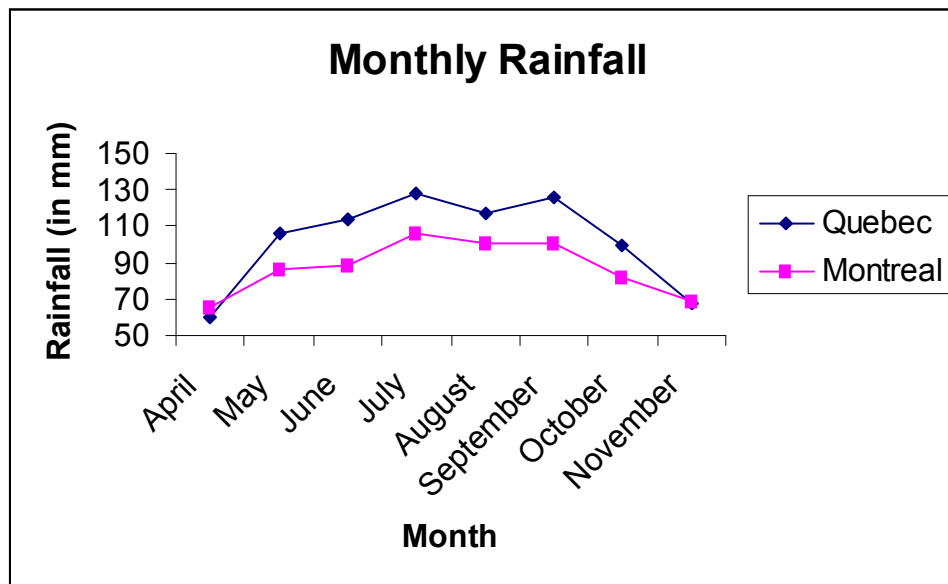
Time Series

In order to compare the average rainfall during the summer months for two cities, Quebec and Montreal, a meteorologist recorded the following data.

| Month | Average rainfall (in mm) | |
|-----------|--------------------------|----------|
| | Quebec | Montreal |
| April | 60 | 65 |
| May | 106 | 86 |
| June | 114 | 88 |
| July | 128 | 106 |
| August | 117 | 101 |
| September | 126 | 101 |
| October | 100 | 82 |
| November | 68 | 69 |

Make the usual heading and labeling of sheet 1. Reproduce the above table in cells A7:C16.

To make the graph, blacken the cells A8:C16 and Insert a LINE GRAPH. Fix your graph so that it looks something like this.



Histograms

An ecologist wishes to know about the size of the trees in a forest. A sample of 50 trees is taken, and each of their height (in decimeters) is measured. Let us organize this data into a frequency distribution table, and then represent it graphically.

Make the usual heading and labeling to sheet 2.

In cell A6 write “Height in dm” in bold (where dm stands for decimeters, 1dm = 10 cm). Adjust the width of the columns so that the words appear in full. Open the worksheet “Data for Organizing Data with Excel” from my web site and copy the sample results from the worksheet “Tree Height” in cells A7:A56.

Frequency Distribution Table

Let us organize the data into a frequency distribution table. To start, we need to see what our classes will be. To do this, let us first find the maximum and the minimum value in our statistical series. In cell C7 write “Min =” and in cell C8 write “Max =”. Go to cell D7 and click on the f_x icon and go to STATISTICAL – MIN – OK. In NUMBER 1, write the range of your data, that is A7:A56 (or, equivalently, darken the cells A7:A56 with your mouse). Repeat the same thing for the max in D8. In cell C9 write “Range =” and in cell D9 calculate the range, using a formula (Max – Min). In cell C10 write “n =” and in cell D10 calculate the size of the sample using the COUNT function. In cell C11 write “Width =” and in cell D10 divide the range by the square root of the sample size using SQRT() for square root). The answer obtained, 4.51, means the width of our interval should be around this. Thus, rounding this number off, we obtain a class width of 5 (or 4.9). Since the smallest number is 19.7, we can make our first class start at 15, giving us the class 15.0 to 19.9. The next class is then 20.0 to 24.9 and so on.

Using the classes found above let us make a relative frequency distribution table in cells C14:H24 that looks like this (ignore the gray lines, they simply outline the cells in Excel):

| Distribution of the height for a sample of 50 trees | | | | | |
|---|----|------|-----------|--------------------|----------------|
| Height (dm) | | | Frequency | Relative Frequency | Class Midpoint |
| 15.0 | to | 19.9 | | | |
| 20.0 | to | 24.9 | | | |
| 25.0 | to | 29.9 | | | |
| 30.0 | to | 34.9 | | | |
| 35.0 | to | 39.9 | | | |
| 40.0 | to | 44.9 | | | |
| 45.0 | to | 49.9 | | | |
| 50.0 | to | 54.9 | | | |
| <i>Total</i> | | | | | |

To compile the data go to cell F16 and use the function FREQUENCY in f_x . For DATA_ARRAY, we give the block of cells representing the height of our trees, so the cells A7:A56, and for BINS_ARRAY the upper class limits, that is, the cells E16:E23. Click OK. Next, to get the other numbers, we darken the block of cells for the frequency, that is, F16:F23 and then press the key F2 followed by CTRL-SHIFT-ENTER. The frequencies should appear. The rest of the numbers in the table can be obtained through formulas.

Your results should look like this:

| | A | B | C | D | E | F | G | H | I |
|----|----------------------------|---|--------------|----|---|---|---|---|---|
| 1 | Business Statistics | | | | | | | | |
| 2 | Organizing Data with Excel | | | | | | | | |
| 3 | By Martin Huard | | | | | | | | |
| 4 | 18-Aug-08 | | | | | | | | |
| 5 | | | | | | | | | |
| 6 | Height in dm | | | | | | | | |
| 7 | 32.9 | | Min = 19.7 | dm | | | | | |
| 8 | 26.1 | | Max = 51.6 | dm | | | | | |
| 9 | 36.7 | | Range = 31.9 | dm | | | | | |
| 10 | 43.9 | | n = | 50 | | | | | |
| 11 | 43.4 | | Width = 4.51 | dm | | | | | |
| 12 | 47.1 | | | | | | | | |
| 13 | 19.7 | | | | | | | | |
| 14 | 35.2 | | | | | | | | |
| 15 | 42.7 | | | | | | | | |
| 16 | 27.4 | | | | | | | | |
| 17 | 30.2 | | | | | | | | |
| 18 | 23.2 | | | | | | | | |
| 19 | 22.1 | | | | | | | | |
| 20 | 28.2 | | | | | | | | |
| 21 | 29.6 | | | | | | | | |
| 22 | 20.2 | | | | | | | | |
| 23 | 31.0 | | | | | | | | |
| 24 | 32.2 | | | | | | | | |
| 25 | 35.9 | | | | | | | | |
| 26 | 20.8 | | | | | | | | |

| Height (dm) | Frequency | Relative Frequency | Class Midpoint |
|--------------|-----------|--------------------|----------------|
| 15.0 to 19.9 | 1 | 2% | 17.45 |
| 20.0 to 24.9 | 7 | 14% | 22.45 |
| 25.0 to 29.9 | 9 | 18% | 27.45 |
| 30.0 to 34.9 | 12 | 24% | 32.45 |
| 35.0 to 39.9 | 8 | 16% | 37.45 |
| 40.0 to 44.9 | 7 | 14% | 42.45 |
| 45.0 to 49.9 | 4 | 8% | 47.45 |
| 50.0 to 54.9 | 2 | 4% | 52.45 |
| <i>Total</i> | 50 | 100% | |

Graphing

Let us make a histogram with the above data. Insert a COLUMN graph with the relative frequency the same way you did previously, adding appropriate titles, and with no legend nor gridlines. For the HORIZONTAL (CATEGORY) AXIS LABELS in SELECT DATA, use the Midpoints. Choose the appropriate CHART LAYOUT and STYLE so that your graph looks something like this.

