# MATH 124 – Creating Histograms
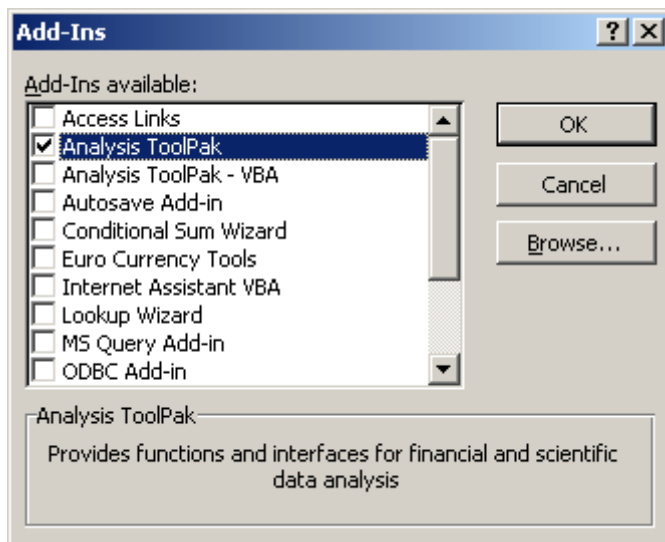## bolstad_math124@bmbolstad.com
## http://math124sfsu.bmbolstad.com

The purpose of this document is to guide you through the steps needed to create histograms using excel. For practice we will use the cereals data for this document (this is the same data used in the loading data document). Excel by itself does not produce histograms, but there is an add-in that does help you along the way. This document will guide you along the steps to installing it if you need it (on campus computers should already have this installed).

**Making sure that you have the Analysis ToolPak installed and how to install it if you don't have it already**

The Analysis Toolpak provides a number of statistical routines/methods that are not by default part of excel. You can check that it is installed by going to the *Tools* menu. If there is a *Data Analysis* option in this menu then it is already installed and you do not need to install it again.
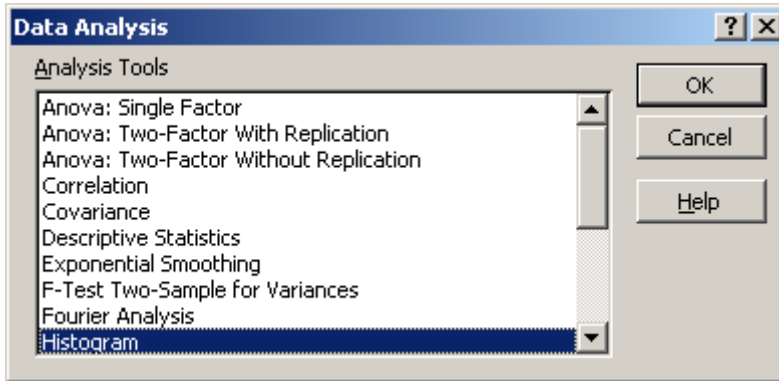
1. To install the Analysis Toolpak go to the *Tools* menu and select add-ins. Then check the box next to *Analysis Toolpak* and click ok.
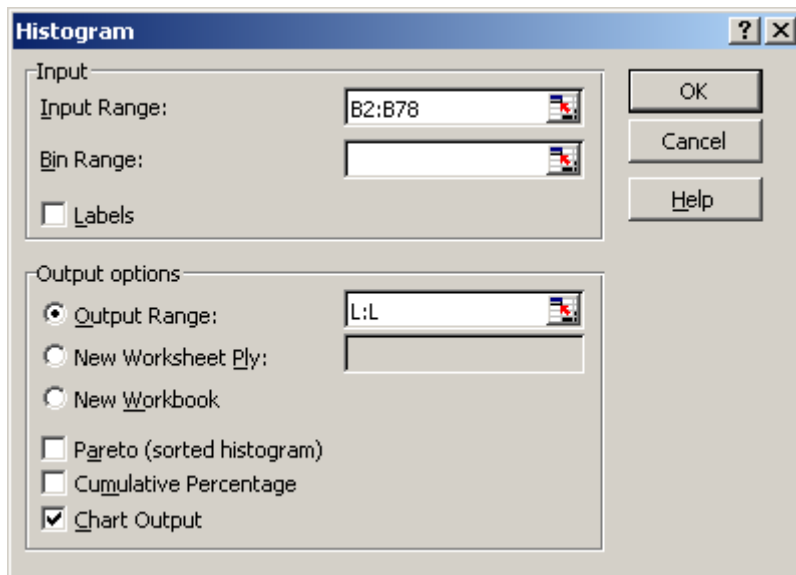


2. Depending on how Excel is installed on your computer you might need to provide your Excel or Office CD for it to install. The installer will load the requisite software from the CD and install it into Excel.
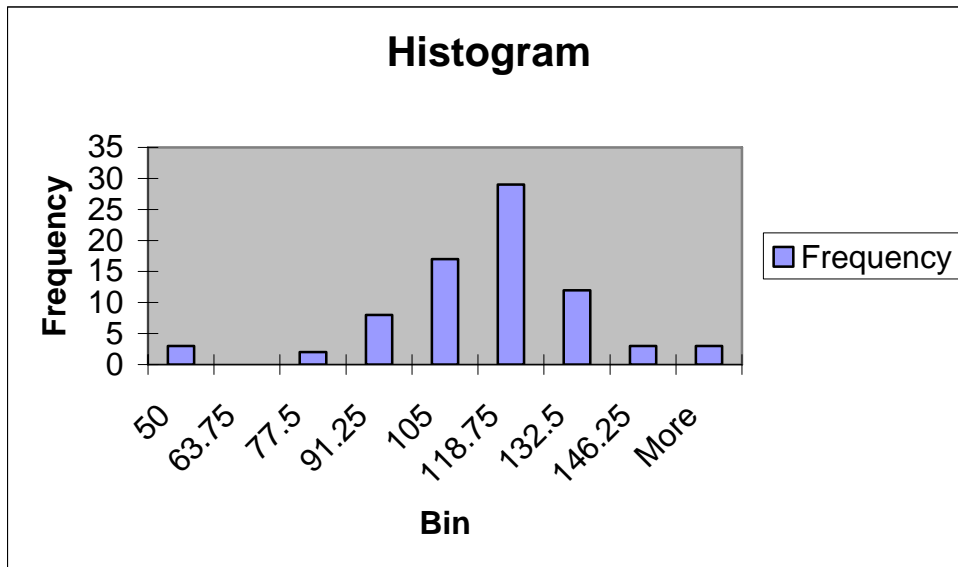
**Drawing Histograms**

1. We will draw a histogram of calories. First go to the *Tools* menu and select the *Data Analysis* option. This should bring up a dialog box. Choose the *Histogram* as your analysis tool and then click on OK.
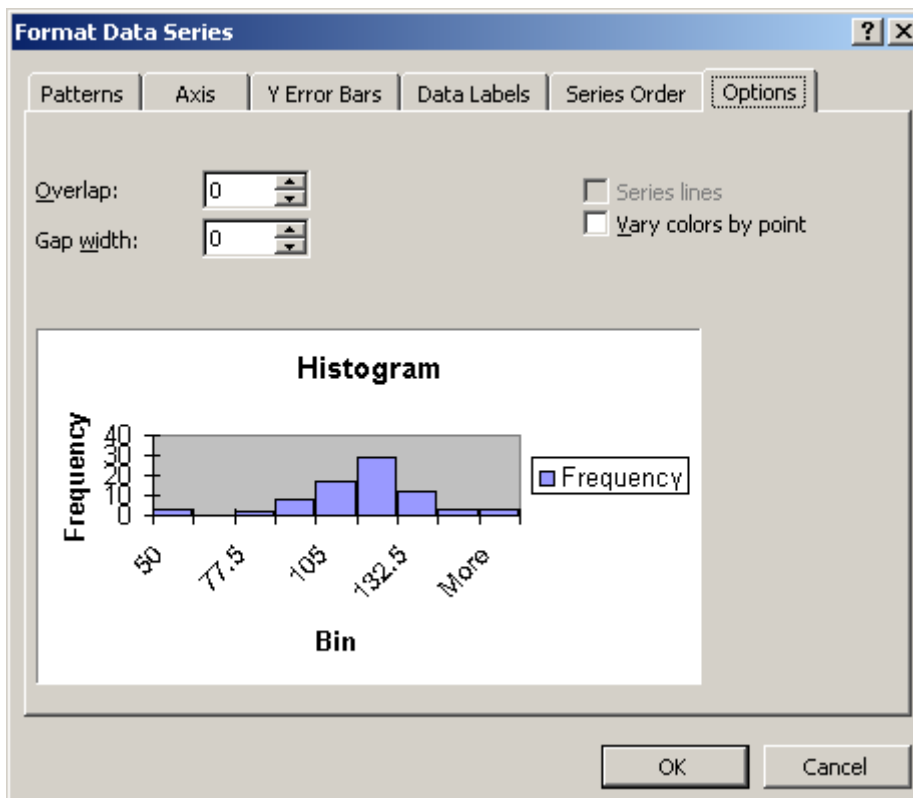


2. The Histogram dialog box will appear. To specify the calories data type B2:B78 into the *Input Range* field (right now leave the Bin Range empty). Go to the Output options and set the *Output range* to some empty part of the spreadsheet (in this example I use L:L to refer to column L. Also make sure that you check the box next to *Chart Output*. Then click on OK.

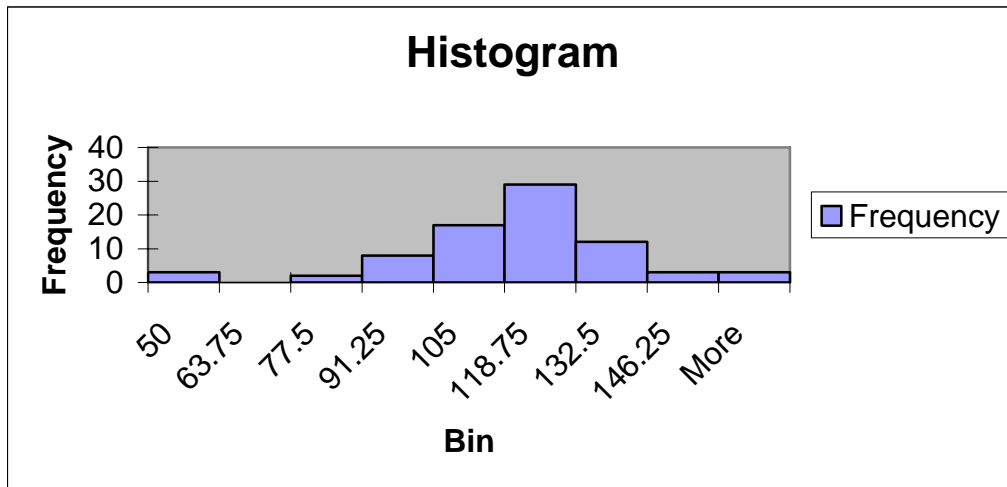3. You should now have a histogram that looks something like the following.



4. We need to do a few things to make it look nicer. The first thing is to make all the bars join together. To do this double click on any bar on the plot to bring up the Format Data Series dialog box. Go to the *Options* tab and set the Gap Width to 0. Then click on ok.
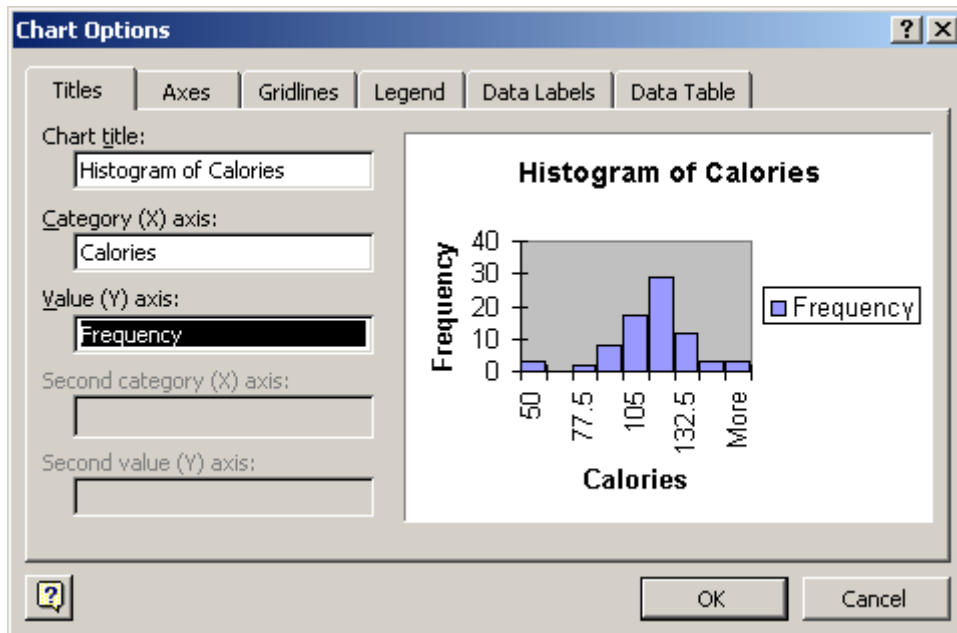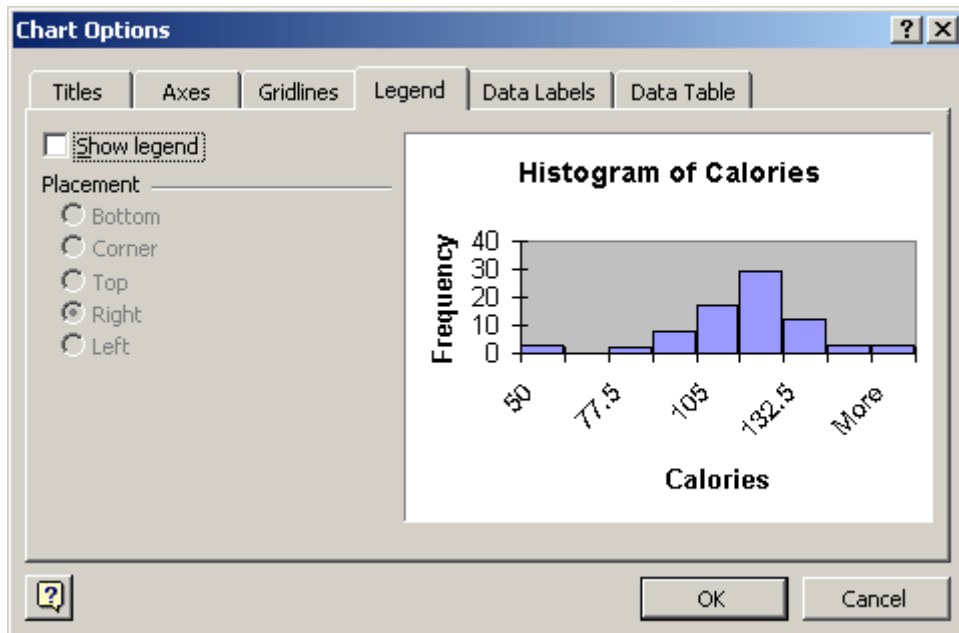


5.

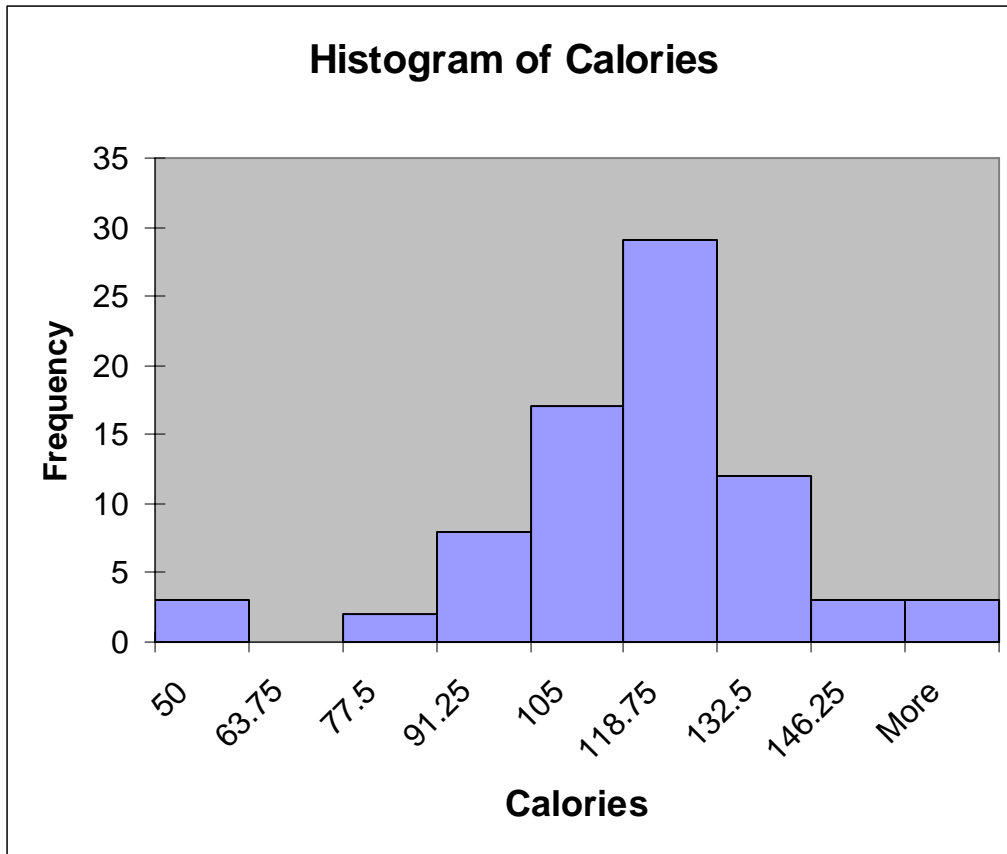6. Your histogram should now look something like the following:



7. To make it look a little nicer right click on the graph and select Chart Options. Change the titles using the options in the Titles tab.

8. It will probably also look a little nicer if you turn off the legend. Do this from the legend tab. Once you are done click ok.
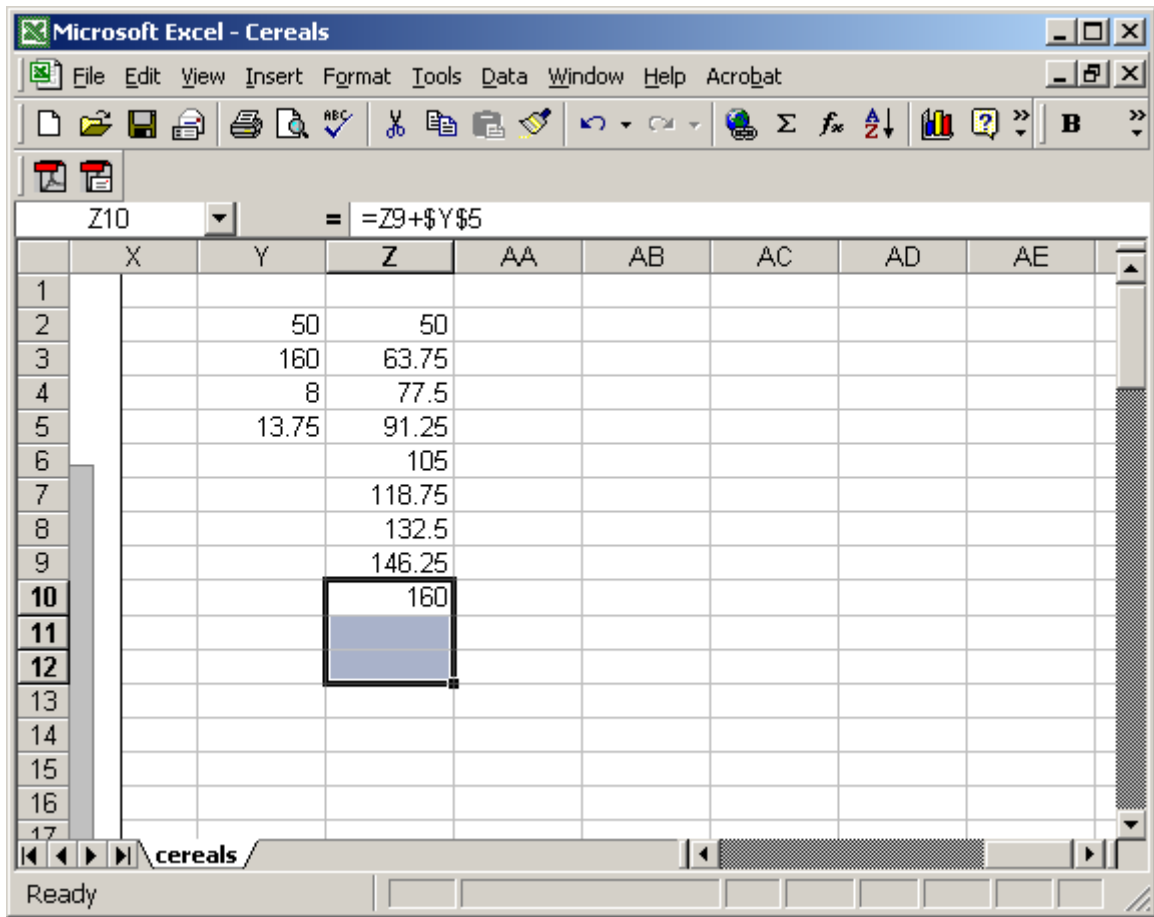


9. You've completed your histogram. It should look something like the following:
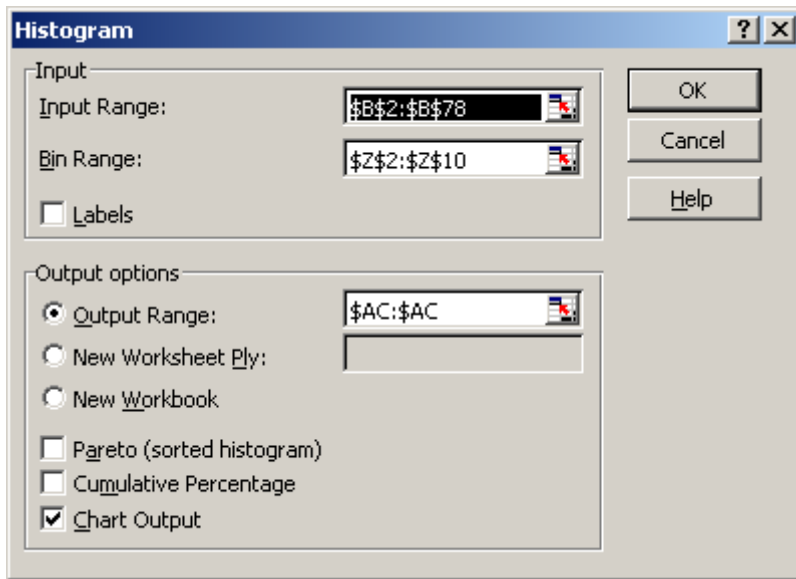
# Histogram of Calories
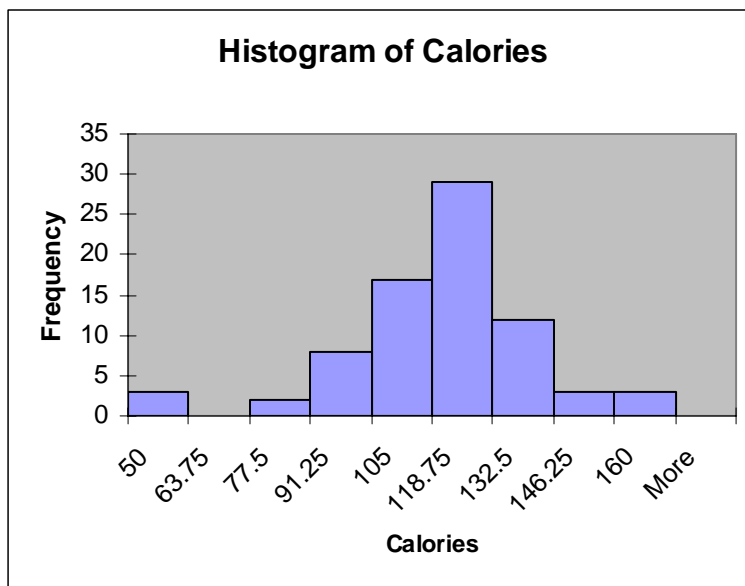


**Controlling the number of bins in the histogram**

1. You might have noticed that there is a "More" bin on your histogram. This is not really statistically a good idea, because we don't know if the "More" observations are just a little bit or a lot larger than the rest of the observations.

2. The first thing you need to do is set up the bin ranges of the data. Go to an empty part of the spreadsheet (in this example we will use cells in the Y column). In the cell Y2 type "=min(B2:B78)". In the cell Y3 type "=max(B2:B78)". In the cell Y4 type a number to represent how many bins you intend to have. (we will type in 8 for this example). Finally go to the cell Y5 and type =(Y3-Y2)/Y4". This number will be the bin width.

3. Next move across to the adjacent column and in the cell Z2 type "=Y2". Then move down to the cell Z3 and type "=Z2+$Y$5".

4. Select cells Z3 through Z10 and select Edit Menu > Fill > Fill Down (or press Ctrl+D). At this point your spreadsheet should look something like this:
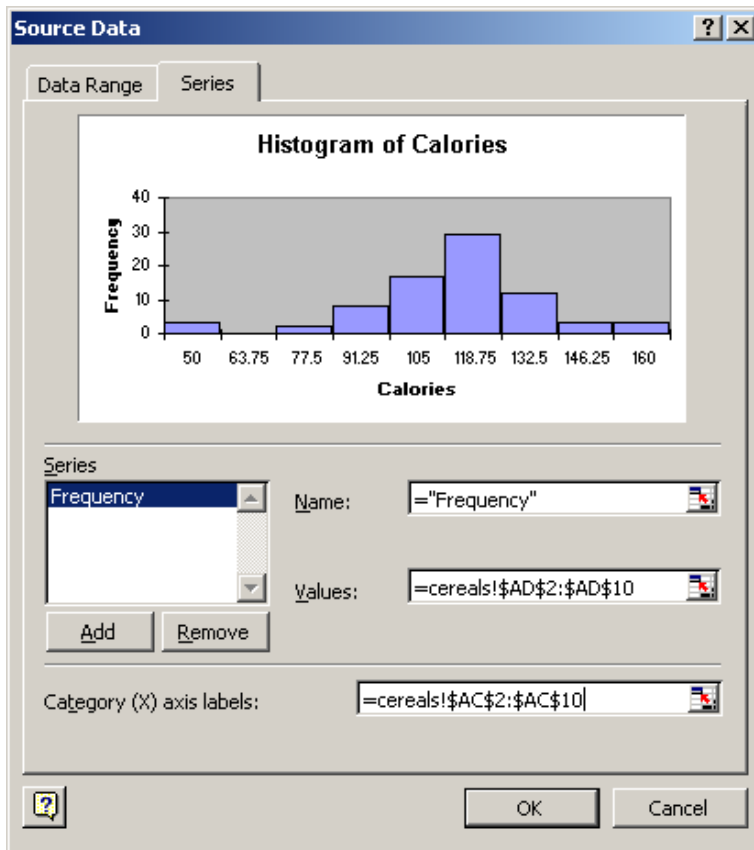
5. Now it is time to draw the histogram. Go to the Tools menu and select Data Analysis and then histogram as before. The main change is now you should enter something into the *Bin Range* field. In this case we will set it to the range we just set up (ie Z2:Z10). Make sure you set an Output range to an empty part of the spreadsheet and select chart output. Click OK to get the chart to appear.

6. Next you should reformat your chart in pretty much the same way you did above (ie make the bars have 0 gap width, change the titles and get rid of the legend). Do this using the same instructions above. Once your done your plot should look something like this



7. You'll notice that the "More" column is still there. To get rid of it right click on the chart and select the *Source Data* option. Go to the series tab and alter the ranges in the Values and Labels fields (this should just be a matter of making the cell one cell earlier). In this case that means you have range AD2:AD10. Click OK to finish.

8. Your plot should now be done and look something like the following.