

Making plots using StatTools

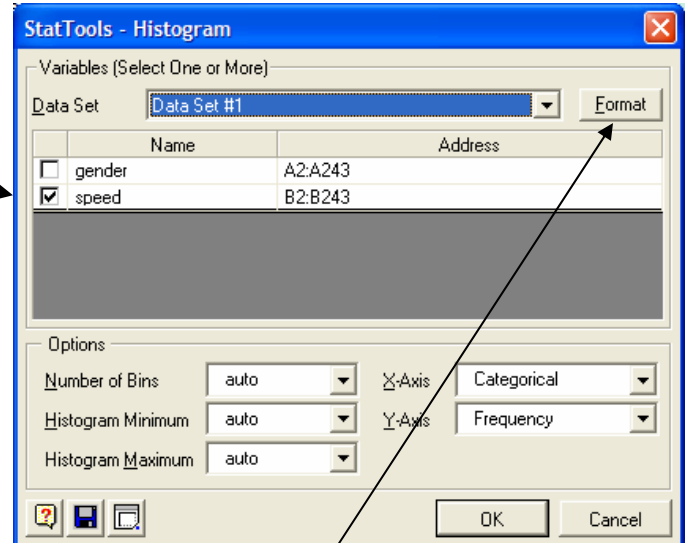
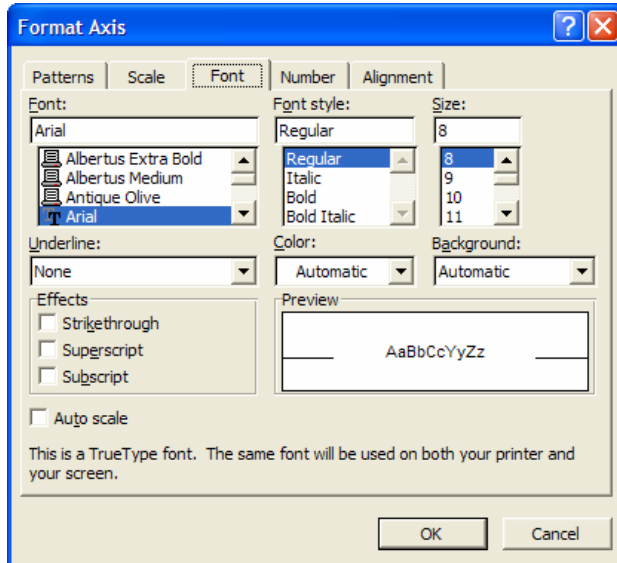
Plots for numerical variables: The variables in my example are: *Gender* and *Speed*.

First open up your data set in StatTools. To make a histogram or Box and whisker plot, go through the following pull down menus: **StatTools** → **Summary Graphs** → your choice of graph

For example, if you want a Histogram, click on **Histogram**. You will get a window like the one below.

I want a histogram of the variable *speed*. I need to check the box **speed** and then click **OK**.

Once you have your graph, you may decide that you'd like to change the title. To do this, click on the title and type in the new one. If you'd like to change the font size of anything, double click on what you'd like to change. You'll get a window like the one below. Choose your size and click **OK**.



Format button

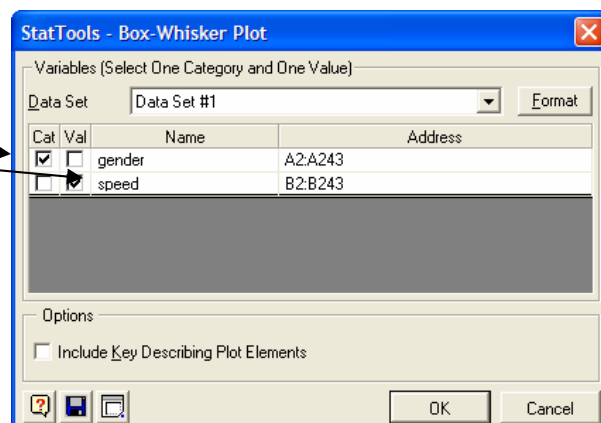
To make a **Q-Q Normal plot**, use **StatTools** → **Normality Tests** → **Q-Q Normal Plot...**

Plots of association when the explanatory is cat and the response is numerical:

The only difference from the above directions is that in addition to the above, you'll have to click the **Format** button and choose **stacked**. When you do this, you'll get a window similar to the one below. Check the **Cat** box next to the explanatory variable and check on the **Val** box next to the response variable. For example, if I wanted to make stacked box plots for speed based on gender, I should have a window like the one below.

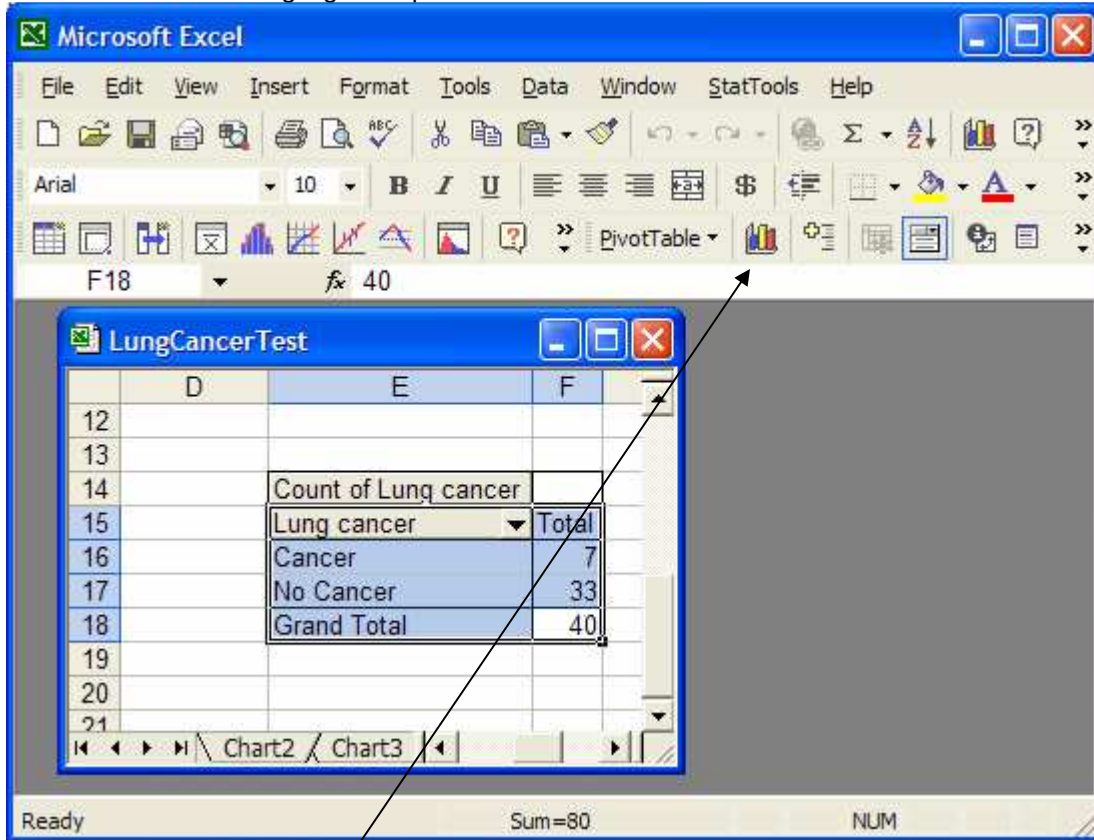
Now click on **OK**, and you are done.

Cat box
Val box

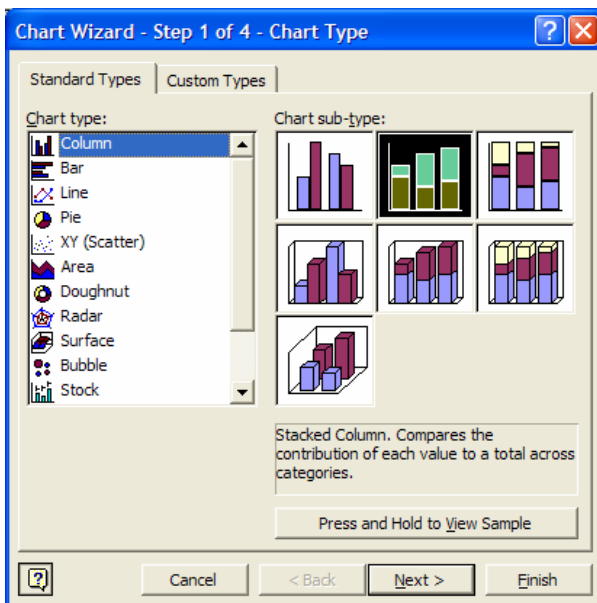


Plots for categorical variables:

Single categorical variable: Before starting, you must make a [pivot chart](#) with just that variable. For my example, I'm going to use the Smoking and Lung Cancer data set. I will use *Lung Cancer* as my single variable. First I must highlight the pivot table as follows:

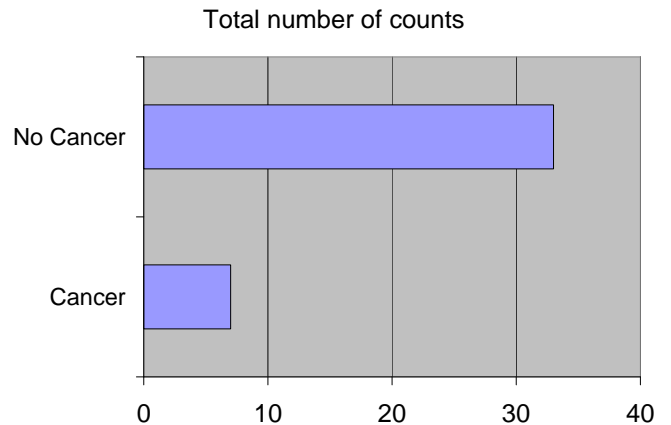
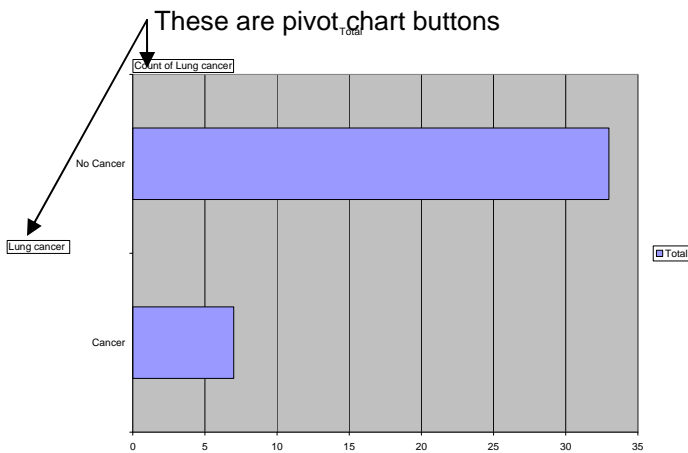


Click once on the charts icon. Now you will have a bar chart. If this is the plot you want, then go to the next page to see how you can make it pretty and readable. If you want another type of chart, click on the icon again. When you do so, you'll get the window below. Choose the type of plot you want, click **Next**, type in the **Chart title** of your choice and click **Finish**.



This may give you a nice chart, but most likely, you'll have to change the font. Please see the next page for directions on how to make your chart pretty.

To change your chart from something unreadable like the chart below left to a readable chart like below right, you need to double click on the numbers and text and change the font size. There are precise directions on pages 6 and 7. Usually a font of 24 works well. To get rid of those annoying pivot chart field buttons, right click on a pivot chart button on the plots and choose **Hide PivotChart Field Button**.



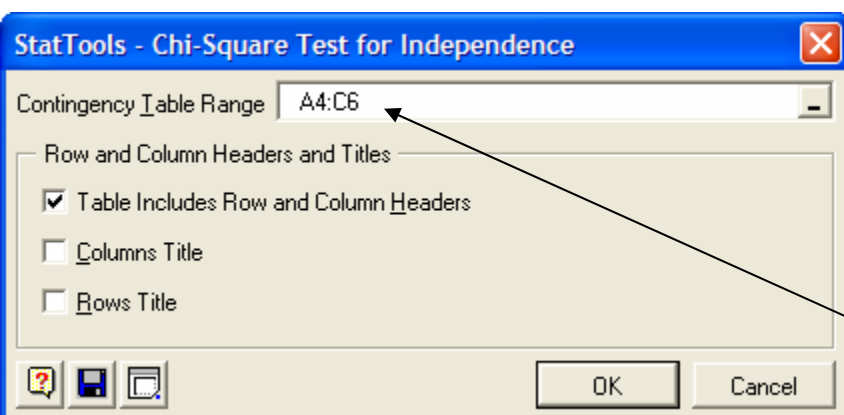
Two categorical variables:

There are various different possible charts and graphs. Most of them require that you do a Chi-Square Statistical test first, which requires a pivot table. Below is my pivot table.

Count of Gender	Diabetes		
Gender	diabetes	no diabetes	Grand Total
female	27	23	50
male	12	38	50
Grand Total	39	61	100

Once you have the pivot table, go through the sequence of drop-down menus

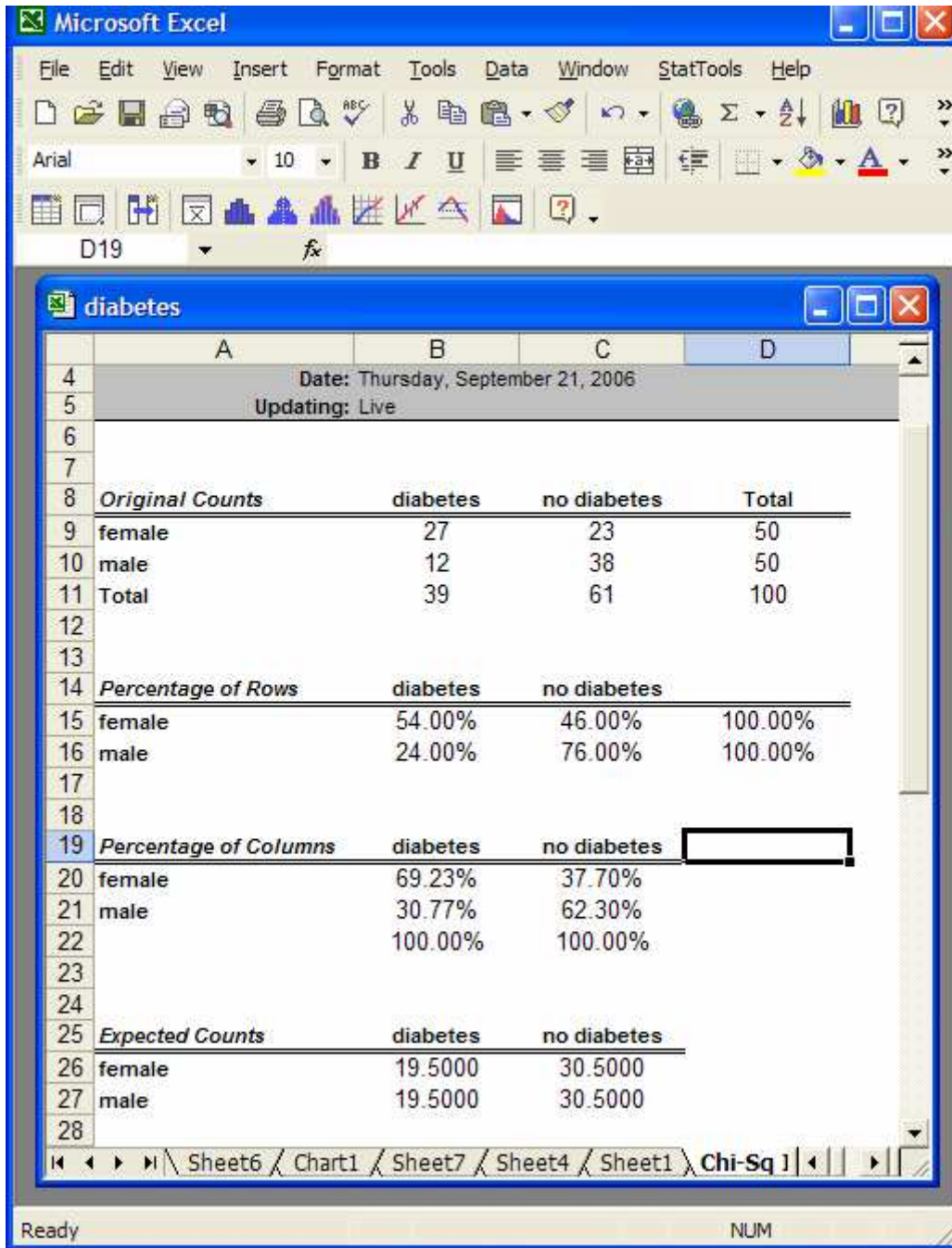
StatTools → Statistical Inference → Chi-square Independence Test ...



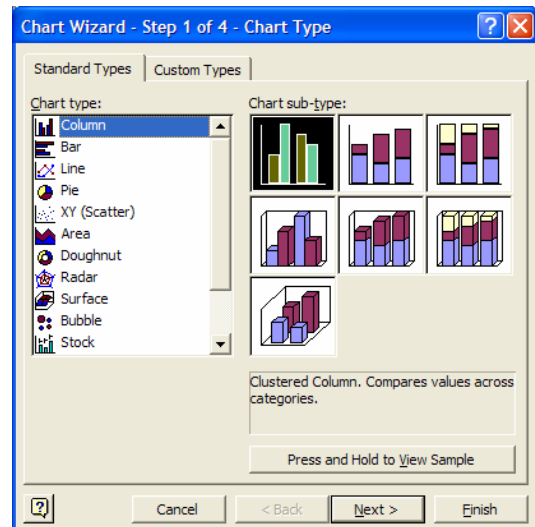
You will get a window like the one to the left. It may or may not have the Contingency Table Range already filled in. But you need to make sure that you only include all Row and Column headers and the counts. None of the totals should be included in the range.

Range of contingency table

After putting in the range and clicking **OK**, you will get a table of charts like the one below.



You are now ready to make your plots. To make a side-by-side bar chart, click the Charts icon and highlight the type of chart you want, which in this case is the one already highlighted in black to the right. Click **Next**.

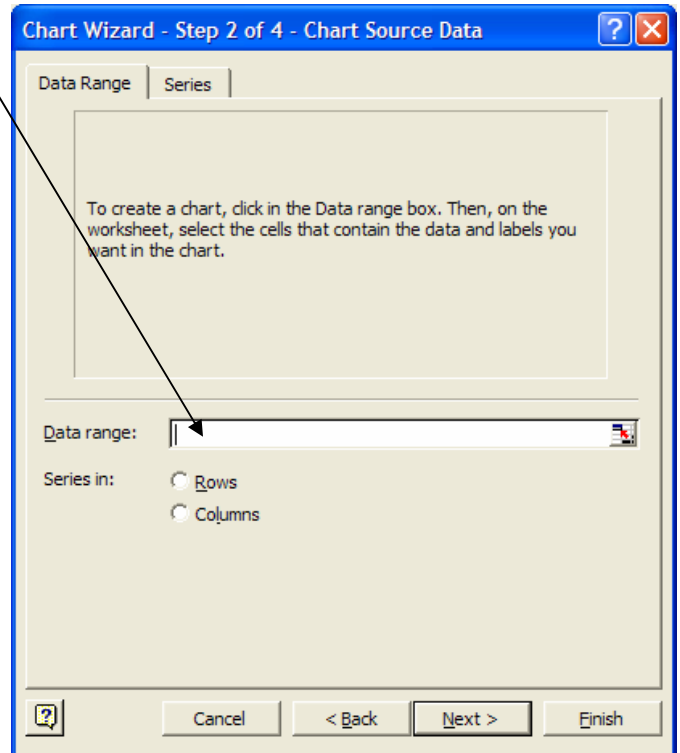


The next window, shown right, asks for the **Data Range**.

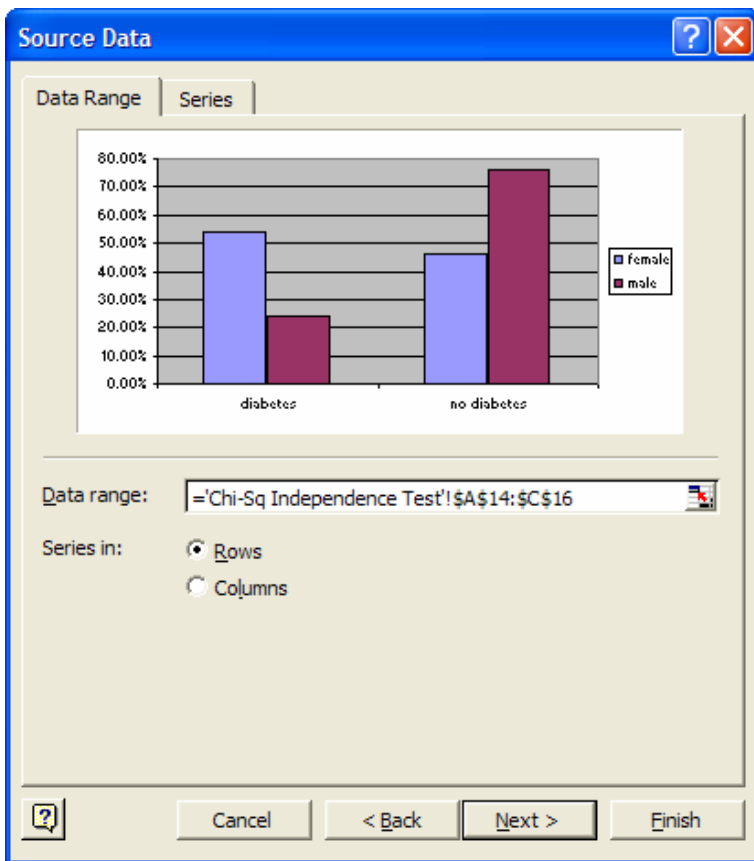
For the side-by-side bar chart, you want the **Percentage of Rows** chart. See below the part of the chart I chose. It is important that you NOT include the Totals. The way to select the part of the graph you want is to left click on a cell (usually a corner cell) and keeping the left button depressed, move the mouse around until the correct part of the chart has been selected.

Percentage of Rows	diabetes	no diabetes
female	54.00%	46.00%
male	24.00%	76.00%

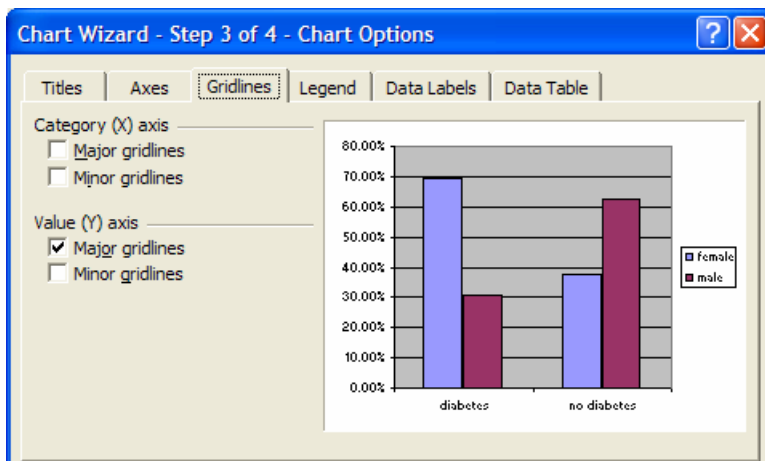
Once you've selected the Data range, you will get the window directly below.



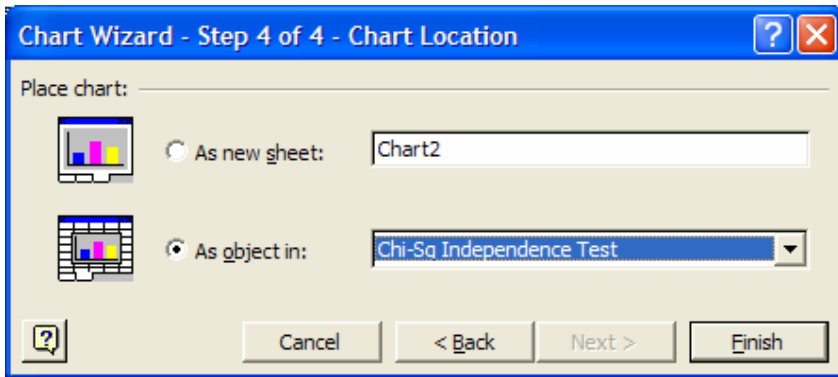
Note that there are these annoying grid lines. You can get rid of these in the next step. To get to the next step, click **Next**.



You will get the window below. To get rid of the grid lines, go to the Gridlines tab and unclick Major grid lines for the Y axis. You can also add a title, data labels, etc now. When done, click **Next**.



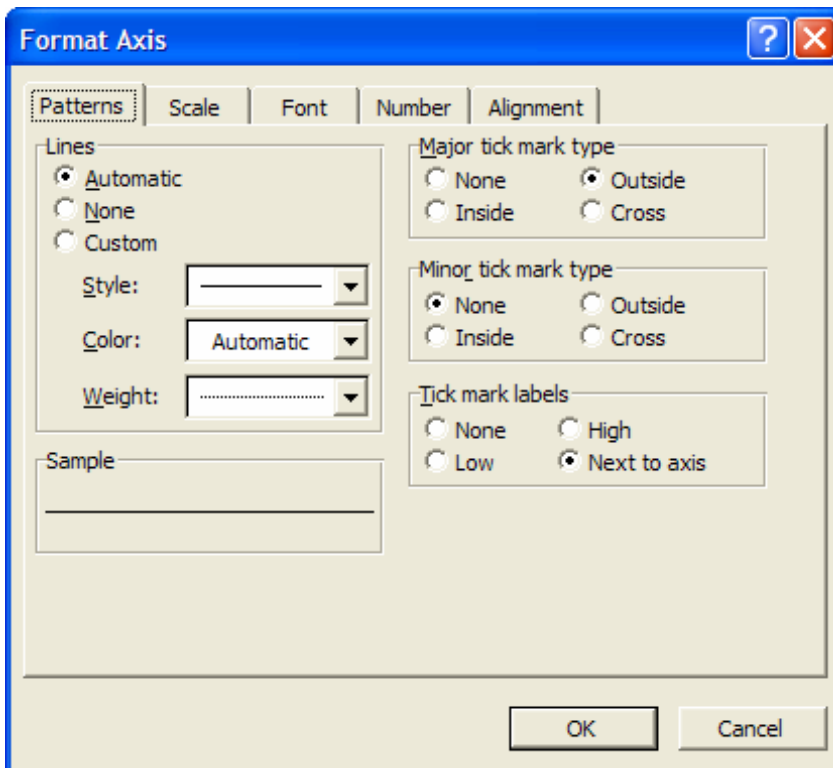
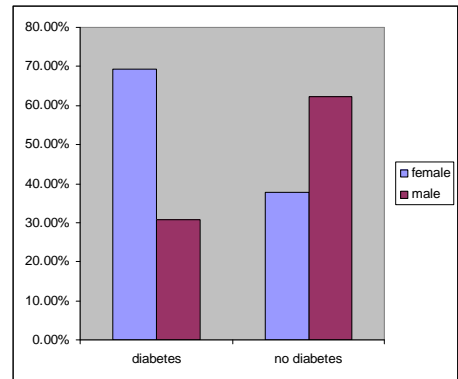
The last window you will get is the one below asking you where you want your chart put. Pick one and click **Finish**.



You'll get a plot like the one to the right. Now, you can change the font size.

To make the font size larger, right click on a word you'd like bigger. You'll get a window asking if you want Format Axis or to Clear. Select Format Axis. You'll get the box below.

Select the Font tab and then the font size you want. You can also put boxes around things, add tick marks, etc. End by clicking **ok**.



Making a scatter plot:

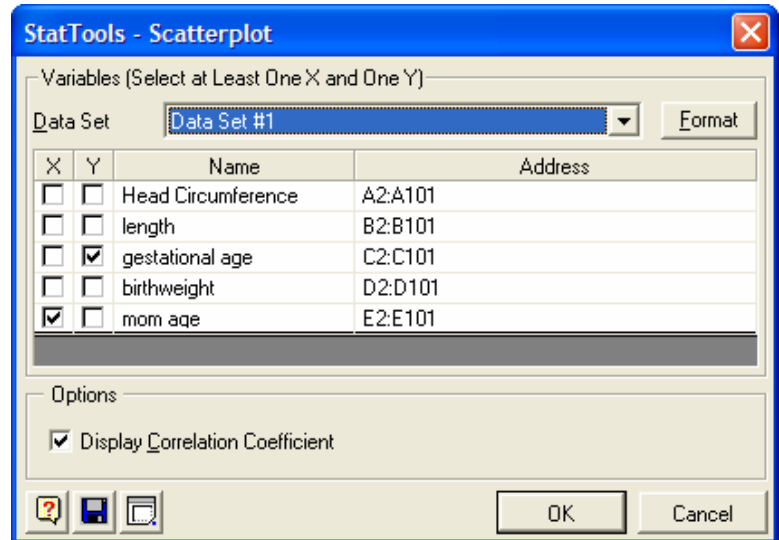
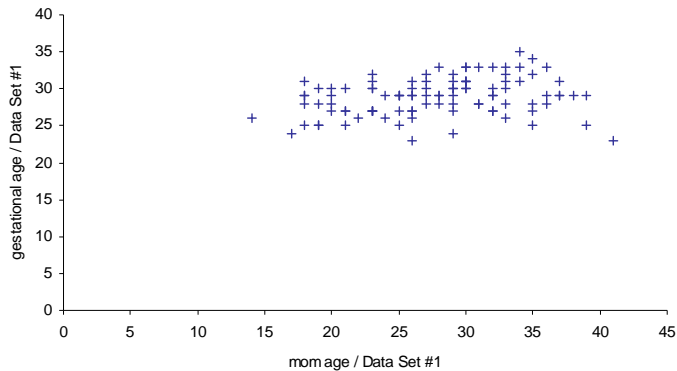
The data set I'll be using is comprised of data taken from premature infants at birth. The variables are: head circumference, length, gestational age (number of weeks mom was pregnant), birth weight and mom's age. The variables of interest are gestational age and mom's age. Mom's age is the explanatory variable (plotted on X axis) and gestational age is the response (plotted on the Y axis).

It is easiest to go through the pull down menu sequence: **StatTool -> Summary Graphs -> Scatterplots.**

Click the variables you wanted to be plotted as shown to the right. Then click **OK.**

You will get something like the below.

Scatterplot of gestational age vs mom age of Data Set #1

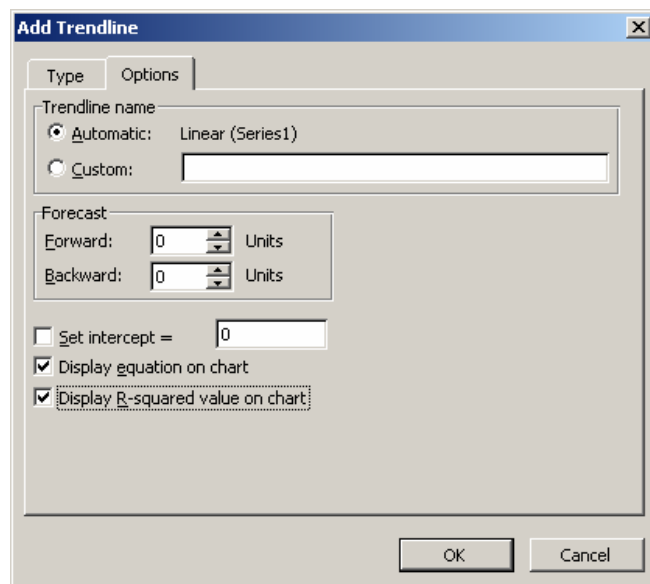
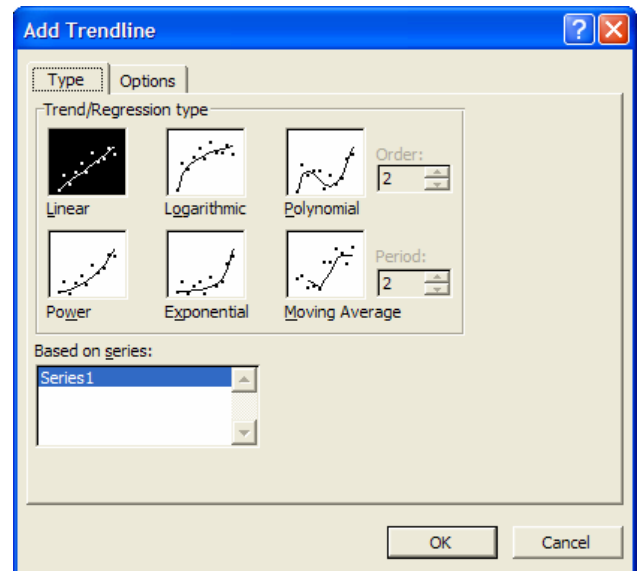


To change the font size of an axis, **right click** on one of the numbers on that axis and choose **Format Axis...** Next click the Font button. Now choose the font you want (12 or 14 works well).

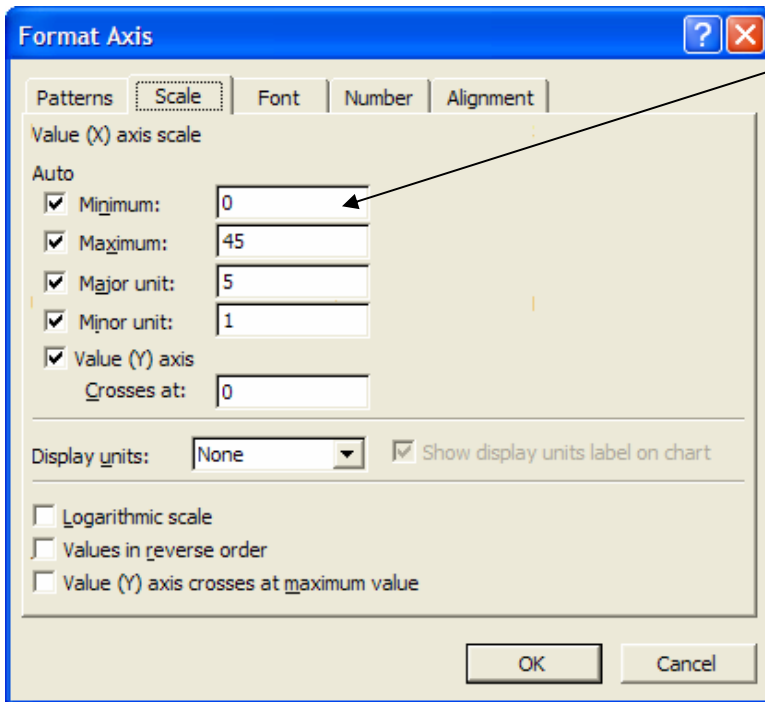
To change the title or label of any axis, left click on it and type in the new title. Now right click on this, choose **Format Axis Title...** and choose an appropriate font size.

To add a regression line, right click on any data point and choose **Add Trendline...** You will get a pop up like the one to the right. Click on **Linear** box and then **OK.**

To get R^2 or the estimated equation, click on the options tab of the **Add Trendline** dialogue box, and select "Display equation on chart" and "Display R-squared value on chart."



Finally, you might want to change the range plotted on the X axis. To change the range of the X axis values, right click on the X axis line and choose **Format Axis...** You will get a pop up like the one below. Click on the Scale button.



Now, type in the minimum and maximum values you want for the X axis in the appropriate boxes.

An improved chart would look like that shown below.

