# MINITAB 15 MANUAL FOR MBA 6120

Created by: Liz Ames, Chris Swanton & Norm Chervany
Carlson School of Management
Full-time MBA Program

Introduction

Welcome to the exciting world of Minitab! This introduction will provide a brief description of how the manual is organized. An outline is provided so that you can easily find what function you are looking for. The manual will assist you with any Minitab use needed to MBA 6120.

The manual describes functions in Minitab by walking you through the commands and providing screen shots. Also, using specific data, an example is provided for each function. Please see the Appendix for a description of each data set. The **commands** are easy to recognize as they are in **regular bold font** and always start off by clicking one of the menu bar options. Here is what the menu bar (located at the top of the Minitab screen) looks like:
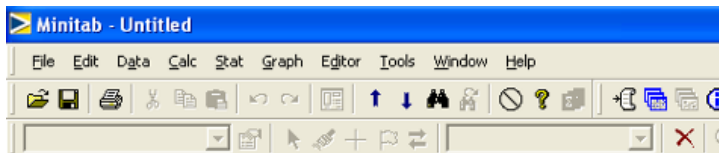

Throughout the manual, after the commands are given, a screenshot of the window you will see appear is shown. Then after all the commands and screenshots are provided, the Minitab output for the example is shown. Also, with many of the functions, you are given different options of what data you want displayed or how you want the data displayed. For example, as you will see with many of the screenshots, buttons like these appear,


and you can, if desired, click on them to edit the function. Throughout the manual, these **buttons** will be designated with ***italicized bold*** letters. However, it is important to note that because these buttons appear on so many different functions, once one of the buttons is explained, it most likely will not be explained again under a different function. Many of the key buttons are shown when the histogram is explained (particularly the buttons shown above), so please reference that section. Another option that is important but is only shown once in the histogram section is the option to edit the graph after it is displayed by double-clicking on different parts of the output graph.

HELP. If you need further assistance or would like more explanation or examples, Minitab has a help function. If you click on the menu bar Help, you have these options:
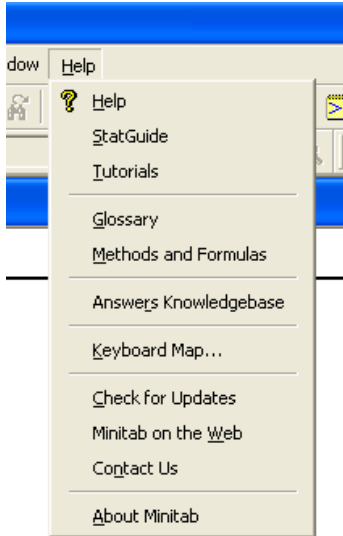
Also, as you will see with the screen shots, the pop-up windows show a help button in the lower, left-hand side.
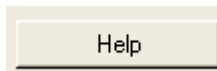
If you click one that *Help* button, you are given the option to search, but are also given explanations and examples of the function that is currently in use. Here are some of the options you can click on to get more information on your function.

OUTLINE

---

[i] Describes **Preview** button.
[ii] Describes **Description** button.
[iii] Describes **Condition** button.
[iv] Describes **By Columns** button.
[v] Describes **Statistics** and **Graphs** buttons.
[vi] Describes **Statistics** and **Options** buttons.
[vii] Describes **Scale, Labels, Data View, Multiple Graphs,** and **Data Options** buttons. This section also shows you how to edit the graph after it is displayed.
[viii] Describes the **Bar Chart Options** button.
[ix] Describes the **Pie Chart Options** button.
[x] Describes the **Graphs** and **Options** buttons.
[xi] Describes the **Graphs** and **Options** buttons.
[xii] Describes the **Options** button.
[xiii] Describes the **Graphs, Options,** and **Results** buttons.
[xiv] Describes the **Options** button.
[xv] Describes the **Data View** button.

I. Basic Functions

a) Entering/Transferring Data

To start a project in Minitab, data can be entered into the worksheet directly or transferred from an Excel file using copy and paste. Many actions in the worksheet are similar to Excel (e.g. Inserting/deleting rows/columns, copy and paste, etc.), however, Minitab worksheets do not accept formulas. You can also transfer data directly from Excel by going to **File → Open Worksheet**
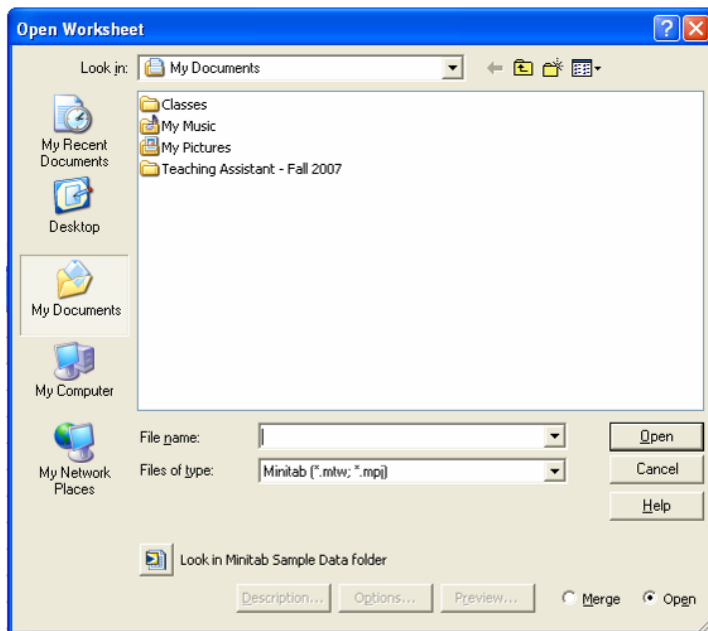
Before clicking **Open**, you will need to specify: (1) where the file is located using the 'Look in:' drop down menu and (2) the type of file using the 'Files of type:' drop down window.
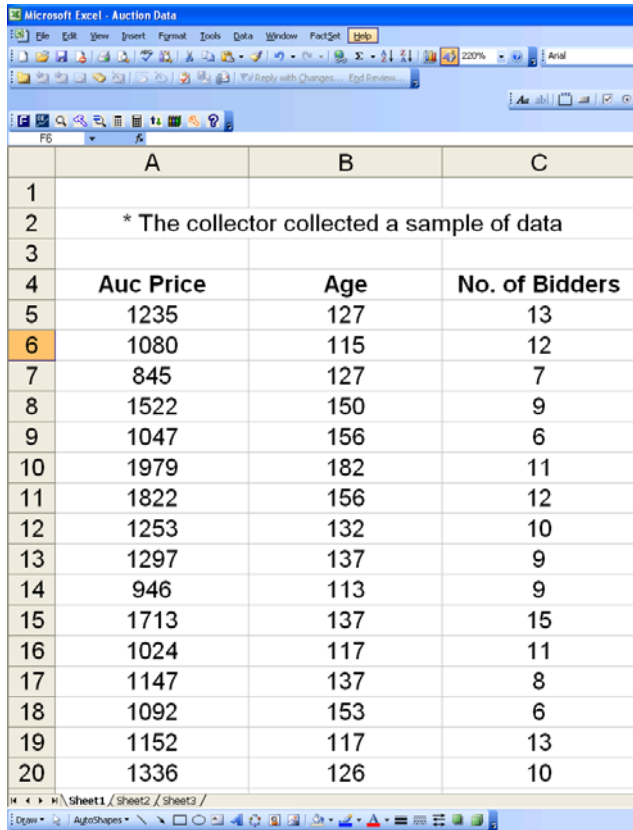
In this manual, the Dell Data, Auction Data, Student Data, and Student (2) Data will be used as examples. The original problem statements and data for each example are contained in the Appendix. You can copy and paste the data from the Appendix into Minitab if desired.

b) Specifying Data Type

Minitab handles numeric data, text data, and date/time data. When transferring data into Minitab, it is assumed that the first row of the file contains the column names. When data is typed or transferred, Minitab decides which type of column it is, depending on the data. In the Data windows and Columns folder, Minitab labels text columns with T and date/time columns with D; numeric columns are not labeled. Sometimes Minitab sets a column as text when you want it to be numeric, and vice versa. If columns are incorrectly labeled as text, you will need to change the column to numeric in order to perform many of the functions.

i)   Change columns from text to numeric

In this example, we want to import the Auction Data, but let's say the top of the excel spreadsheet looked like this:

	A	B	C
1			
2	* The collector collected a sample of data		
3			
4	Auc Price	Age	No. of Bidders
5	1235	127	13
6	1080	115	12
7	845	127	7
8	1522	150	9
9	1047	156	6
10	1979	182	11
11	1822	156	12
12	1253	132	10
13	1297	137	9
14	946	113	9
15	1713	137	15
16	1024	117	11
17	1147	137	8
18	1092	153	6
19	1152	117	13
20	1336	126	10

To transfer the file to Minitab, you would go to **File → Open Worksheet→** (find your file under 'Look in') **Open.** In Minitab, your screen would look like this:

Because the columns are labeled with a "T", we know that Minitab interpreted this file as text instead of numeric. It did this because there was text in a row besides the first row. In this situation, you need to delete the first two rows and cut and paste row 3 into the heading row. Then highlight the three columns, go to **Data → Change Data Type → Text to Numeric →**
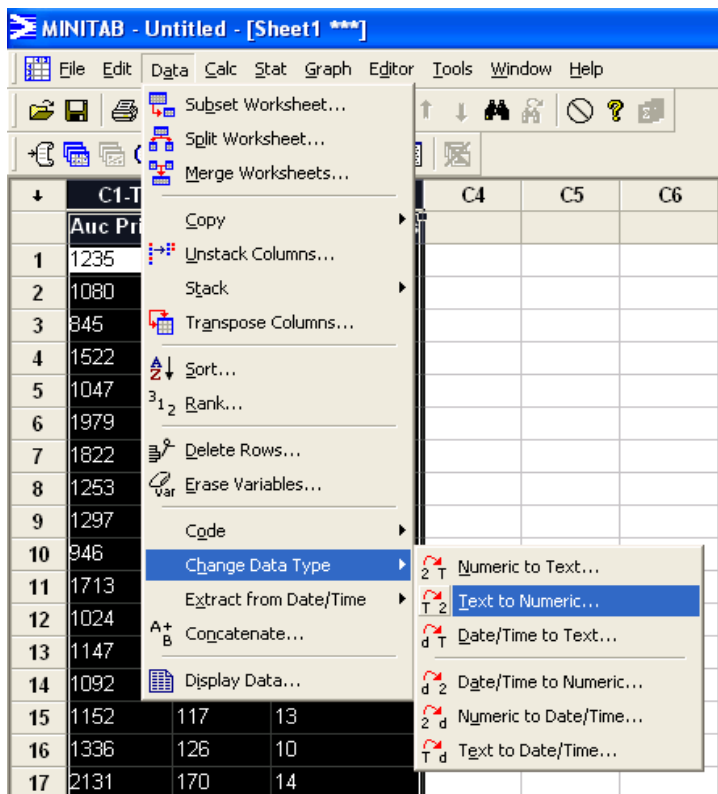
Then this screen will appear. **Double-click on the columns you want to change under Change text columns: → Double-click on the columns where you want to store the changed data under Store numeric columns in: → OK.**

The revised numeric data will now be shown:

ii)   Choose Data Type when Transferring Data

Here is another way to ensure your data is specified as numeric. Go to **File → Open Worksheet→** (find your file under 'Look in') **Click *Preview* →**

Open Worksheet

Look in: MBA 6120

My Recent Documents
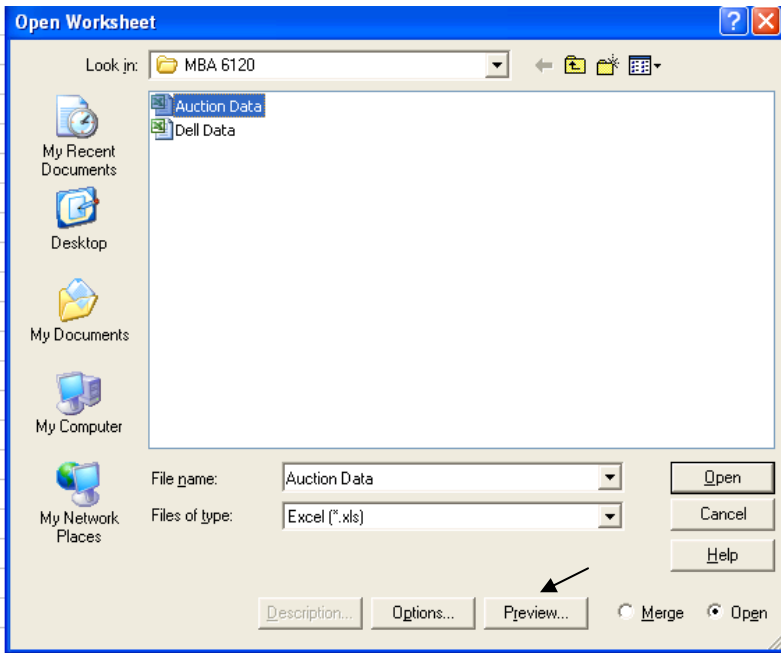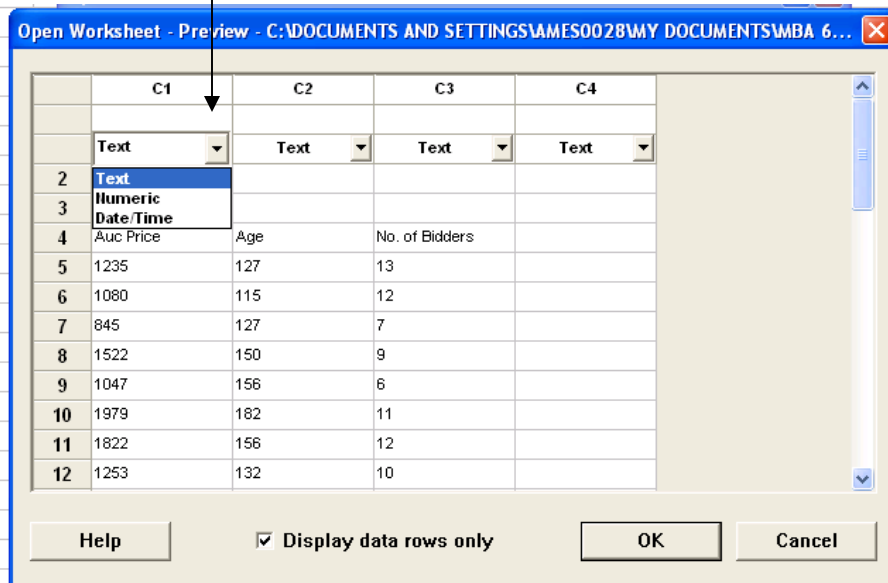
Desktop

My Documents

My Computer

My Network Places

Auction Data
Dell Data

File name: Auction Data   Open

Files of type: Excel (*.xls)   Cancel

Help

Description...   Options...   Preview...   ○ Merge   ● Open

As you can see below, you can choose the data type for each column by clicking on the drop down box.

Open Worksheet - Preview - C:\DOCUMENTS AND SETTINGS\AMES0028\MY DOCUMENTS\MBA 6...

	C1	C2	C3	C4
	Text	Text	Text	Text
2	Text			
3	Numeric			
	Date/Time			
4	Auc Price	Age	No. of Bidders	
5	1235	127	13	
6	1080	115	12	
7	845	127	7	
8	1522	150	9	
9	1047	156	6	
10	1979	182	11	
11	1822	156	12	
12	1253	132	10	

Help   ☑ Display data rows only   OK   Cancel

Make the appropriate changes and click **OK.**

c)   Saving

To save your work in Minitab, you can choose from three options: Saving the worksheet, saving the graph, and saving the project. Saving the project is recommended because it saves everything associated with the project: the worksheet (data), graphs, and the session window (displays history). To do this, go to **File → Save Project As**
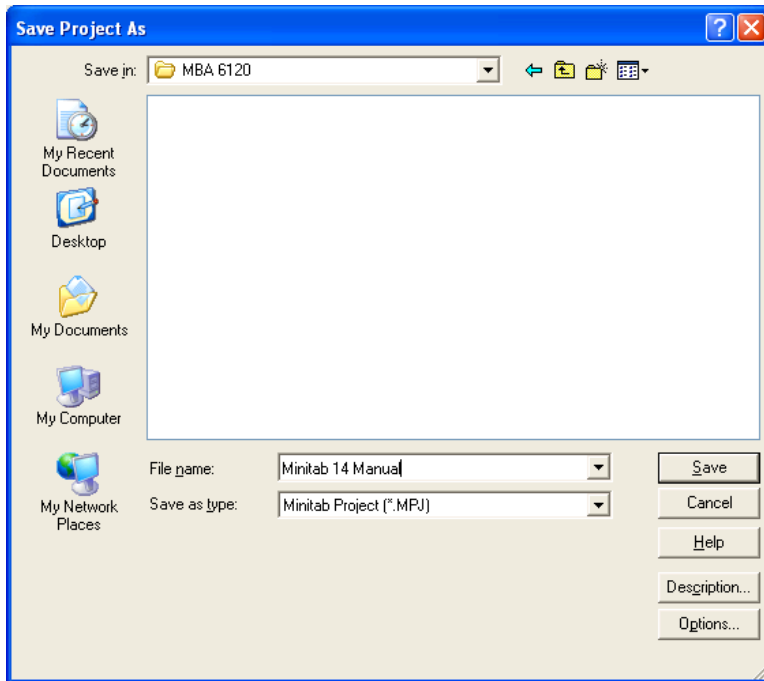
Before clicking **Save**, you will need to specify: (1) where you want to save the project using 'Save in:' drop down menu and (2) the name of the file using the 'File name:' drop down window. Using the *Description* button, you can provide details describing the project being saved.
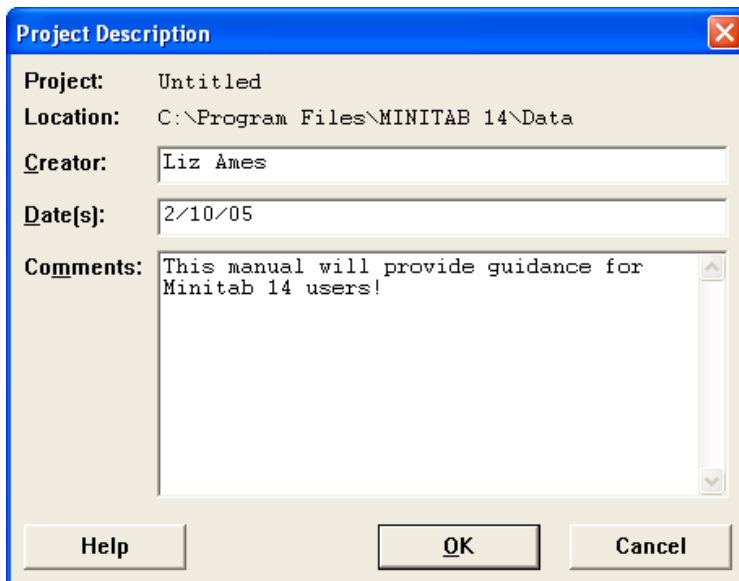
d)   *Printing*

To print, select the worksheet, graph, or session to be printed. Then go to **File → Print Worksheet**, or, **File → Print Graph**, or, **File → Print Session**, depending on what window is being printed.
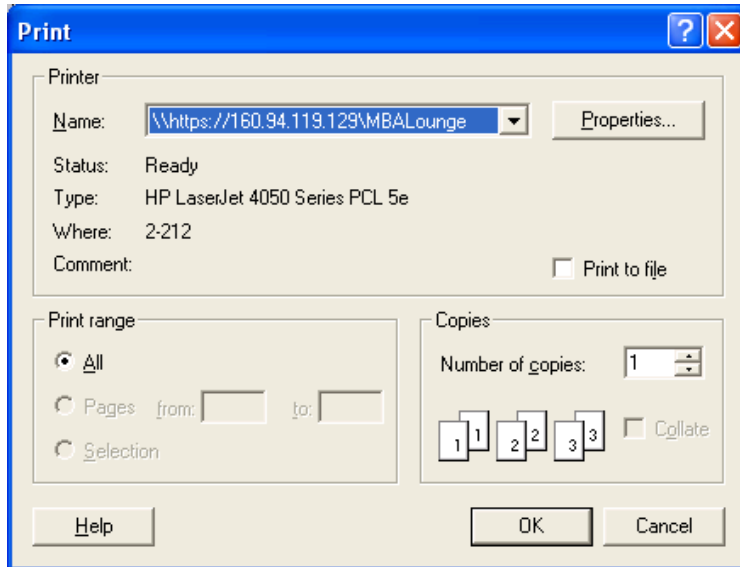
If you are printing a worksheet, you see an additional screen before the one shown above. This screen gives you some options on how you want the worksheet to be printed.
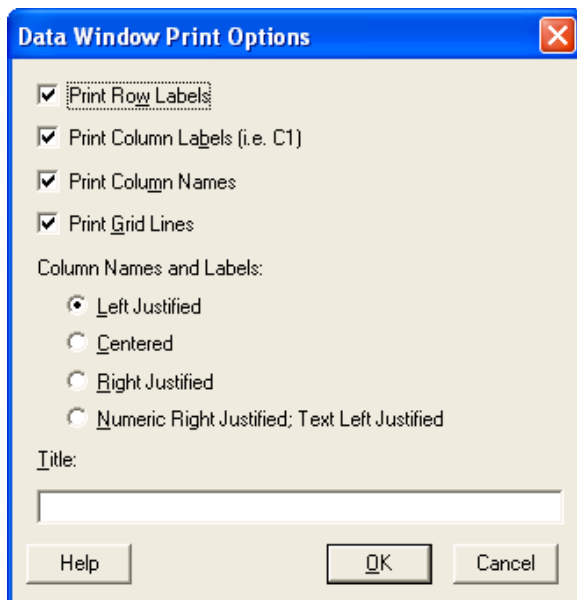
You can also copy and paste what you want to print into other programs, such as Word, and print from there.

II.   Manipulating Data

a)   Sort

You can sort one or more columns of data according to values in the associated column(s) you select. Sorting alphabetizes or numerically orders the data and carries along the associated columns. You can sort in ascending or descending order, and you can specify whether the sorted data should be stored in the original columns, other columns you specify, or in a new worksheet.

Data → Sort → Double-click on appropriate variables for Sort columns and By column → Click on appropriate selection under Store sorted data in → OK.
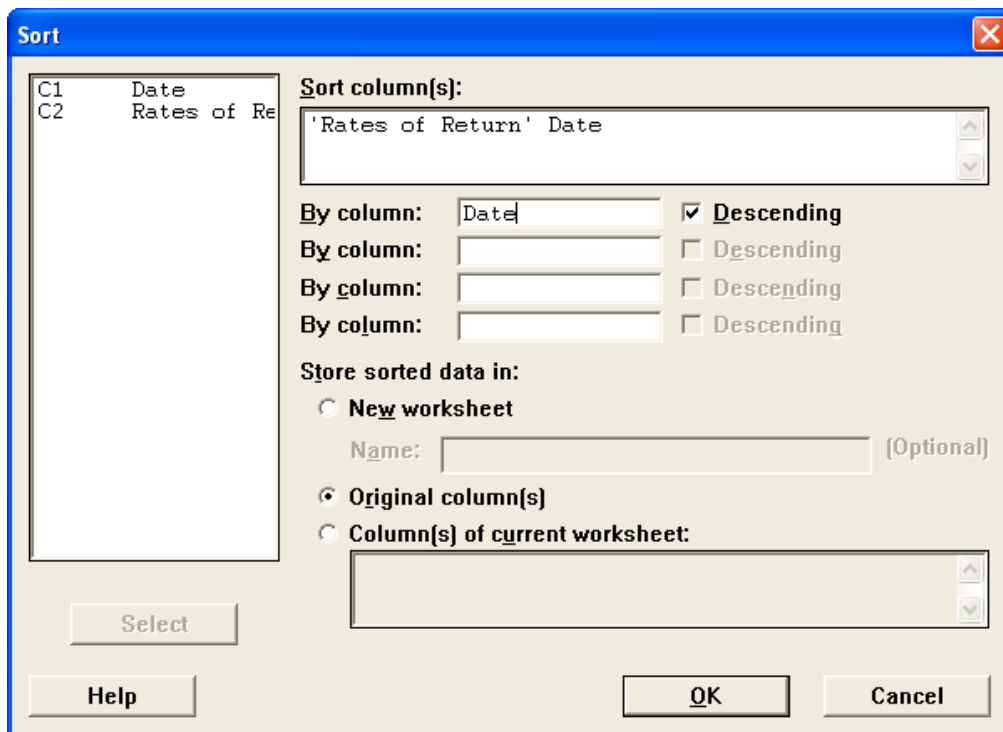
By making the choices shown above, the Dell Data will have the Rates of Return sorted in descending order by the Date in the original columns.

Here is a glimpse of the first 8 rows (there are really 60 rows) of the worksheet:

Date	Rates of Return
31-Dec-98	0.2035
30-Nov-98	-0.0716
30-Oct-98	-0.0038
30-Sep-98	0.315
31-Aug-98	-0.0791
31-Jul-98	0.17
30-Jun-98	0.1263
29-May-98	0.0205

b) *Rank*

You can assign rank scores to values in a column: 1 to the smallest value in the column, 2 to the next smallest, and so on. Ties are assigned the average rank for that value. Missing values are left as missing

Data → Rank → Double-click on appropriate variable in Rank data in (For Dell data, double-click on Rates of Return) **→ Type in name of column heading in Store ranks in** (Type in Rank) **→ OK.**

```
Rank                                              ✕

 C1      Date
 C2      Rates of Re      Rank data in:    'Rates of Retu

                          Store ranks in:  Rank




         Select

         Help                    OK              Cancel
```
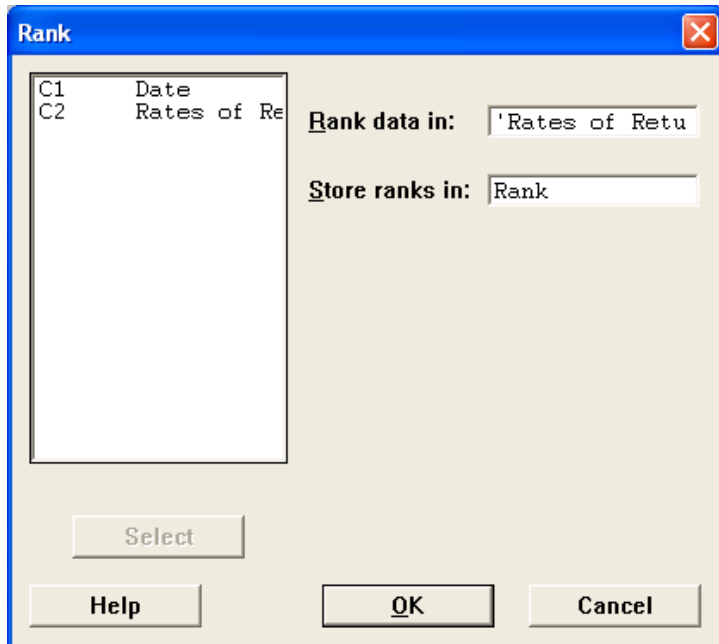
Here is a glimpse of the same first 10 rows as the Sort example:

Date	Rates of Return	Rank
31-Dec-98	0.2035	47
30-Nov-98	-0.0716	9
30-Oct-98	-0.0038	19
30-Sep-98	0.315	56
31-Aug-98	-0.0791	7
31-Jul-98	0.17	40
30-Jun-98	0.1263	35
29-May-98	0.0205	22
30-Apr-98	0.1919	45
31-Mar-98	-0.0313	15

c)   *Delete Rows*

You can delete specified rows from columns in the worksheet and move the remaining rows up.

Data → Delete Rows → Enter which rows you want deleted under Delete rows (For Dell Data, type in 1:3 8 10 to delete rows 1,2,3,8,10) **→ Double-click on appropriate column names under From columns** (Double-click on Date, Rates of Return, and Rank to delete rows from all three columns) **→ OK.**
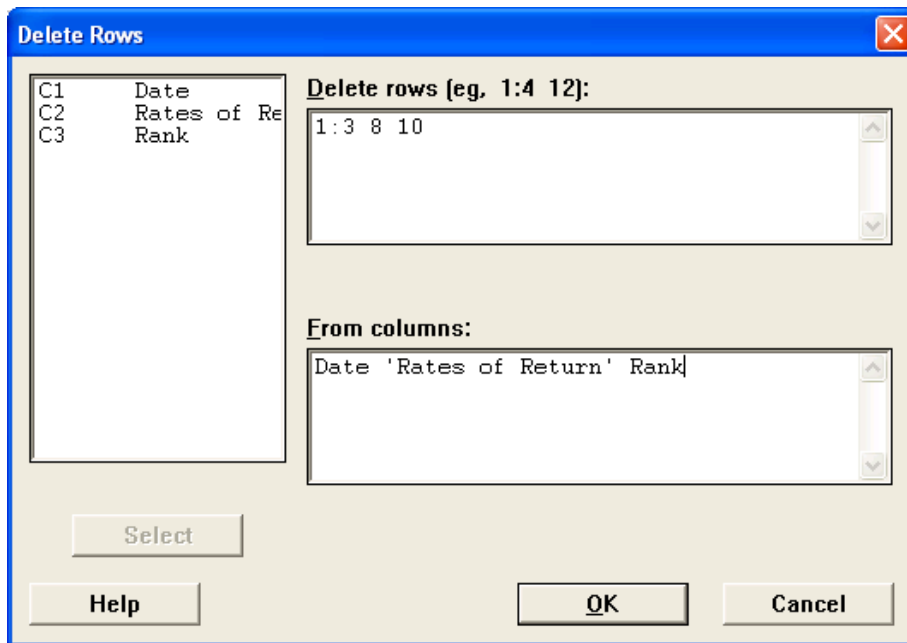
Here is a glimpse of the five rows that are left from the 10 shown in previous Rank example:

Date	Rates of Return	Rank
30-Sep-98	0.315	56
31-Aug-98	-0.0791	7
31-Jul-98	0.17	40
30-Jun-98	0.1263	35
30-Apr-98	0.1919	45

Note: You can also delete rows in Minitab the same way as you would in Excel.

d)   *Erase Variables*

You can erase any combination of columns, constants, and matrices (including their names).

Data → Erase Variables → Double-click on what you want to erase (For Dell Data, double-click on Rank column that was added previously) **→ OK.**
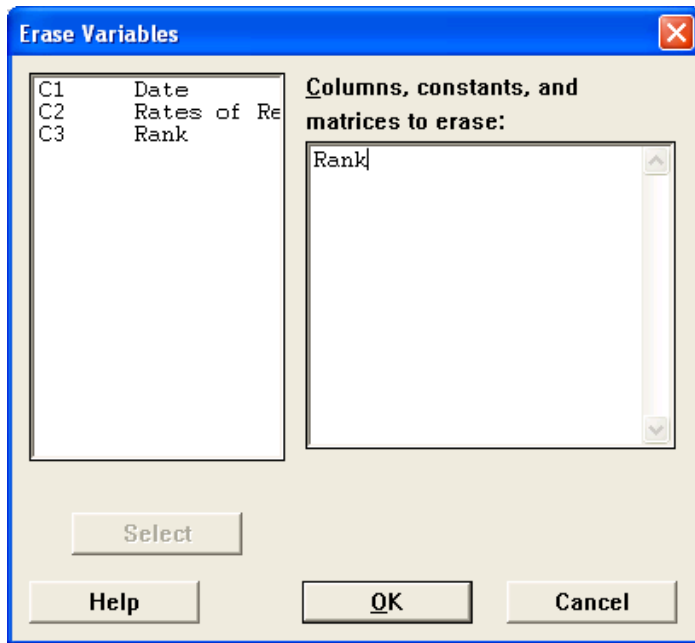
Here is a glimpse of what is left of five rows shown in the Previous Delete example:

Date	Rates of Return
30-Sep-98	0.315
31-Aug-98	-0.0791
31-Jul-98	0.17
30-Jun-98	0.1263
30-Apr-98	0.1919

Note: You can also delete columns in Minitab the same way as you would in Excel.

e)   Code

Use Code to change a value or set of values to new values.

Data → Code → This screen will appear:
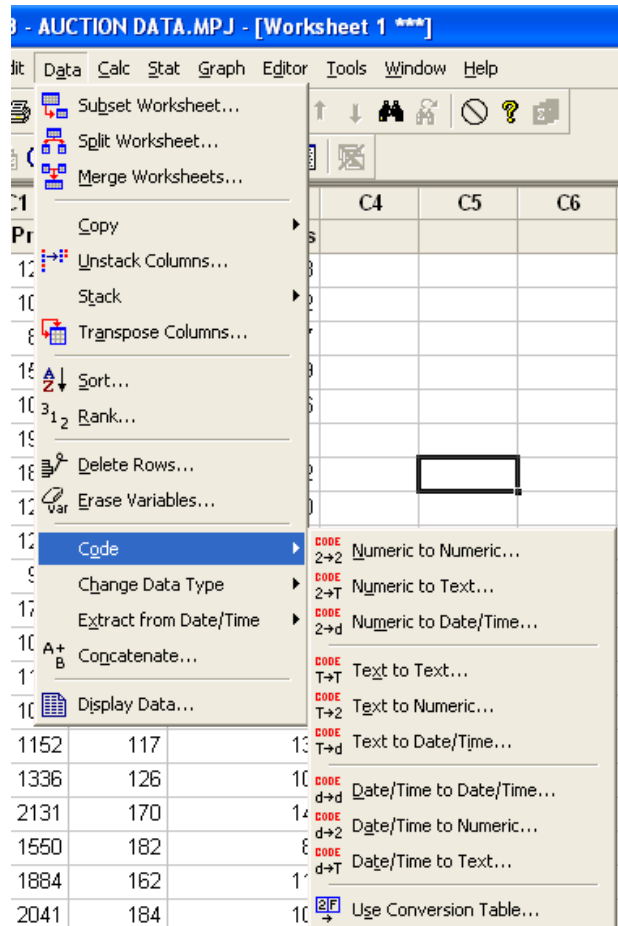
As you can see from this picture, there are many options to choose from at this point. Only an example of Numeric to Text will be shown here, but use this as a reference for other choices since the format of each one is essentially the same. Below are some examples of when you might want to use the various code commands.

Here are some examples of when you might use the various code commands.

Here are some examples of when you might use the various code commands.

Use...	To code...
Numeric to Numeric	• Test scores from 91 through 100 to a 4 (grade of A), from 81 through 90 to a 3 (grade of B), from 71 through 80 to a 2 (grade of C), from 61 through 70 to a 1 (grade of D), and 60 or below to a 0 (grade of F) • All occurrences of -99 to $*$ (the missing value symbol)
Numeric to Text	Test scores from 91 through 100 to an A, from 81 through 90 to a B, from 71 through 80 to a C, from 61 through 70 to a D, and 60 or below to an F.
Numeric to Date/Time	People's ID numbers to their birthdays; for example, 45234 to 11/15/75; 45235 to 12/4/65; etc.
Text to Text	Tennessee to TN, North Carolina to NC, Alabama to AL, Georgia to GA.

Text to Numeric	The letter grade A to a 4, B to a 3, C to a 2, D to a 1, and F to a 0.
Text to Date/Time	Month names to dates. For example, code January to 1/1/02, February to 02/01/02, etc.
Date/Time to Date/Time	All dates in January 2002 to January 1, 2002.
Date/Time to Numeric	Any date that falls between January and March to a 1 for the first quarter.
Date/Time to Text	Any date that falls in January to "January."

Ok, now for the example:

Data → Code → Numeric to Text → Double-click on variable you want to code (For Auction Data, double-click on Age) **→ Type name of new column in Into Columns box** (Type in Clocks) **→ Type in values or range of values you want to change in Original values box** (Type 100:150 in first rows and 151:200) **→ Type the new value with which you want to replace each original value in the New box** (Type New in the first row and Old in the second row) **→ OK.**

Code - Numeric to Text

C1	Auc Price
C2	Age
C3	No. of Bidd

Code data from columns:

Age

Into columns:

Clocks

Original values (eg, 1:4 12):

| 100:150 | New |
| 151:200 | Old |

Select   Help   OK   Cancel

Here are the first ten rows of the data that is displayed:

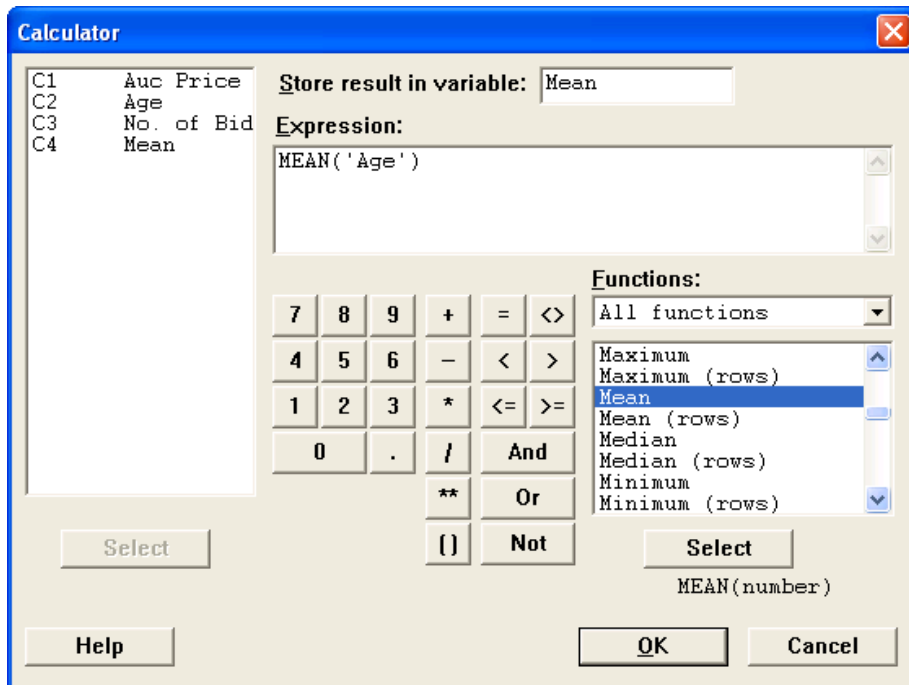| Auc Price	Age	No. of Bidders	Clocks
1235	127	13	New
1080	115	12	New
845	127	7	New
1522	150	9	New
1047	156	6	Old
1979	182	11	Old
1822	156	12	Old
1253	132	10	New
1297	137	9	New
946	113	9	New

As you can see, the word New is associated with ages 100-150 and Old is associated with ages 151-200.

f)   *Calculator*

Use the Calculator to do arithmetic operations, comparison operations, logical operations, functions, and column operations. Expressions may include columns, stored constants, numbers, and text, but not matrices.

Note: You cannot type formulas in the cells as you would in Excel.

Calc → Calculator → Type in name of column heading (For Auction Data, type in Mean) **→ Enter the mathematical operation you want Minitab to perform, selecting variables, buttons, and functions to build your expression** (For this example, double-click on mean under the Functions dropdown box, then double-click on Age) **→ OK.**

Here are the first few rows of the data displayed:

Auc Price	Age	No. of Bidders	Mean
1235	127	13	144.938
1080	115	12	
845	127	7	

Here is another example of an expression you can do. Let's say that we discovered that the grandfather clock ages were underestimated by 5 years. Thus, to add 5 years to each age, use this expression:
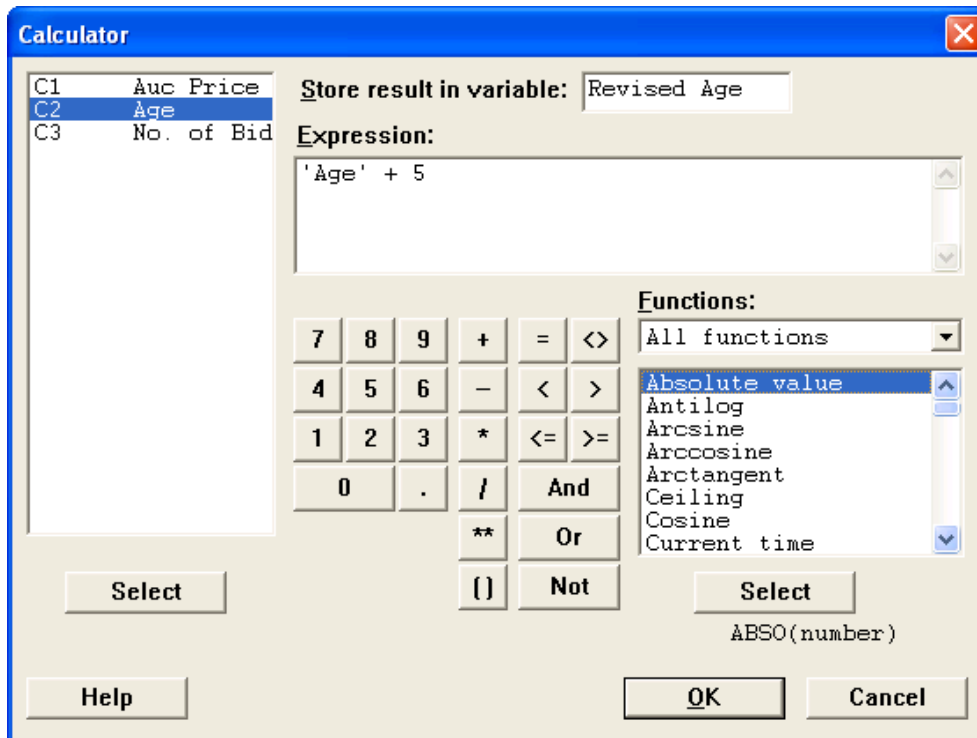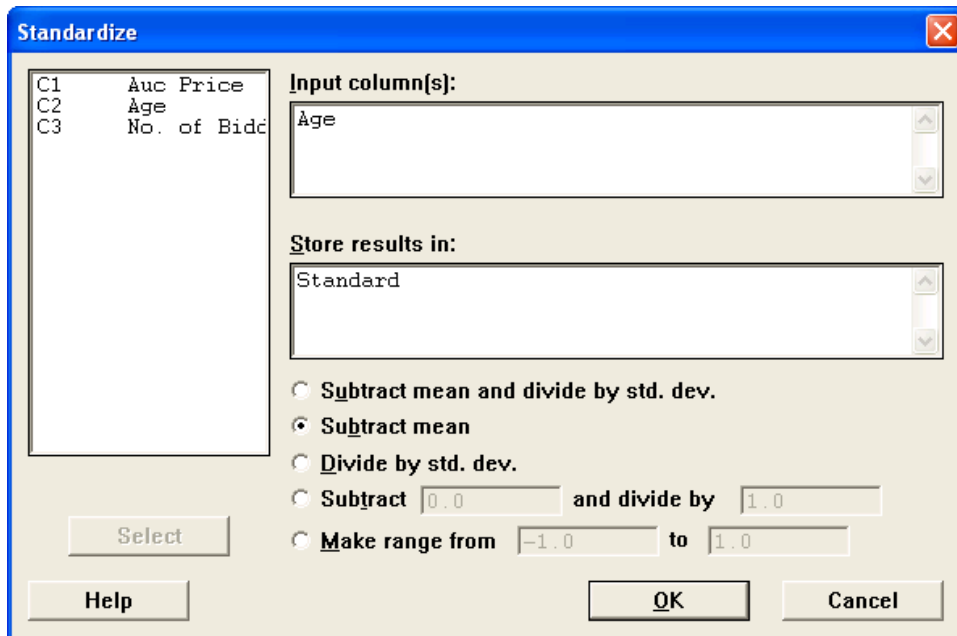
Here are the first few rows of the data displayed:

Auc Price	Age	No. of Bidders	Revised Age
1235	127	13	132
1080	115	12	120
845	127	7	132
1522	150	9	155
1047	156	6	161

g)   Standardize

The Standardize function centers and scales columns of data. Standardization is useful because it allows you to transform values into comparable units. For example, if you click on Subtract mean and divide by std. dev., the output column will tell you how many standard deviations the value is from the mean.

For the Auction Data, we can see how far the age of each clock is from the mean (in years) by:

Calc → Standardize → Double-click on the column you want to standardize (Double-click on Age) **→ Type in name of new column under Store results in** (Type in Standard) **→ Click in appropriate circle** (Click by Subtract mean) **→ OK.**

Here are the first five rows of data displayed:

| Auc Price	Age	No. of Bidders	Standard
1235	127	13	-17.9375
1080	115	12	-29.9375
845	127	7	-17.9375
1522	150	9	5.0625
1047	156	6	11.0625

h)   Subset Worksheet

Use this to copy specified rows from the active worksheet to a new worksheet. Using the Auction Data as an example, let's say that you only wanted to look at auction prices that were over $1000.

Data → Subset Worksheet →
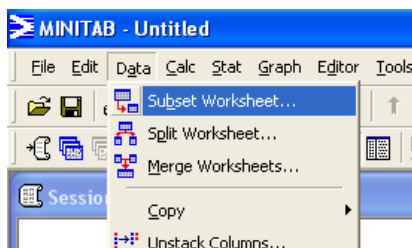
Then this screen will appear:
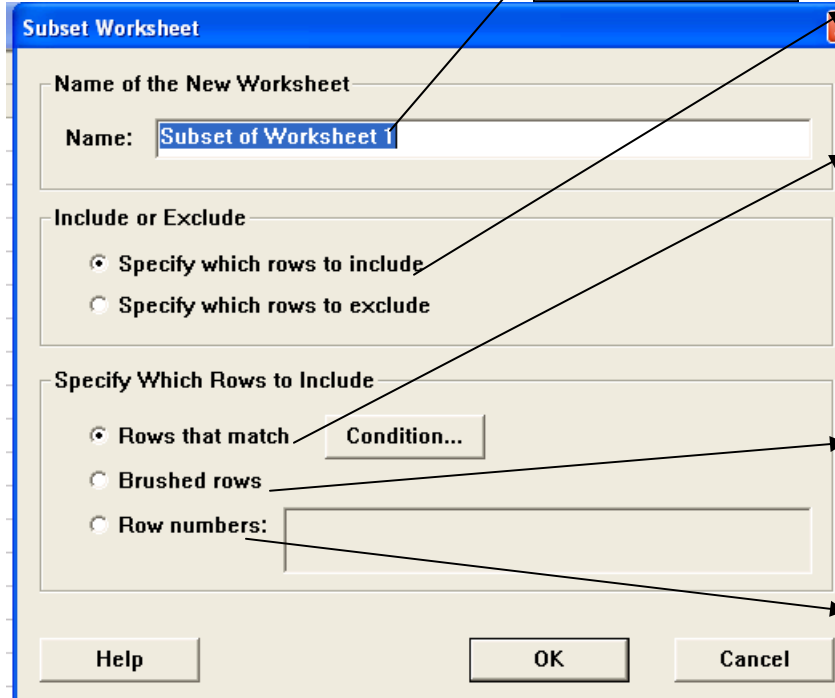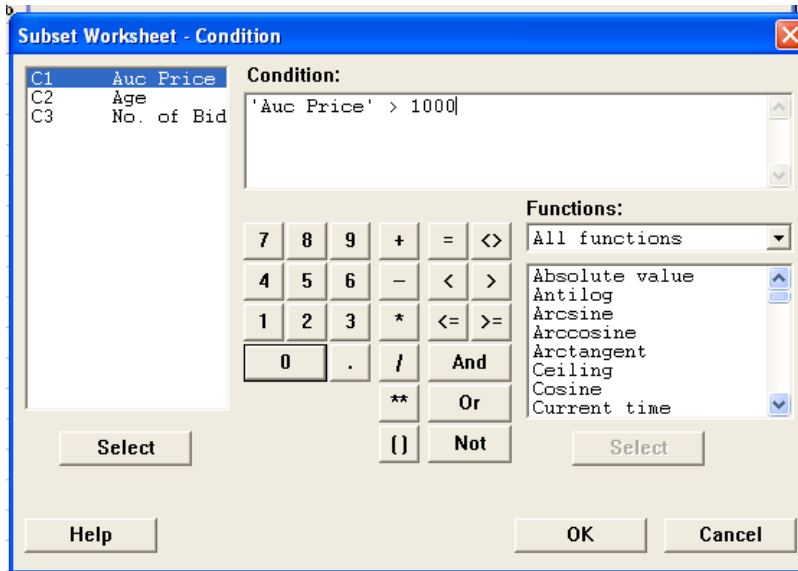
Name the new worksheet

Choose to include or exclude rows. (In this example, we'll choose the rows to include.)

Rows that match will base your subset on an expression. (which we will do in this example)

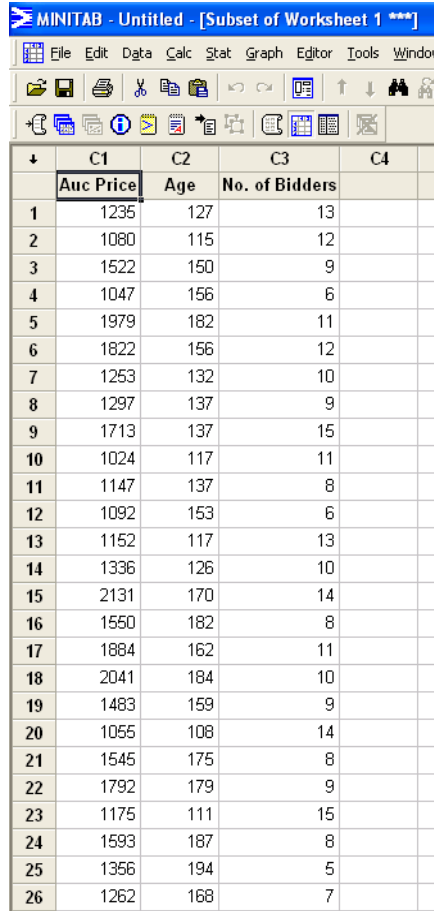Choose to base your subset on brushed points on a graph.

Choose to base your subset on specific row numbers, then type the row numbers.

Subset Worksheet

Name of the New Worksheet

Name: Subset of Worksheet 1

Include or Exclude
- Specify which rows to include
- Specify which rows to exclude

Specify Which Rows to Include
- Rows that match   Condition...
- Brushed rows
- Row numbers:

Help   OK   Cancel

Click on **Condition** to write your expression.

Subset Worksheet - Condition

C1   Auc Price
C2   Age
C3   No. of Bid

Condition:
'Auc Price' > 1000

Functions:

7	8	9	+	=	<>
4	5	6	−	<	>
1	2	3	*	<=	>=
0	.	/	And
		**	Or
		{}	Not

All functions

Absolute value
Antilog
Arcsine
Arccosine
Arctangent
Ceiling
Cosine
Current time

Select   Select

Help   OK   Cancel

In this example, **Double-click on Auc Price** so that it is listed under 'Condition' and then use the buttons to type > 1000 → **OK** → **OK.**

This new worksheet will appear:

↓	C1	C2	C3	C4
	Auc Price	**Age**	**No. of Bidders**	
1	1235	127	13	
2	1080	115	12	
3	1522	150	9	
4	1047	156	6	
5	1979	182	11	
6	1822	156	12	
7	1253	132	10	
8	1297	137	9	
9	1713	137	15	
10	1024	117	11	
11	1147	137	8	
12	1092	153	6	
13	1152	117	13	
14	1336	126	10	
15	2131	170	14	
16	1550	182	8	
17	1884	162	11	
18	2041	184	10	
19	1483	159	9	
20	1055	108	14	
21	1545	175	8	
22	1792	179	9	
23	1175	111	15	
24	1593	187	8	
25	1356	194	5	
26	1262	168	7	

i)   Split Worksheet

Using the Student Data, let's say we wanted to split the worksheets into two worksheets: one with the 'Male' data and one with the 'Female' data.

Data → Split Worksheet →

19
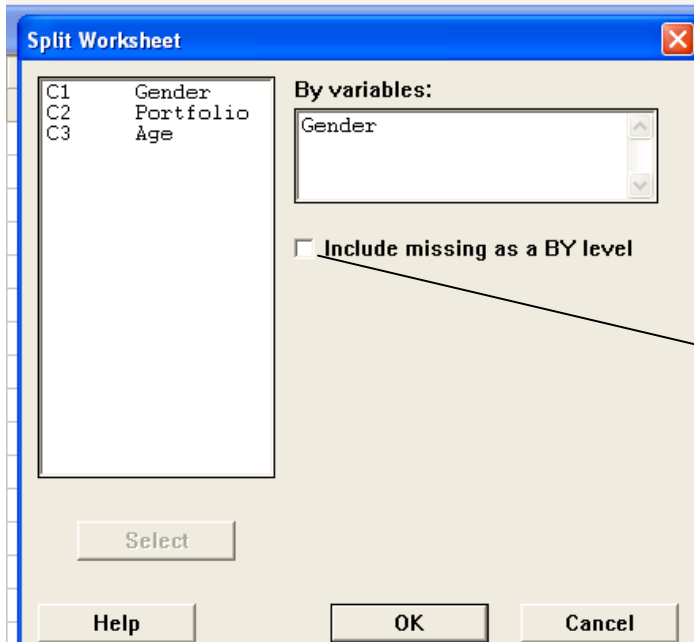
> By default, Split Worksheet ignores rows where the By variable contains a missing value. To treat missing values as an additional distinct value, check Include missing as a BY level.

Put cursor in 'By variables' box and **Double-click on appropriate variable → OK.**
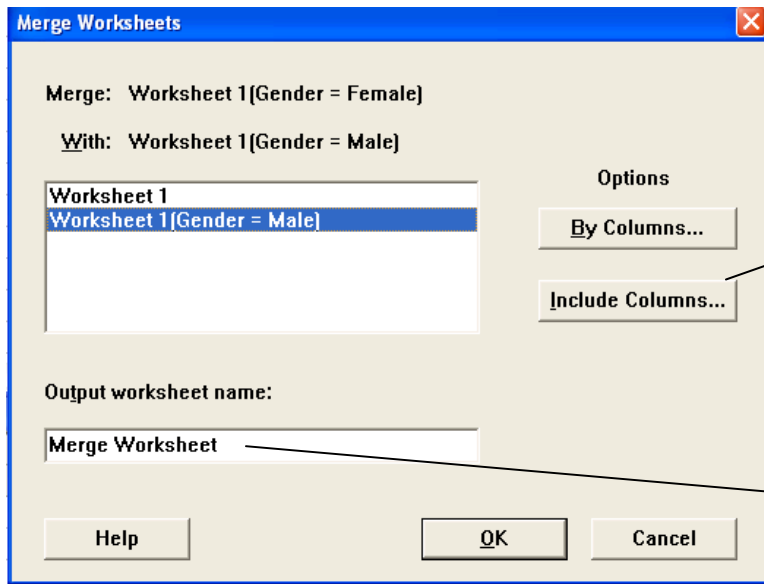
Two new worksheets will appear:

j)   Merge Worksheet

Use to combine two open worksheets into one new worksheet. For this example, we'll undo what was just shown above (i.e., the gender split example) and merge the two new worksheets we just created.

Data → Merge Worksheet →

Merge Worksheets

Merge: Worksheet 1(Gender = Female)

With: Worksheet 1(Gender = Male)

Options

Worksheet 1
Worksheet 1(Gender = Male)

By Columns...

Include Columns...

Output worksheet name:

Merge Worksheet

Help   OK   Cancel

Choose **Include Columns** to specify the columns from each original worksheet to include/exclude. By default, all columns will be included.

Name the new merged worksheet.

Click the appropriate worksheet to merge → OK.

The new merged worksheet will appear:

	C1-T	C2-T	C3	C4-T	C5-T	C6
	Gender:Worksheet 1(G	Portfolio:Worksheet 1(G	Age:Worksheet 1(Gend	Gender:Worksheet 1(G	Portfolio:Worksheet 1(G	Age:Worksheet 1(Gend
1	Female	Marketing	25	Male	Finance	29
2	Female	Marketing	25	Male	IST	33
3	Female	Consulting	29	Male	Finance	29
4	Female	Marketing	29	Male	Finance	35
5	Female	Finance	32	Male	Consulting	25
6				Male	Consulting	27
7				Male	Operations	27
8				Male	Operations	33
9				Male	Marketing	38
10				Male	Marketing	28
11				Male	Marketing	28
12				Male	IST	24
13				Male	Finance	25
14				Male	Consulting	24
15						

If you click on the *By Columns…* button, you can customize how the worksheets merge. By Columns options allow you to match observations from the two worksheets to be merged. For example, let's say that two samples of students were taken. Each sample contains one student from each of five groups. Here are the two samples:

21

	Sample 1					Sample 2		
Group	**Gender**	**Portfolio**	**Age**		**Group**	**Gender**	**Portfolio**	**Age**
1	Male	Finance	29		5	Male	Operations	33
2	Male	IST	33		4	Male	Operations	27
3	Male	Finance	29		3	Female	Marketing	25
4	Female	Marketing	25		2	Male	Consulting	27
5	Male	Finance	35		1	Male	Consulting	25

Let's say you would like to combine the results into one worksheet so that all the information for each group is in the same row.

Data → Merge Worksheet →

Click on *By Columns…*

Move appropriate variables from the left to the right by clicking on them and then hitting
> (Here we clicked on Group and moved it over to the By columns: box) **→ OK → OK.**

The new merged worksheet will appear:

	C1	C2-T	C3-T	C4	C5	C6-T	C7-T	C8
	Group:Worksheet 2	Gender:Worksheet 2	Portfolio:Worksheet 2	Age:Worksheet 2	Group:Worksheet 1	Gender:Worksheet 1	Portfolio:Worksheet 1	Age:Worksheet 1
1	1	Male	Consulting	25	1	Male	Finance	29
2	2	Male	Consulting	27	2	Male	IST	33
3	3	Female	Marketing	25	3	Male	Finance	29
4	4	Male	Operations	27	4	Female	Marketing	25
5	5	Male	Operations	33	5	Male	Finance	35
6								
7								

k)   Deleting Observations with Missing Data

Minitab replaces a missing data point with an asterisk (*). If you want to eliminate all observations with missing data points, here is one way to do it – there may be other ways. Let's say that the Auction Data was missing some data points and looked like this:

	C1	C2	C3	C4
↓	Auc Price	Age	No. of Bidders	
1	1235	127	13	
2	1080	*	12	
3	845	127	7	
4	1522	150	*	
5	1047	156	6	
6	*	182	11	
7	1822	156	12	
8	1253	132	10	
9	1297	137	9	
10	946	113	9	
11	1713	137	15	
12	1024	117	11	
13	1147	*	8	
14	1092	153	6	
15	1152	117	13	
16	1336	126	10	
17	2131	170	14	
18	1550	182	8	
19	1884	162	*	
20	2041	184	10	
21	845	143	6	
22	1483	159	9	
23	*	108	14	
24	1545	175	8	
25	729	108	6	
26	1792	179	9	
27	1175	111	15	
28	1593	187	*	
29	785	111	7	
30	744	115	7	
31	1356	194	5	
32	1262	168	7	
33				

It is easy to see which observations have missing data here, but it would be more difficult if the sample size was larger. So we are going to Sort each column so that the asterisks are at the top of the columns, and then delete those rows.

Data → Sort → Double-click on all variables for Sort columns → Double-click on the first variable for By column (Auc Price in this example) **→ Click on box next to Descending → Click on appropriate selection under Store sorted data in → OK.**

Descending order will place the asterisks at the top of the column.

In this example, I am going to create and name a new worksheet every time I sort the data. However, you can click on Original column(s) to keep the same worksheet.
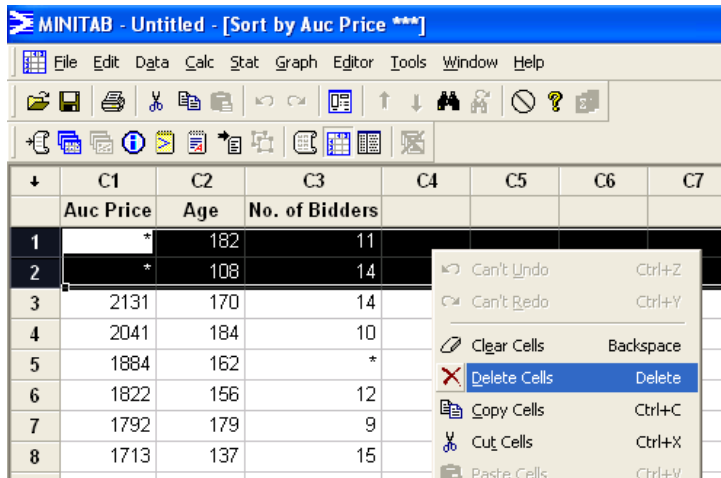
The new worksheet will appear:
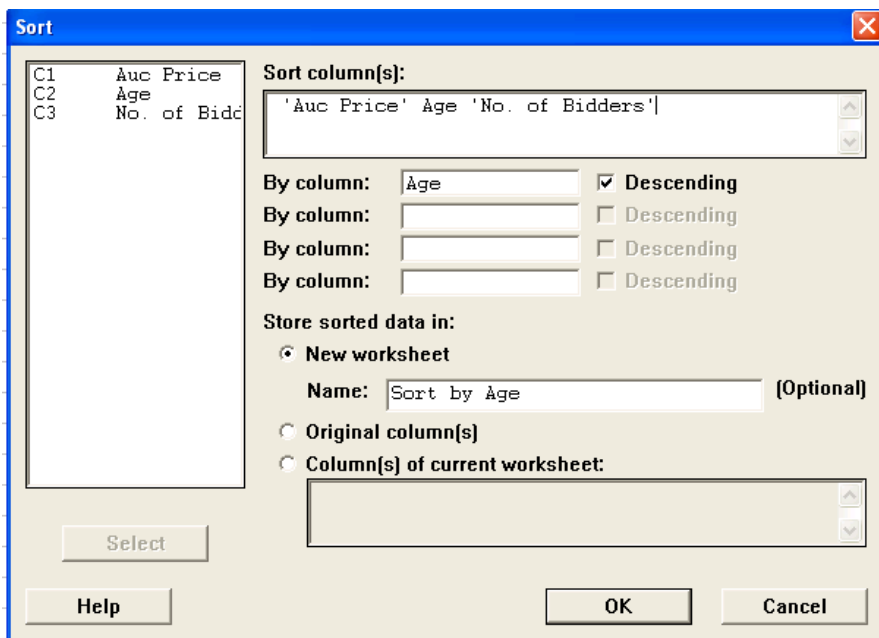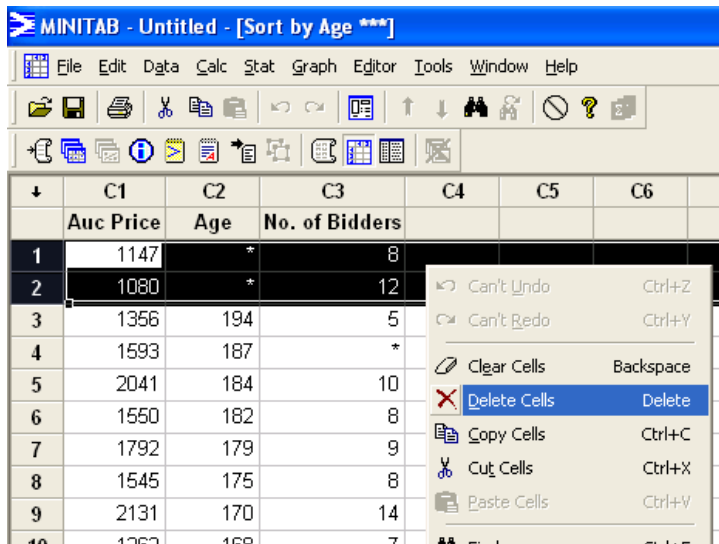
MINITAB - Untitled - [Sort by Auc Price *]**

File Edit Data Calc Stat Graph Editor Tools

	C1	C2	C3	C4
	Auc Price	Age	No. of Bidders	
1	*	182	11	
2	*	108	14	
3	2131	170	14	
4	2041	184	10	
5	1884	162	*	
6	1822	156	12	
7	1792	179	9	
8	1713	137	15	
9	1593	187	*	
10	1550	182	8	
11	1545	175	8	
12	1522	150	*	
13	1483	159	9	
14	1356	194	5	
15	1336	126	10	
16	1297	137	9	
17	1262	168	7	
18	1253	132	10	
19	1235	127	13	
20	1175	111	15	
21	1152	117	13	
22	1147	*	8	
23	1092	153	6	
24	1080	*	12	
25	1047	156	6	
26	1024	117	11	
27	946	113	9	
28	845	127	7	
29	845	143	6	
30	785	111	7	
31	744	115	7	
32	729	108	6	
33				

Highlight the rows with the asterisks → Right-click and a list will appear so that you can click on Delete Cells (Or you can go to **Edit → Delete Cells**). Then those rows will disappear. Now we will follow the same procedure the rest of the columns.

Data → Sort → Double-click on all variables for Sort columns → Double-click on the second variable for By column (Age in this example) **→ Click on box next to Descending → Click on appropriate selection under Store sorted data in → OK.**

The new worksheet will appear:

↓	C1	C2	C3	C4
	Auc Price	**Age**	**No. of Bidders**	
1	1147	*	8	
2	1080	*	12	
3	1356	194	5	
4	1593	187	*	
5	2041	184	10	
6	1550	182	8	
7	1792	179	9	
8	1545	175	8	
9	2131	170	14	
10	1262	168	7	
11	1884	162	*	
12	1483	159	9	
13	1822	156	12	
14	1047	156	6	
15	1092	153	6	
16	1522	150	*	
17	845	143	6	
18	1713	137	15	
19	1297	137	9	
20	1253	132	10	
21	1235	127	13	
22	845	127	7	
23	1336	126	10	
24	1152	117	13	
25	1024	117	11	
26	744	115	7	
27	946	113	9	
28	1175	111	15	
29	785	111	7	
30	729	108	6	

Highlight the rows with the asterisks → Right-click and a list will appear so that you can click on Delete Cells (Or you can go to **Edit → Delete Cells**). Then those rows will disappear.

Data → Sort → Double-click on all variables for Sort columns → Double-click on the third variable for By column (No. of Bidders in this example) **→ Click on box next to Descending → Click on appropriate selection under Store sorted data in → OK.**

The new worksheet will appear:

	C1	C2	C3	C4
	Auc Price	**Age**	**No. of Bidders**	
1	1593	187	*	
2	1884	162	*	
3	1522	150	*	
4	1713	137	15	
5	1175	111	15	
6	2131	170	14	
7	1235	127	13	
8	1152	117	13	
9	1822	156	12	
10	1024	117	11	
11	2041	184	10	
12	1253	132	10	
13	1336	126	10	
14	1792	179	9	
15	1483	159	9	
16	1297	137	9	
17	946	113	9	
18	1550	182	8	
19	1545	175	8	
20	1262	168	7	
21	845	127	7	
22	744	115	7	
23	785	111	7	
24	1047	156	6	
25	1092	153	6	
26	845	143	6	
27	729	108	6	
28	1356	194	5	
29				

Highlight the rows with the asterisks → Right-click and a list will appear so that you can click on Delete Cells (Or you can go to **Edit → Delete Cells**). Then those rows will disappear.

The newest worksheet does not contain any observations with missing data points.

III.   Descriptive Statistics

Displays N, N*, Mean, SE Mean, StDev, Min, Q1, Median, Q3, and Max

a)   Descriptive Statistics for one variable

Stat → Basic Statistics → Display Descriptive Statistics → Double-click on appropriate variable (For Dell Data, double-click on Rates of Return so that it is displayed under Variables).
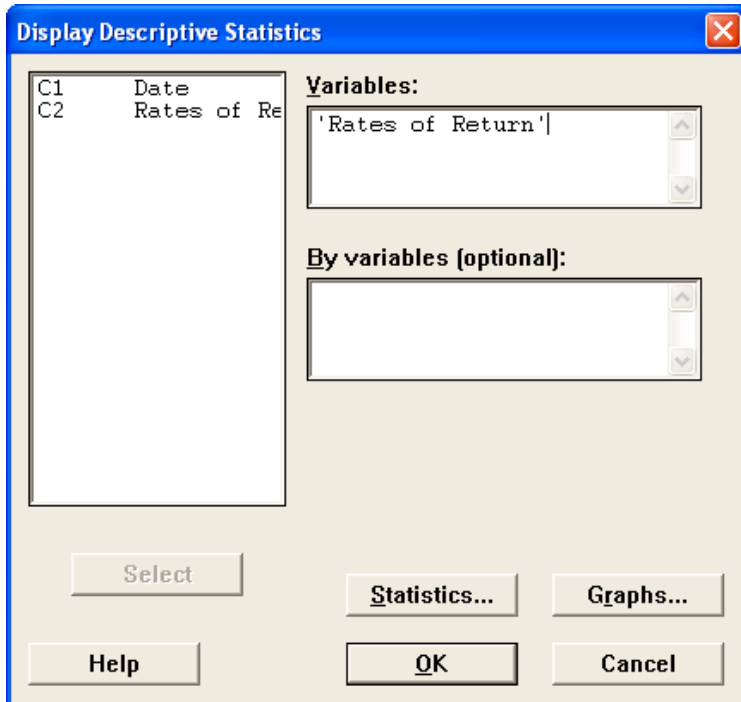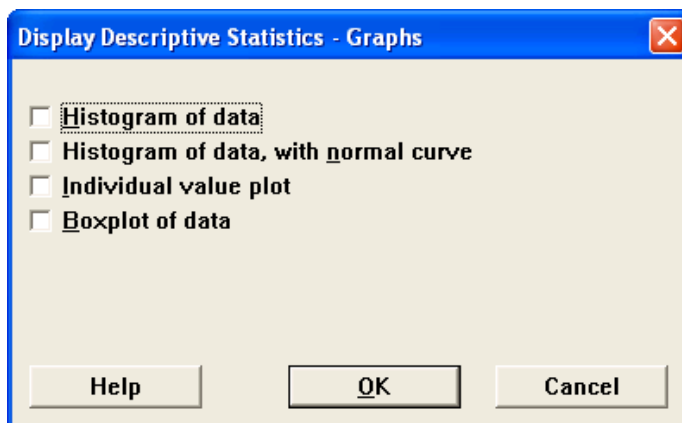

As you can see from the screen above, you are given the option to alter the output by clicking on the buttons. If you click on the *Statistics* button, this screen will appear:
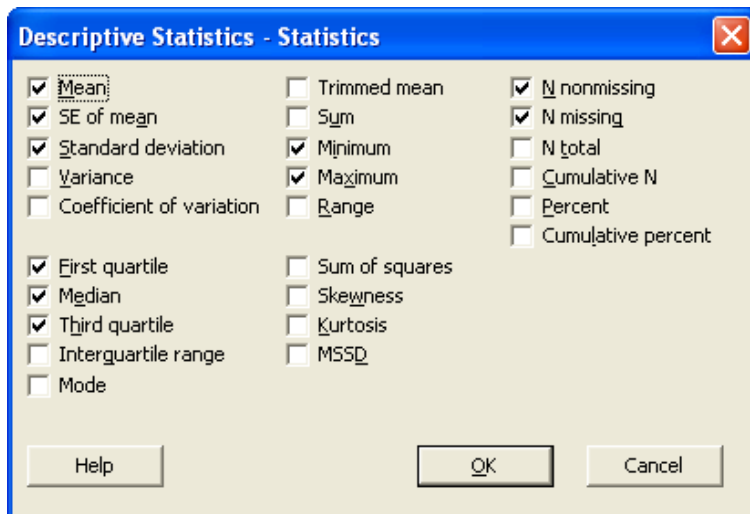
The checked items will be displayed in the output. To check or uncheck an item, click in the box to the left of the word.

If you click on the *Graphs* button, this screen will appear:

To display any of these graphs (in addition the descriptive statistics displayed in the session window), click in the box. (For purposes of this example, I have not clicked on any graphs since graphs will be explained in the next section.)

To display the data, click on **OK.** For the Dell example, this information is displayed in the session window:

Descriptive Statistics: Rates of Return

```
Variable          N  N*    Mean  SE Mean   StDev  Minimum       Q1  Median
Rates of Return  60   0  0.0907   0.0195  0.1511  -0.2175  -0.0304  0.0784

Variable            Q3  Maximum
Rates of Return  0.1931   0.4561
```

b) *Descriptive statistics for one variable, grouped by a second variable*

Stat → Basic Statistics → Display Descriptive Statistics → Double-click on appropriate variable → Click in By variables (optional) box and then double-click on appropriate variable → OK. (For Auction Data, double-click on Auction Price so that it is displayed under Variables. Then move the cursor into the By variables (optional) box and double-click on No. of bidders so that it is displayed under By variables (optional).)

```
Display Descriptive Statistics                             ✕

C1     Auc Price      Variables:
C2     Age
C3     No. of Bidd   ┌─────────────────────────────┐
                     │ 'Auc Price' │              ▲│
                     │                             ││
                     │                             ▼│
                     └─────────────────────────────┘

                     By variables (optional):
                     ┌─────────────────────────────┐
                     │ 'No. of Bidders'           ▲│
                     │                             ││
                     │                             ▼│
                     └─────────────────────────────┘


       Select              Statistics...      Graphs...


       Help                   OK              Cancel
```
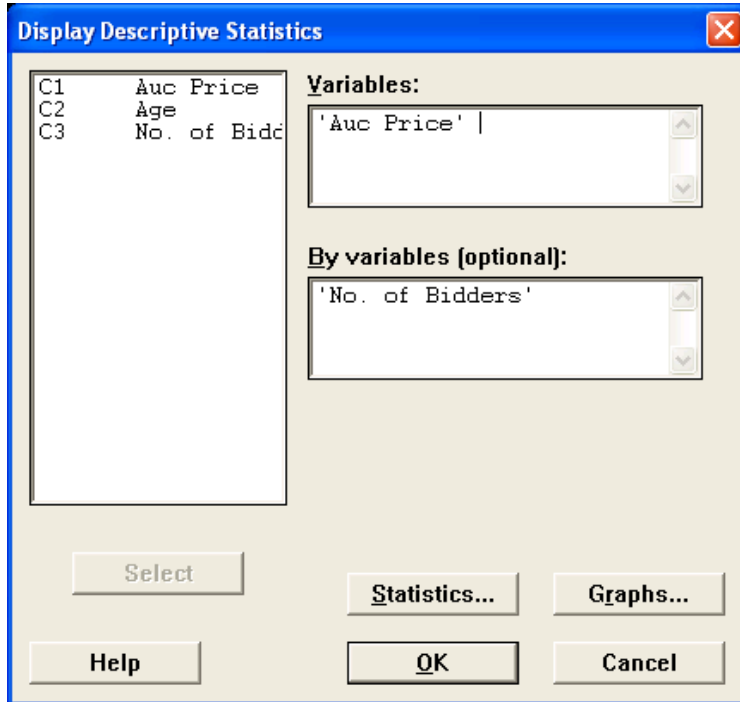
For the Auction Data example, this information is displayed in the session window:

Descriptive Statistics: Auc Price

```
           No. of
Variable   Bidders   N   N*     Mean   SE Mean   StDev   Minimum      Q1   Median
Auc Price     5      1   0    1356.0         *       *    1356.0       *   1356.0
              6      4   0     928.3      85.4   170.8     729.0   758.0    946.0
              7      4   0      909        119     239       744     754      815
              8      4   0     1459        104     209      1147    1247     1548
              9      5   0     1408        140     313       946    1122     1483
             10      3   0     1543        250     433      1253    1253     1336
             11      3   0     1629        304     526      1024    1024     1884
             12      2   0     1451        371     525      1080       *     1451
             13      2   0    1193.5      41.5    58.7    1152.0       *   1193.5
             14      2   0     1593        538     761      1055       *     1593
             15      2   0     1444        269     380      1175       *     1444
```

```
              No. of
Variable     Bidders      Q3   Maximum
Auc Price     5            *    1356.0
              6        1080.8   1092.0
              7          1158     1262
              8          1582     1593
              9          1657     1792
             10          2041     2041
             11          1979     1979
             12            *      1822
             13            *    1235.0
             14            *      2131
             15            *      1713
```

Note: If you see a * in the output, that indicates that the value could not be calculated. In this example, the numerous * appear because N is not large enough in each group to calculate all the descriptive statistics. (e.g. There is only one instance where the number of bidders equals 5, and thus SE Mean, StDev, Q1, and Q3 could not be calculated with only one data point)

c)   Store Descriptive Statistics

This feature adds the descriptive statistics to the data worksheet instead of displaying the output in the session window:

Stat → Basic Statistics → Store Descriptive Statistics → Double-click on appropriate variable (For Dell Data, double-click on Rates of Return so that it is displayed under Variables).

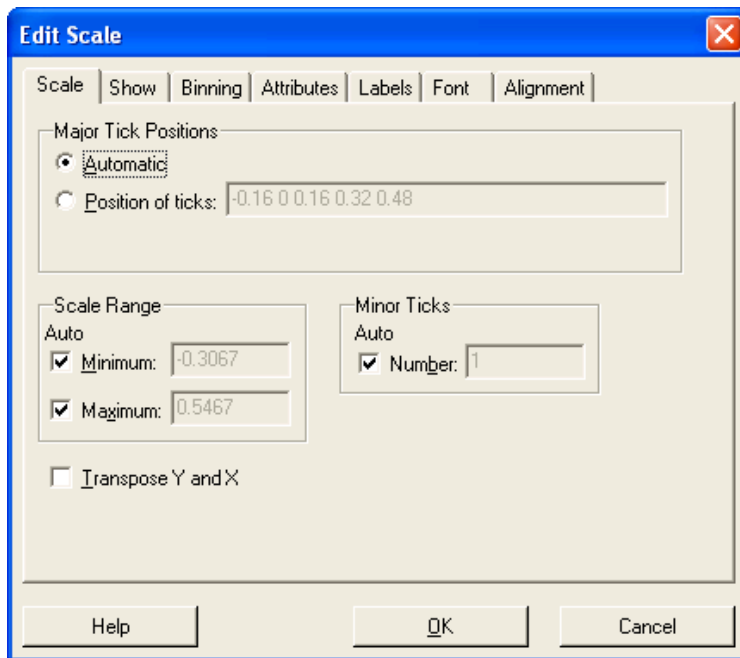The By variables option will not be explained again here. See Descriptive statistics for one variable, grouped by a second variable for information.
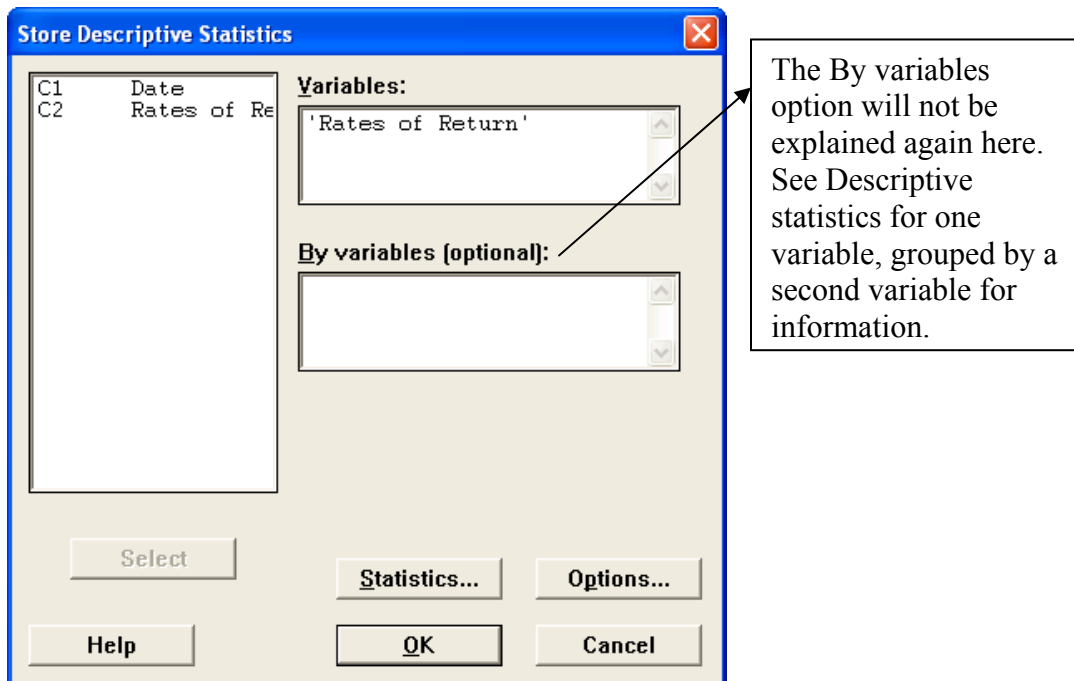
As you can see from the screen above, you are again given the option to alter the output by clicking on the buttons. If you click on the *Statistics* button, this screen will appear:
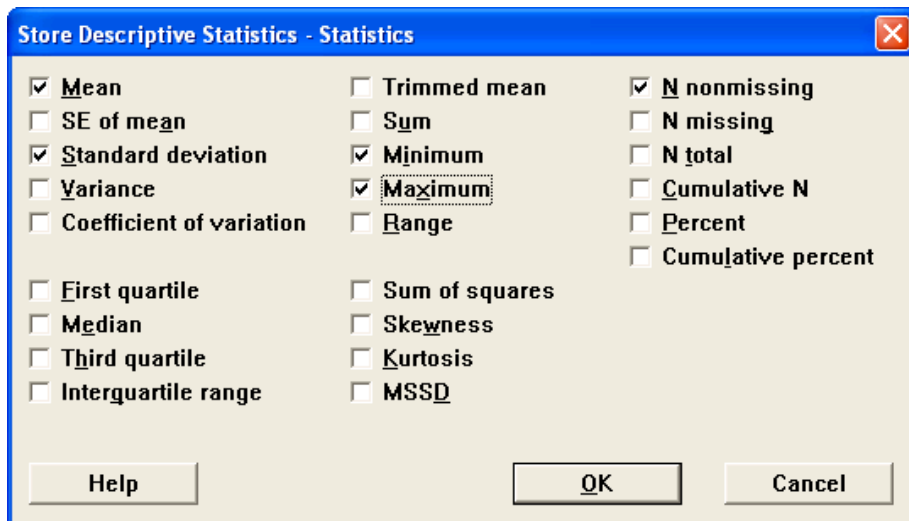
If you click on *Options* button, this screen will appear:

Check this option if you want Minitab to append the appropriate statistics to each row of input data.

Uncheck this option if you do not want Minitab to include columns in the summary data that indicate the levels of the By variables.

If you choose more than one By variable, Minitab stores summary statistics for all combinations of the By variable levels, including combinations for which there are no data (called empty cells). If you do not want to store empty cells, uncheck this option.

Check this option to include missing values as a distinct level of the By variable.

 Since this box is not checked in this example, Minitab stores the requested statistics at the top of the worksheet only, by default.

To display the data, click on **OK.** For the Dell example, here are the first five rows of data displayed in the worksheet:

Date	Rates of Return	Mean1	StDev1	Minimum1	Maximum1	N1
31-Jan-94	-0.0276	0.0907033	0.151094	-0.2175	0.4561	60
28-Feb-94	0.1364					
31-Mar-94	0.01					
29-Apr-94	-0.1287					
31-May-94	0.3011					

You can calculate various statistics on columns. Column statistics are displayed in the Session window, and are optionally stored in a constant.

Calc → Column Statistics → Click by the Statistic you want calculated (For Auction Data, click by Standard Deviation) **→ Double-click on appropriate column in Input variable box** (Double-click on No. of Bidders) **→ OK.**
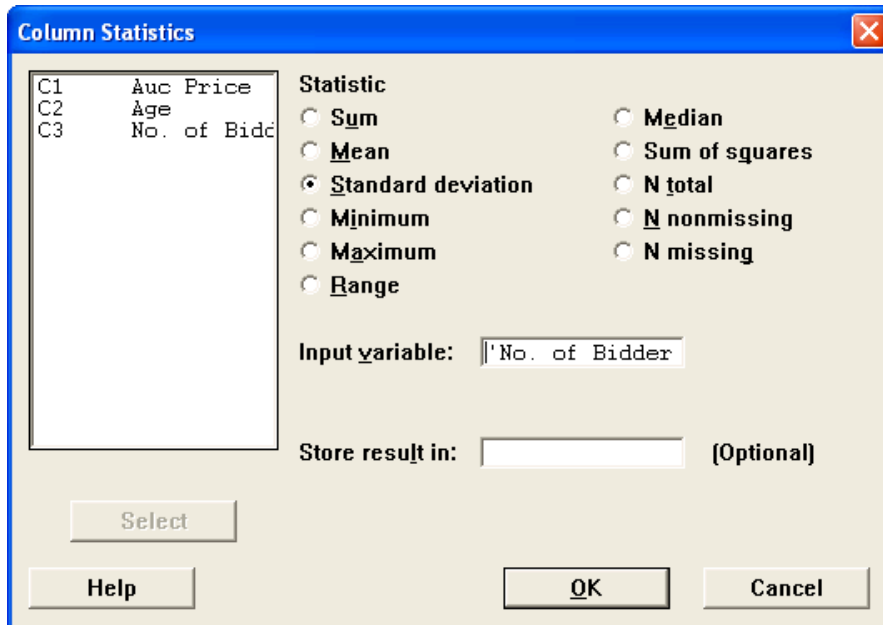
This output is displayed in the session window:

Standard Deviation of No. of Bidders

```
Standard deviation of No. of Bidders = 2.83963
```

e) *Row Statistics*

You can compute one value for each row in a set of columns. The statistic is calculated across the rows of the column(s) specified and the answers are stored in the corresponding rows of a new column.

Calc → Row Statistics → Click by the Statistic you want calculated → Double-click on appropriate variable(s) in Input variables box → Type the name of the new column that will be created → OK.

Calculating Row Statistics does not make sense using the example data because it is not meaningful in context. Thus, an example is not given here. However, in order to see what row statistics are able to be calculated, the screen shot is shown below.
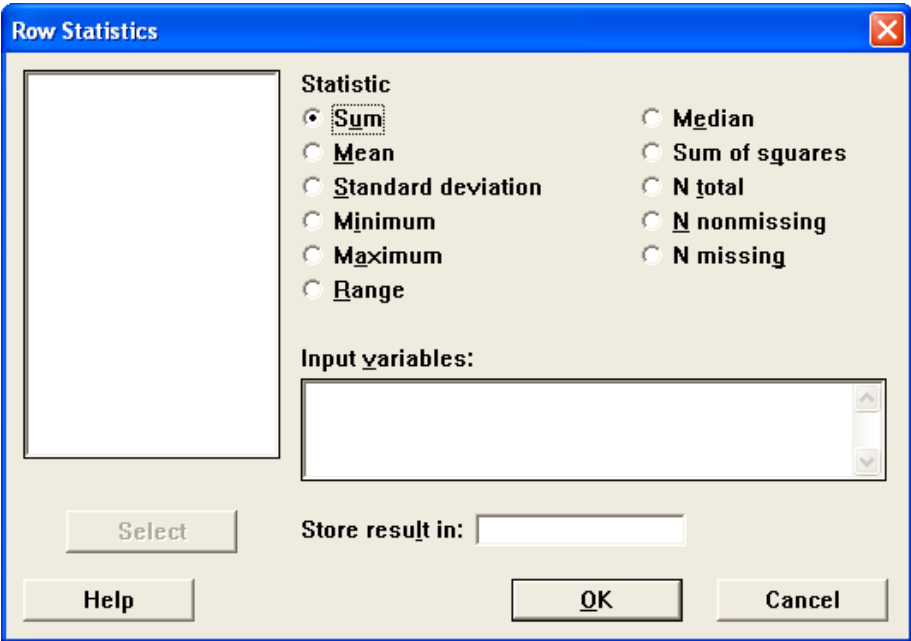
IV. Graphs

a) *Histogram*

Using the Dell Data that is now inserted into Minitab, a histogram can be made by going to **Graph → Histogram →** Then this screen will appear:
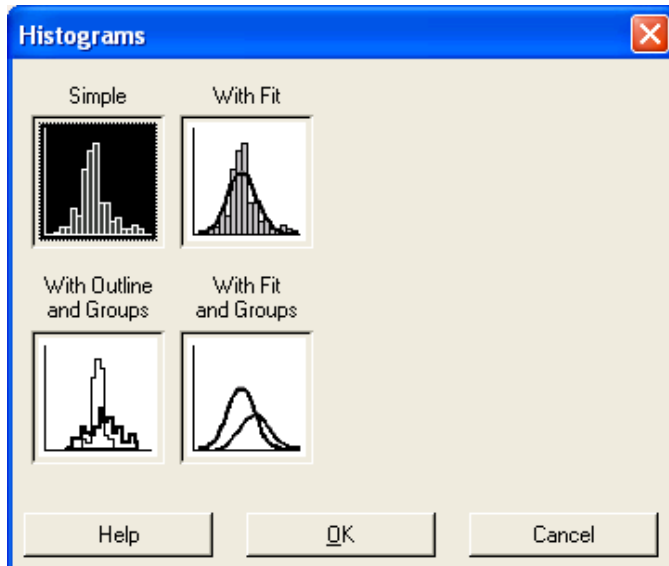
Click on appropriate graph and then click **OK.** (For this example, we will display the simple histogram). **→ Double-click on appropriate variable** (For Dell Data, double-click on Rates of Return so that it is displayed under Graph Variables) **→ OK.**

This histogram will display:

Histogram of Rates of Return
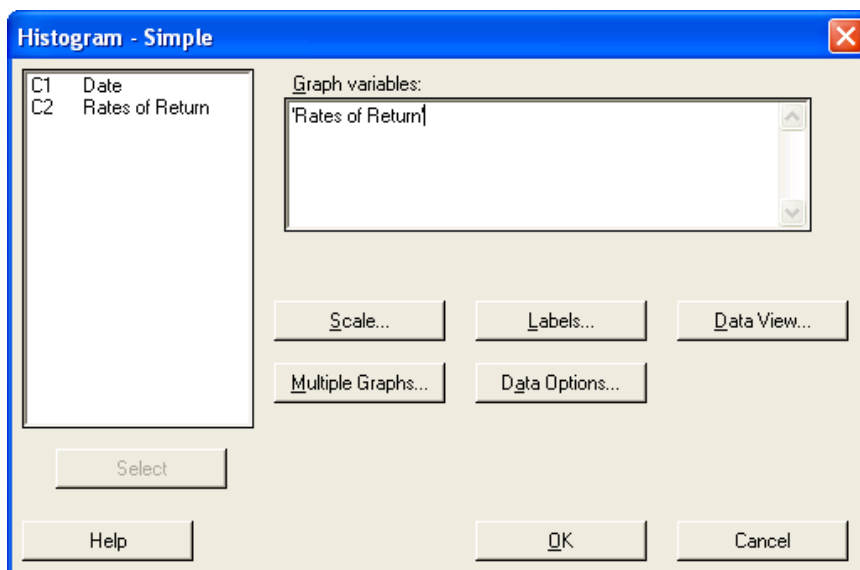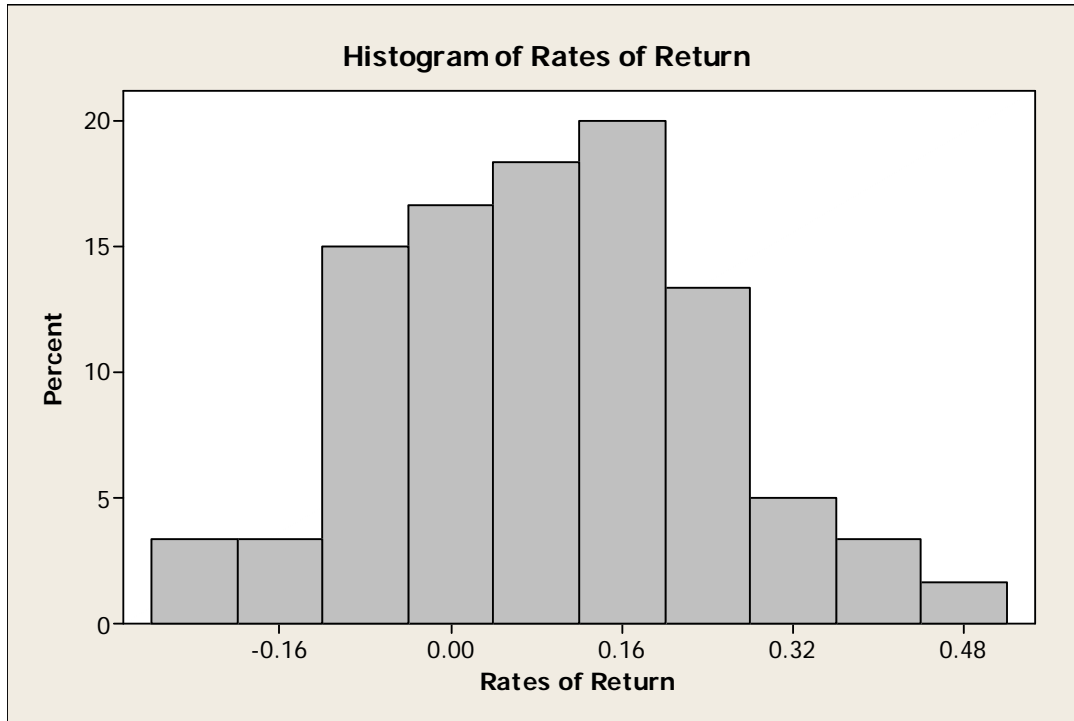
Note: You are able to edit the graph at this point. On the graph below, the arrows represent where you can double-click to make changes to the graph. You can do this type of editing on most graphs.

Histogram of Rates of Return
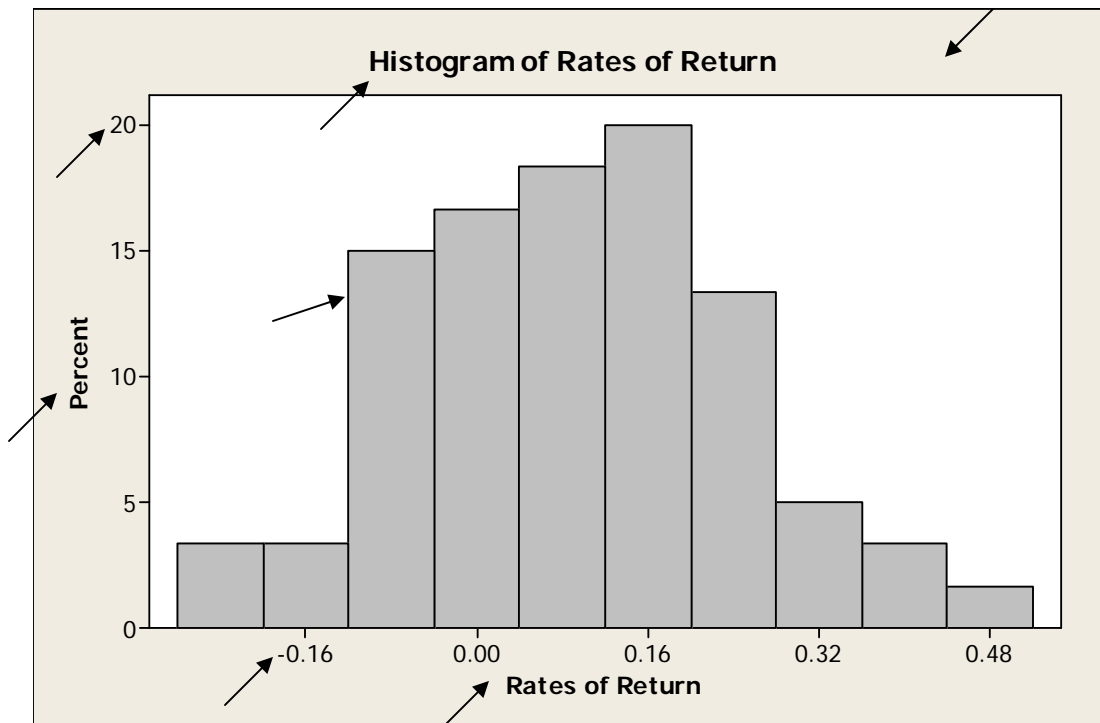
Let's say you wanted to edit the scale on the x-axis. By double-clicking on any of the x-axis numbers (For this Dell example, you could double-click on -0.16), this screen will then appear:

Edit Scale

Scale | Show | Binning | Attributes | Labels | Font | Alignment

Major Tick Positions
- ● Automatic
- ○ Position of ticks: `-0.16 0 0.16 0.32 0.48`

Scale Range
Auto
- ☑ Minimum: `-0.3067`
- ☑ Maximum: `0.5467`

Minor Ticks
Auto
- ☑ Number: `1`

- ☐ Transpose Y and X

Help | OK | Cancel

> Use to specify the range of the scale, the number of major and minor ticks, and their placement.

This screen shows the *Scale* tab. Another way to edit the scale is to click on the *Binning* tab. By doing so, this screen will appear:

Edit Scale

Scale | Show | Binning | Attributes | Labels | Font | Alignment

Interval Type
- ● Midpoint
- ○ Cutpoint

Interval Definition
- ○ Automatic
- ○ Number of intervals: `8`
- ● Midpoint/Cutpoint positions:

`-0.2 -0.1 0 0.1 0.2 0.3 0.4 0.5`

Help | OK | Cancel
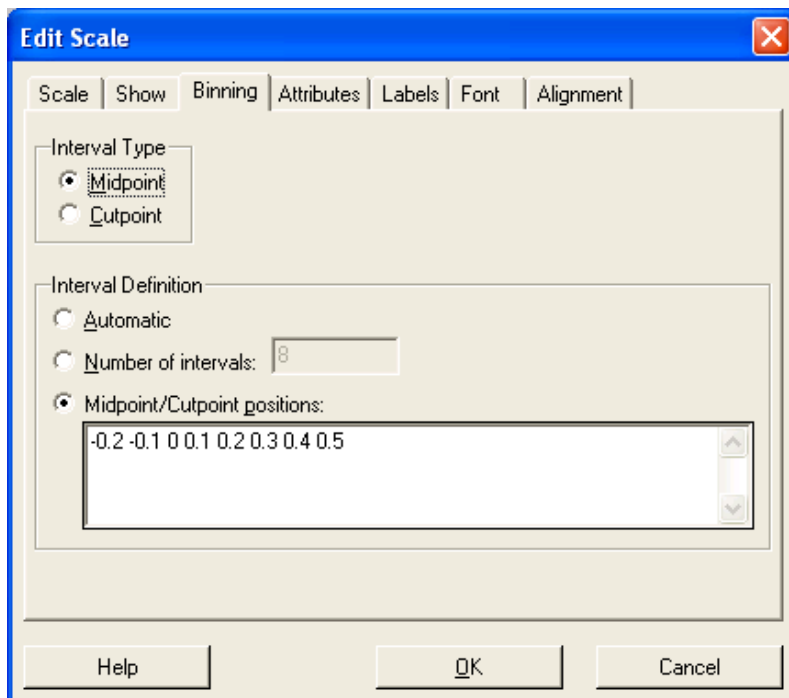
(The default sets the Interval Definition to Automatic. However, for this Dell example, click by Midpoint/Cutpoint positions and replace the numbers given with the new numbers shown above.)

If you click on the *Show* tab, this screen will appear:
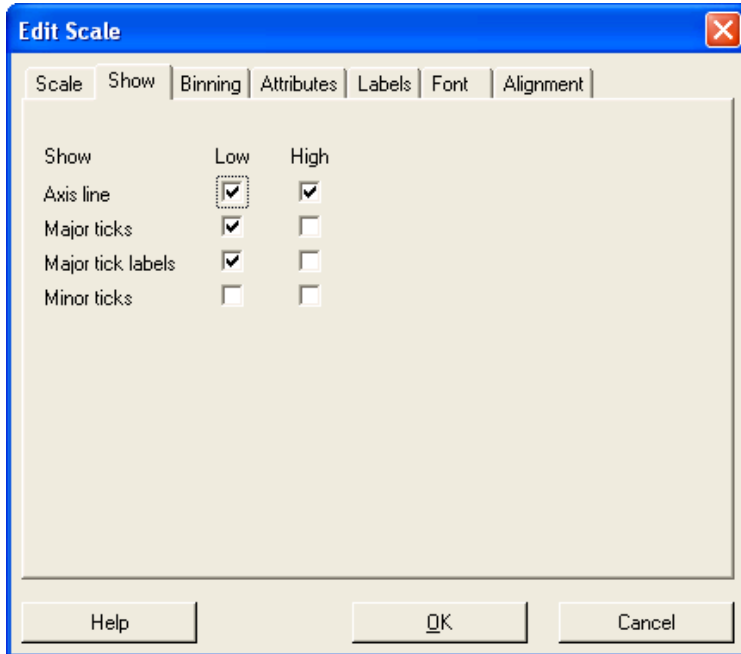
Use to show or hide axis lines, ticks, and labels.

If you click on the *Attributes* tab, this screen will appear:
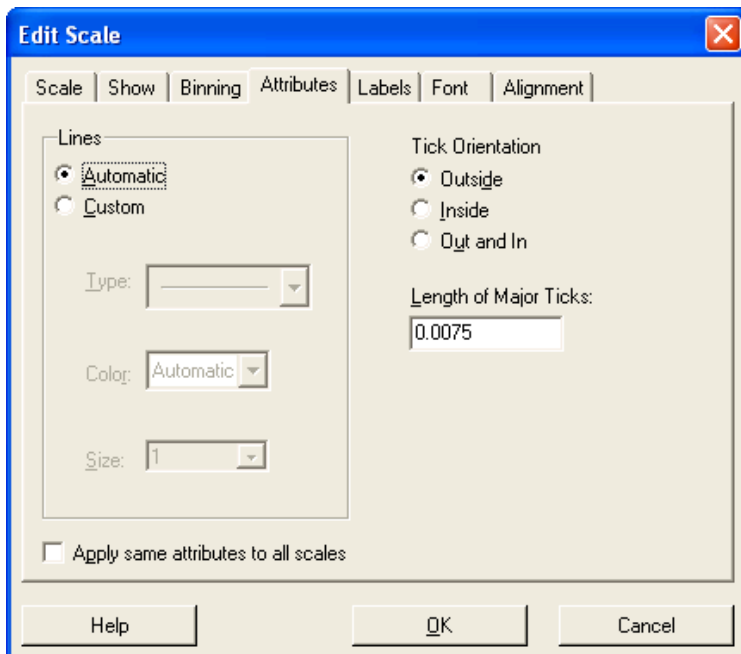
Use to change the type, color, and size of the axis line, as well as the tick orientation and length.

If you click on the *Labels* tab, this screen will appear:
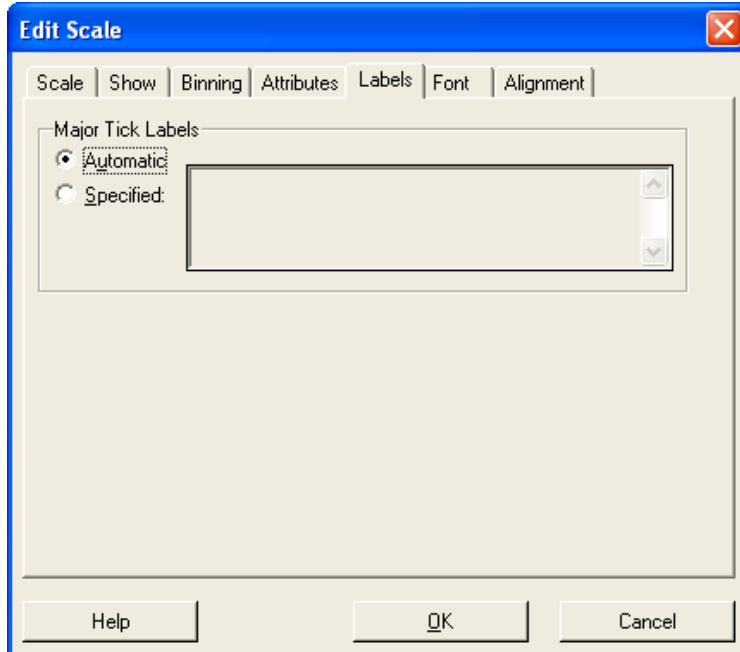
Use to change tick labels.

If you click on the *Font* tab, this screen will appear:
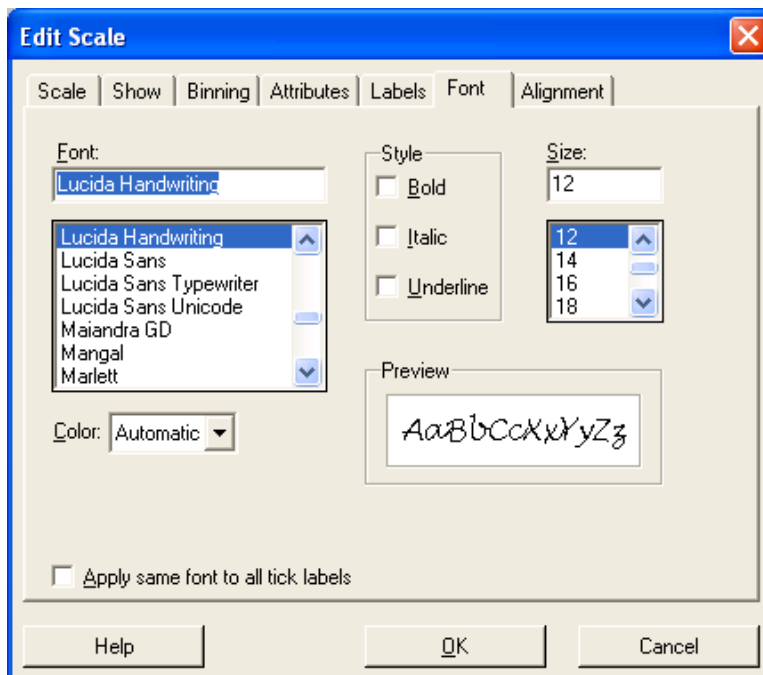
Use to change the tick label font or attributes.

(The default is set to Tahoma Font, Size 10. For this example, choose Lucida Handwriting Font, Size 12.)

If you click on the *Alignment* tab, this screen will appear:
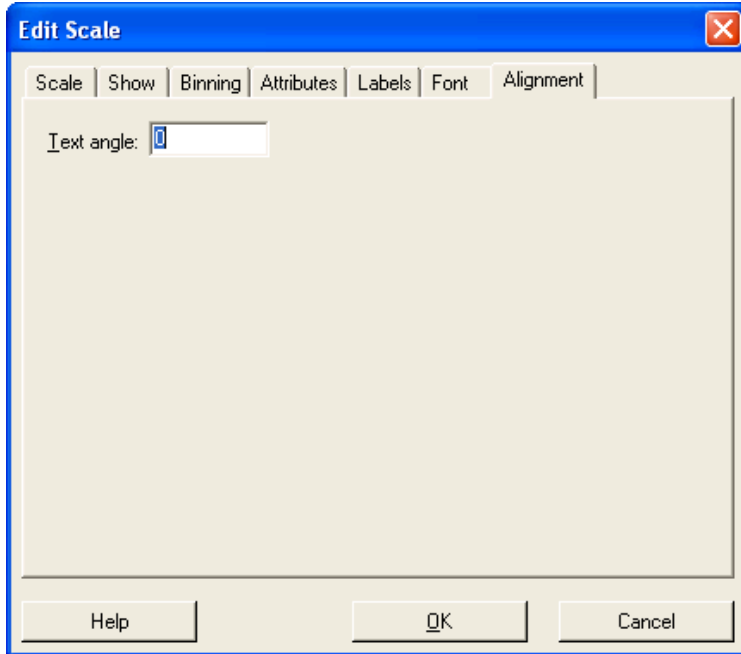
Use to change the angle of the tick labels to improve legibility.

If you click on **OK**, the new histogram will display:
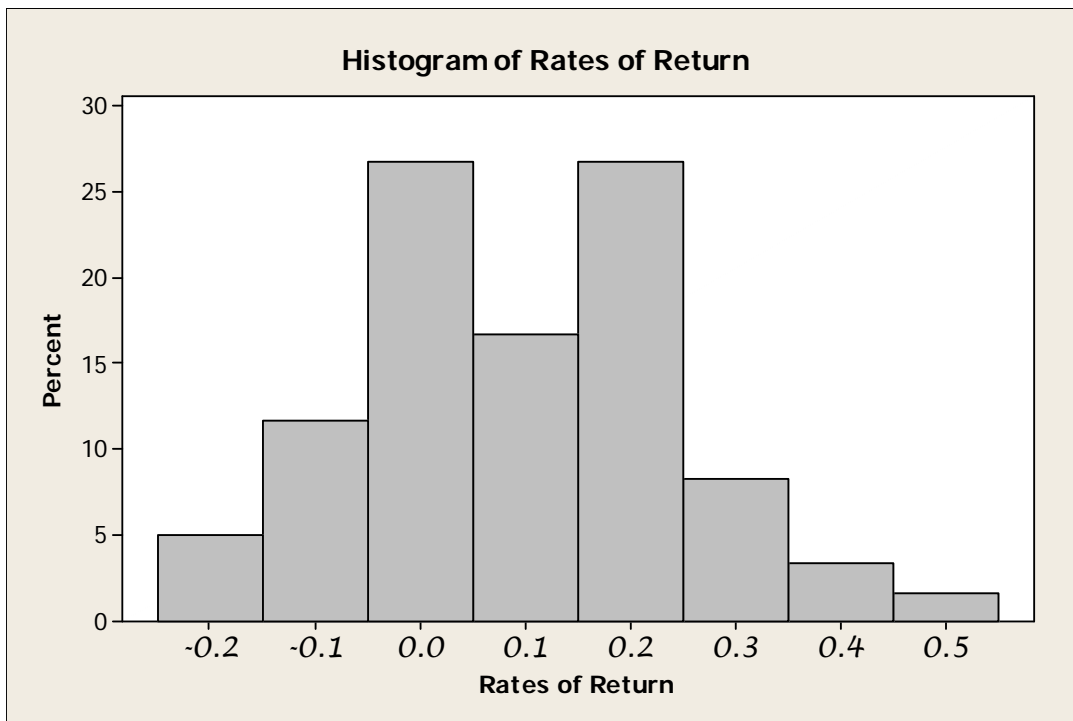
As you can see, the binning, size, and font have been changed in this example. Since we originally double-clicked on one of the x-axis numbers, we were able to make changes regarding that aspect of the graph. Likewise, you can make changes to other parts of the graph by double-

clicking on the appropriate spot. The details for all the other arrows (displayed on page 26) are not going to be explained here. Basically, you can change the way the text, bars, and background are displayed.

Another way to alter graphs is to use the buttons. If we go back to our original histogram example, after going to **Graph → Histogram → OK → Double-clicking on appropriate variable**, we are back to this screen:
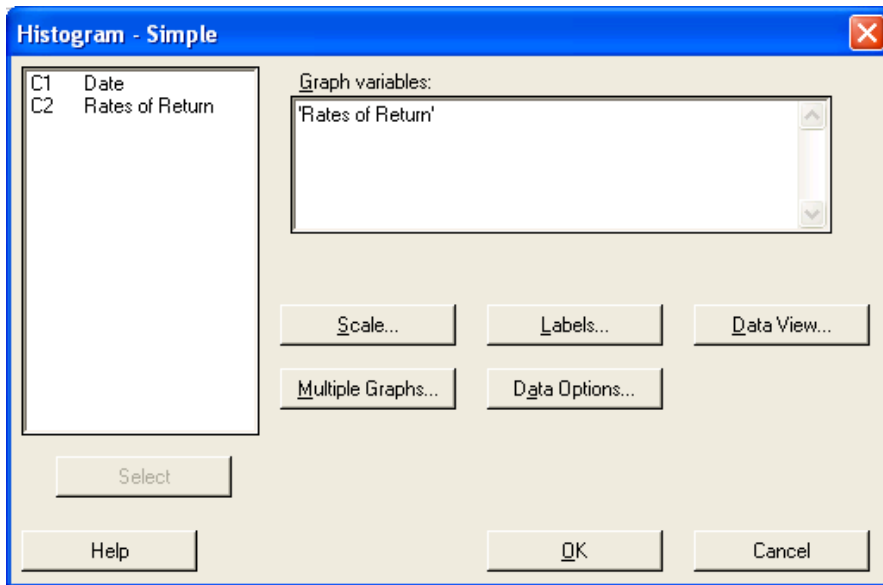
Here you are given the option to alter the output by clicking on the buttons. If you click on the *Scale* button, this screen will appear:
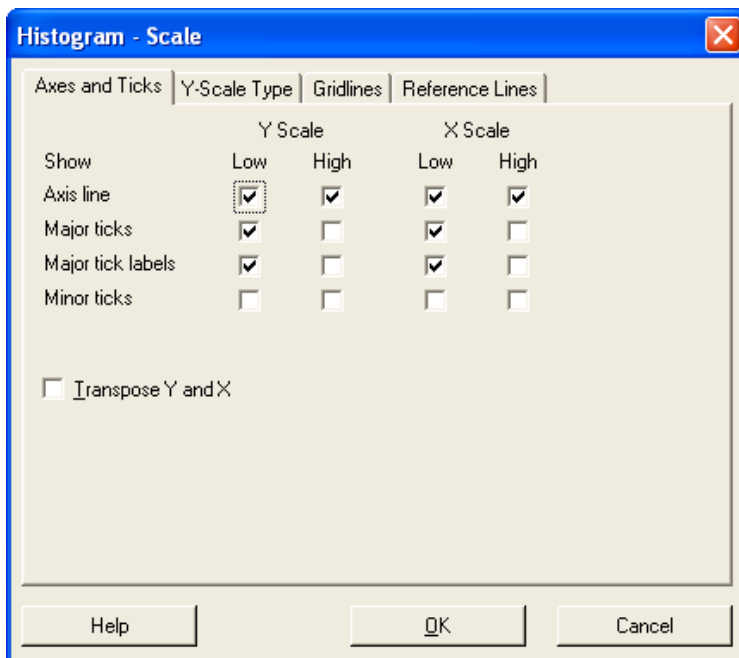
Use to specify which scale elements to display and where to display them.

This screen shows the *Axes and Ticks* tab. If you click on the *Y-Scale Type* tab, this screen will appear:
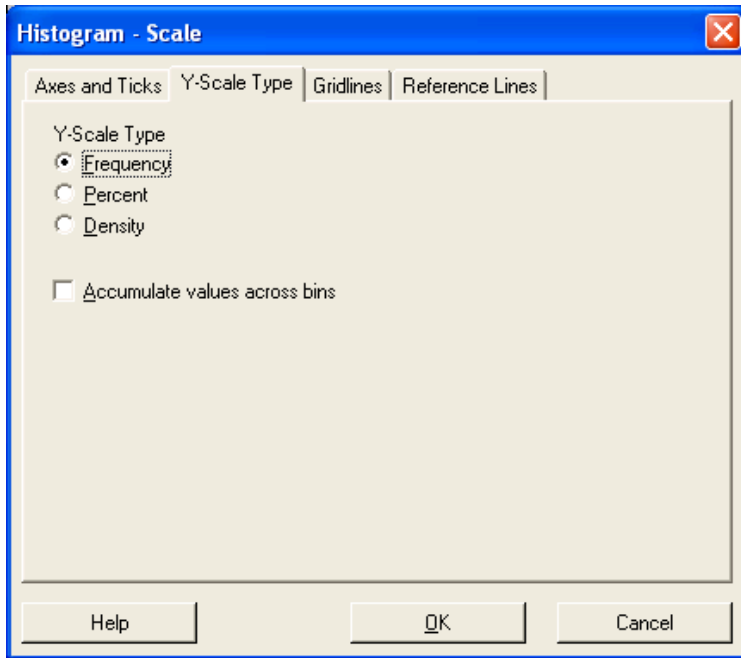
Use to choose the y-scale type.

(The default is set for Percent, but for this Dell example, click by Frequency.)
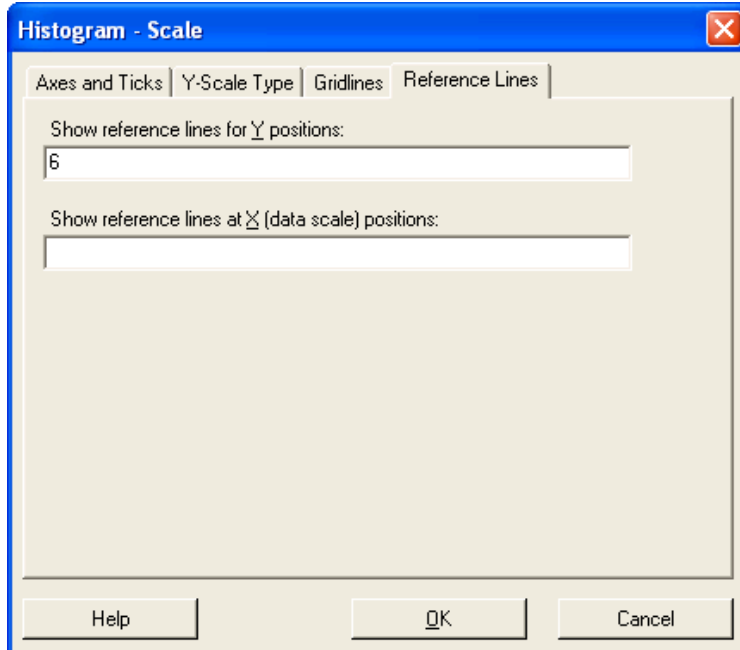
If you click on the *Gridlines* tab, this screen will appear:
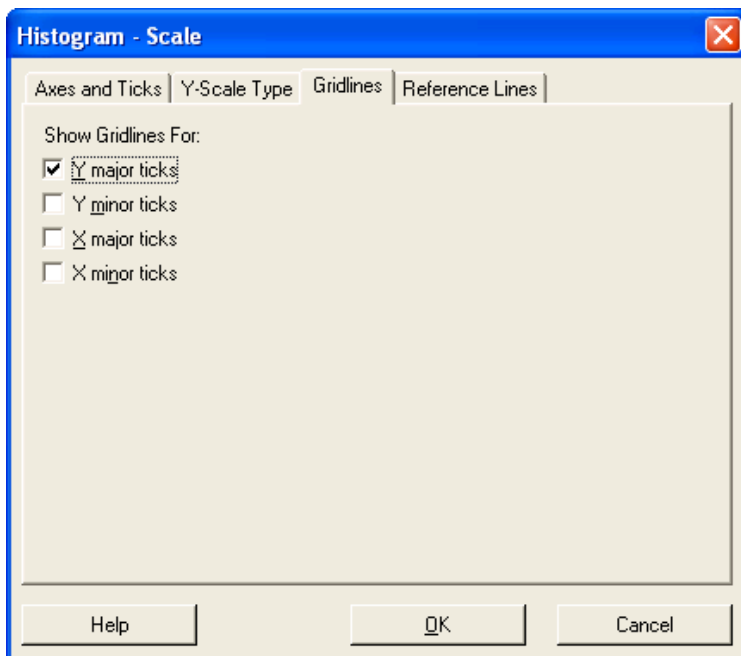
Use to display gridlines at the major and minor tick positions. If you choose gridlines at the minor tick positions, gridlines are also drawn at the major tick positions.

(None of the boxes are checked by default, but for this example, click by Y major ticks.)

If you click on the *Reference Lines* tab, this screen will appear:

Histogram - Scale

Axes and Ticks | Y-Scale Type | Gridlines | Reference Lines

Show reference lines for Y positions:

```
6
```

Show reference lines at X (data scale) positions:

```

```

Help   OK   Cancel

Use to display reference lines.

(There are no references lines by default, but for this example, type 6 to show a reference line at y = 6.)

If you click on the **Labels** button, this screen will appear:

Histogram - Labels

Titles/Footnotes | Data Labels

Title:

Subtitle 1:

Subtitle 2:

Footnote 1:

Footnote 2:

Help   OK   Cancel

Use to display titles and footnotes.

This screen shows the *Titles/Footnotes* tab. If you click on the *Data Labels* tab, this screen will appear:
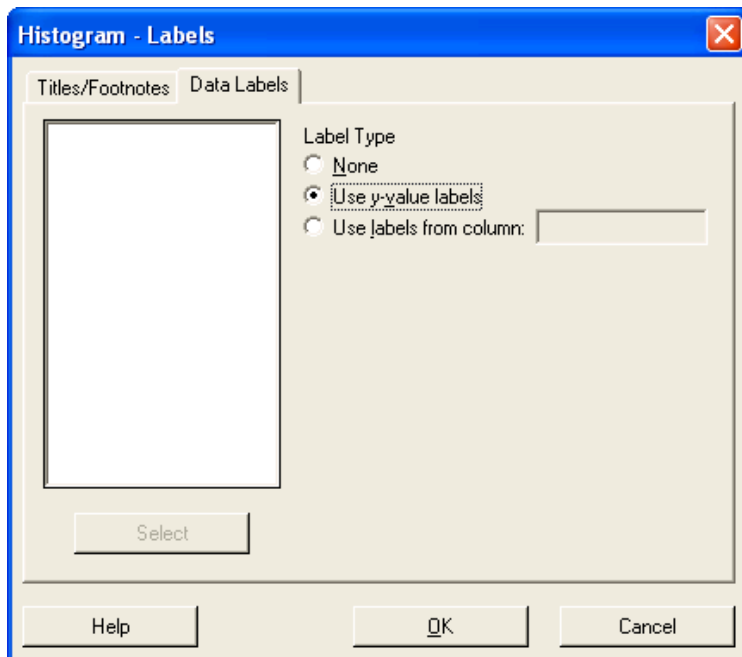
Use to label each data point.

(The default is set for None, but click by Use y-value labels for this example.)

If you click on the **Data View** button, this screen will appear:
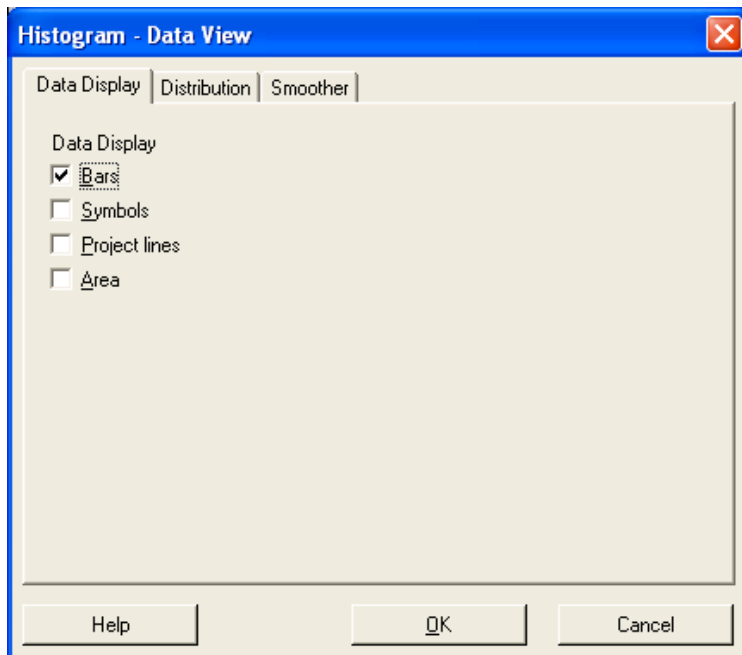
Use to represent the data with one or more data display types, including bars, symbols, project lines, and areas.

This screen shows the *Data Display* tab. If you click on the *Distribution* tab, this screen will appear:
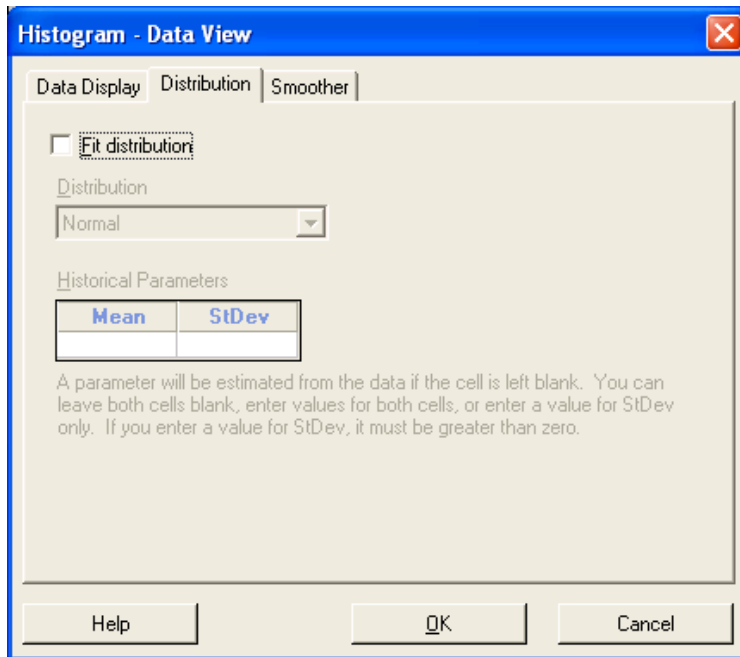
Use to fit a distribution to your histogram data to help you determine whether the data can be adequately modeled by the selected distribution.

If you click on the *Smoother* tab, this screen will appear:
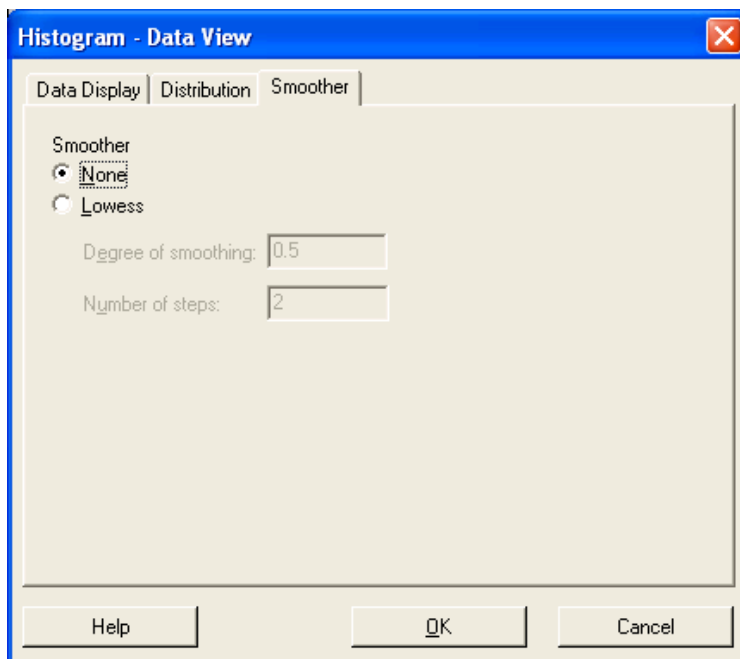
Use to fit a lowess smoother. The lowess routine fits a smoothed line to the data.

If you click on the *Multiple Graphs* button, this screen will appear:
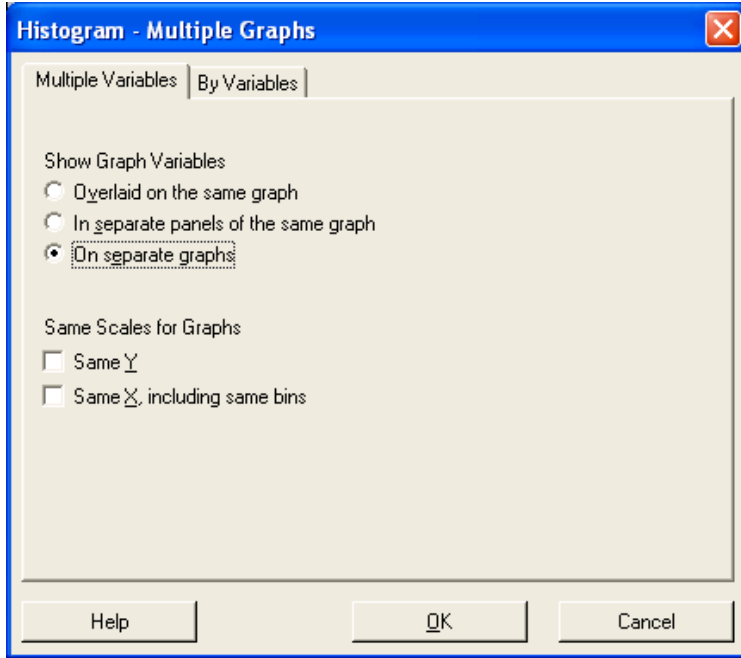
Controls the placement and scales of multiple graphs created in the same graph dialog box.

This screen shows the *Multiple Variables* tab. If you click on the *By Variables* tab, this screen will appear:
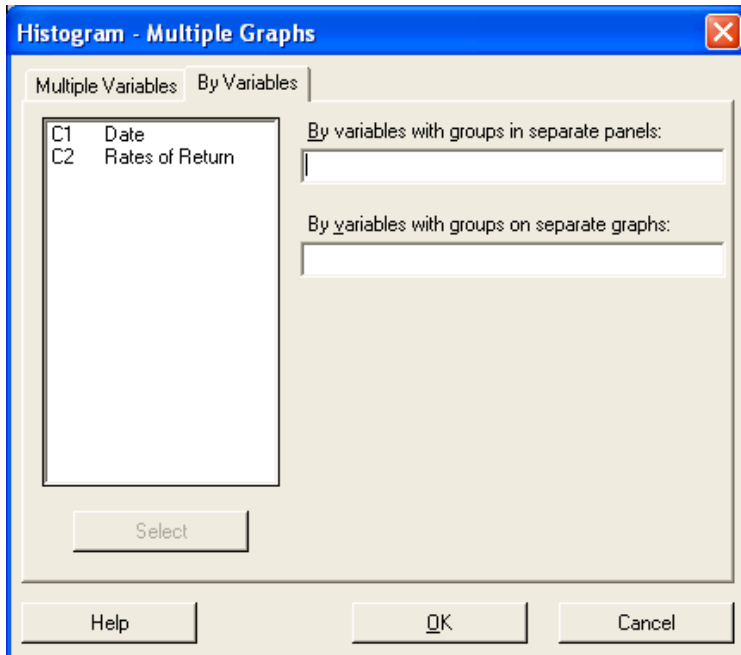
Creates multiple graphs based on a type of grouping variable called a By variable, which divides the data into distinct graphs, either in separate panels or on separate pages. Separating a graph into panels is useful when the number of groups would make an overlaid graph difficult to read.

If you click on the **Data Options** button, this screen will appear:
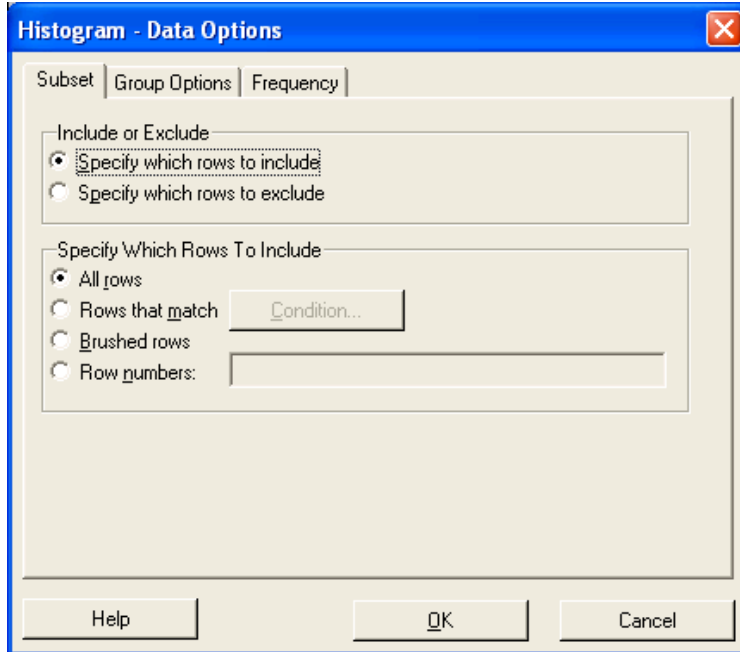
Use to include or exclude rows when creating a graph.

This screen shows the *Subset* tab. If you click on the *Group Options* tab, this screen will appear:
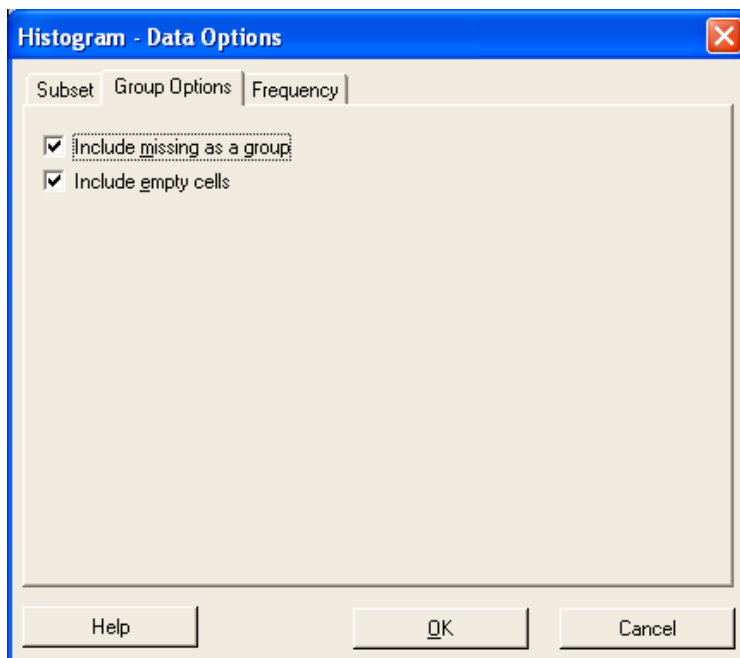
Use to include or exclude unrepresented groups and observations with missing group values when creating a graph.

If you click on the *Frequency* tab, this screen will appear:
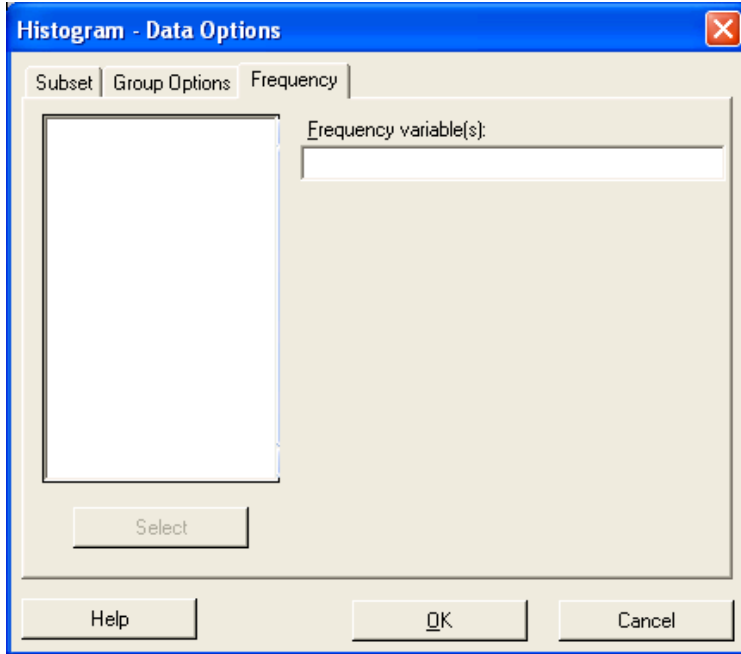
To display the graph, click on **OK.** The histogram will display:
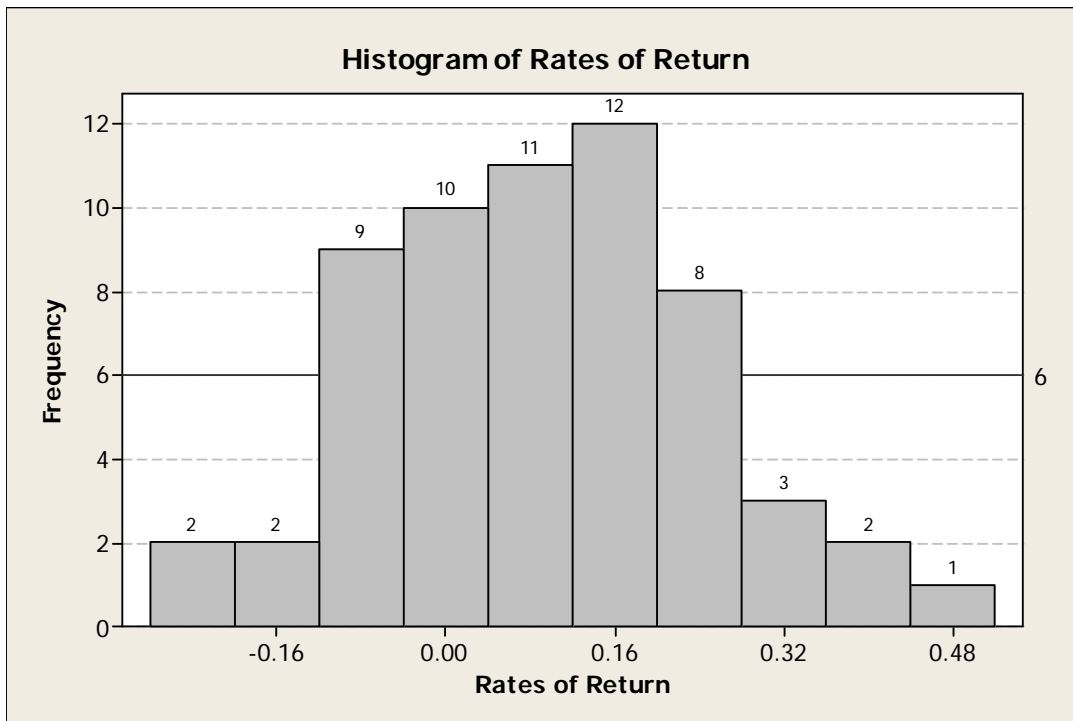
b)   *Dotplot*

Graph → **Dotplot** → Then this screen will appear:
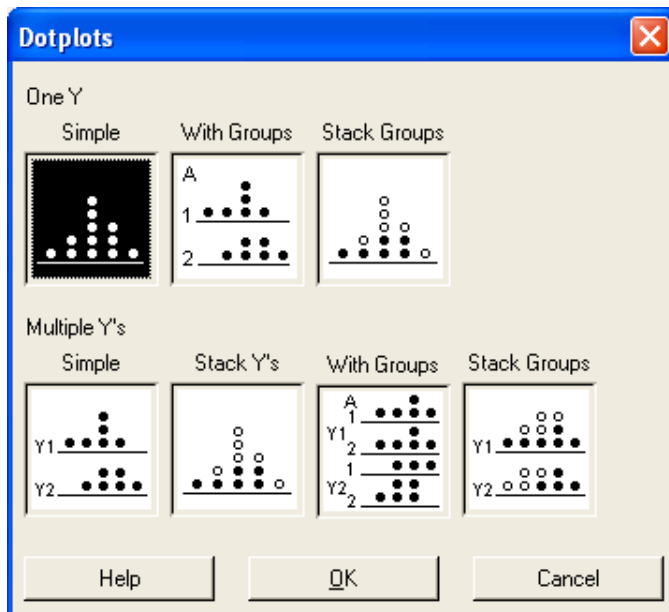
Click on appropriate graph and then click **OK.** (For this example, we will display the simple dotplot). → **Double-click on appropriate variable** (For Dell Data, double-click on Rates of Return so that it is displayed under Graph Variables) → **OK.**
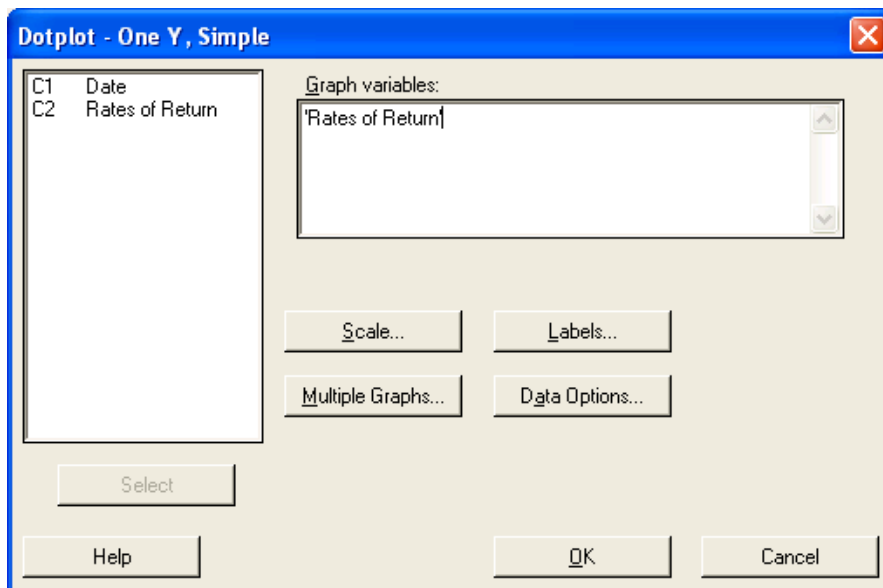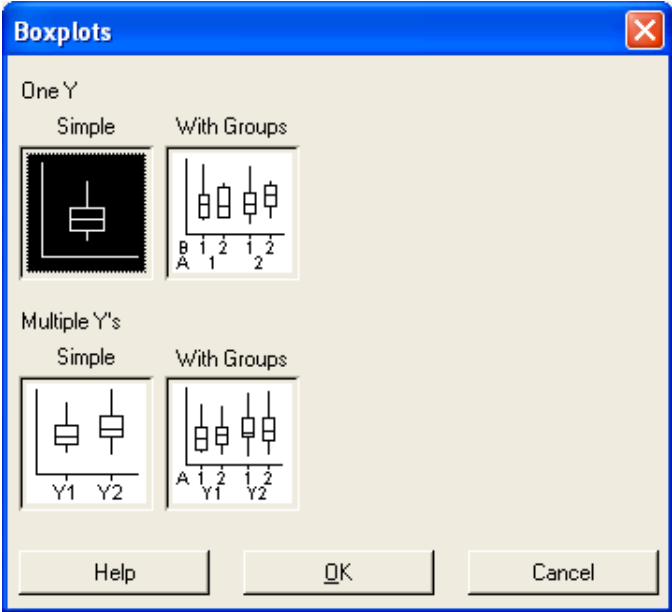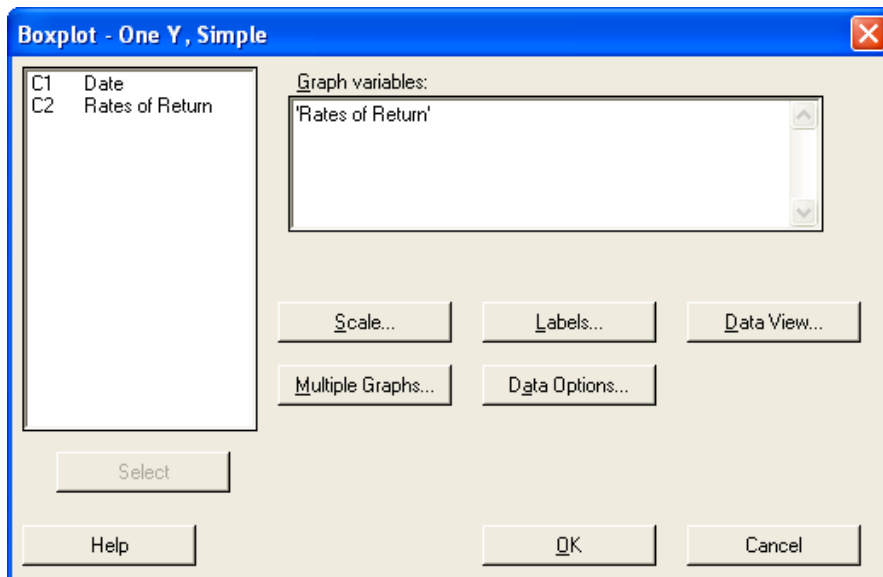
This Dotplot will display:
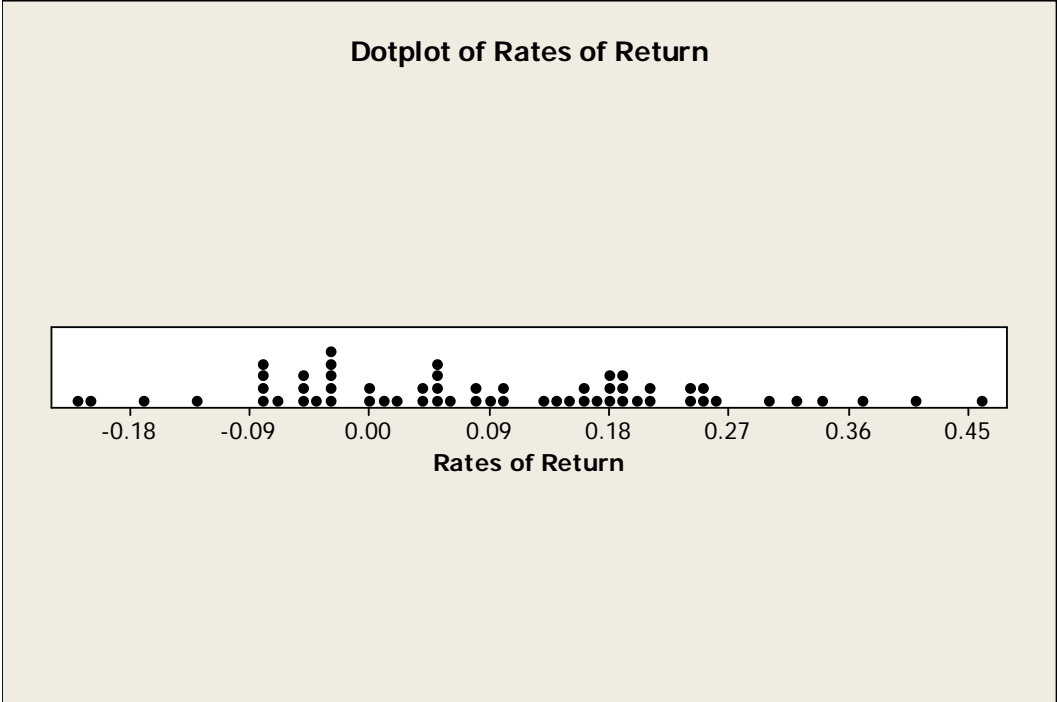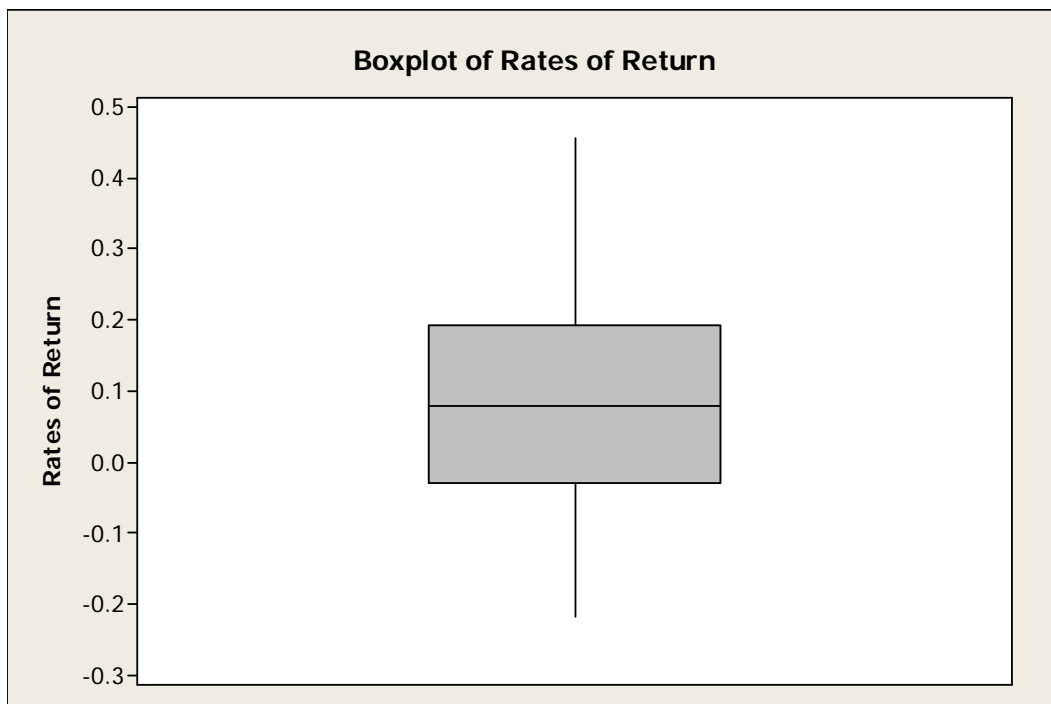

c) *Box-and-Whisker Plot (Boxplot)*

Graph → Boxplot → Then this screen will appear:

Click on appropriate graph and then click **OK.** (For this example, we will display the simple boxplot). ➔ **Double-click on appropriate variable** (For Dell Data, double-click on Rates of Return so that it is displayed under Graph Variables) ➔ **OK.**
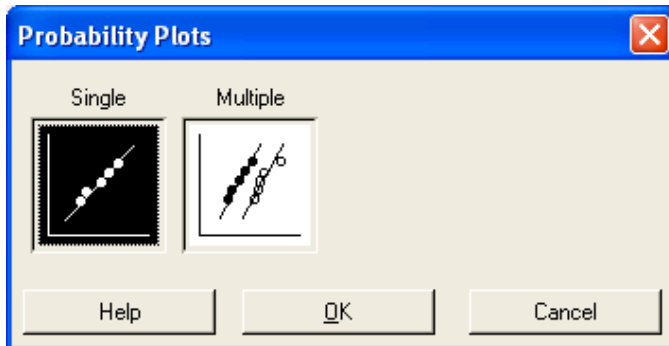

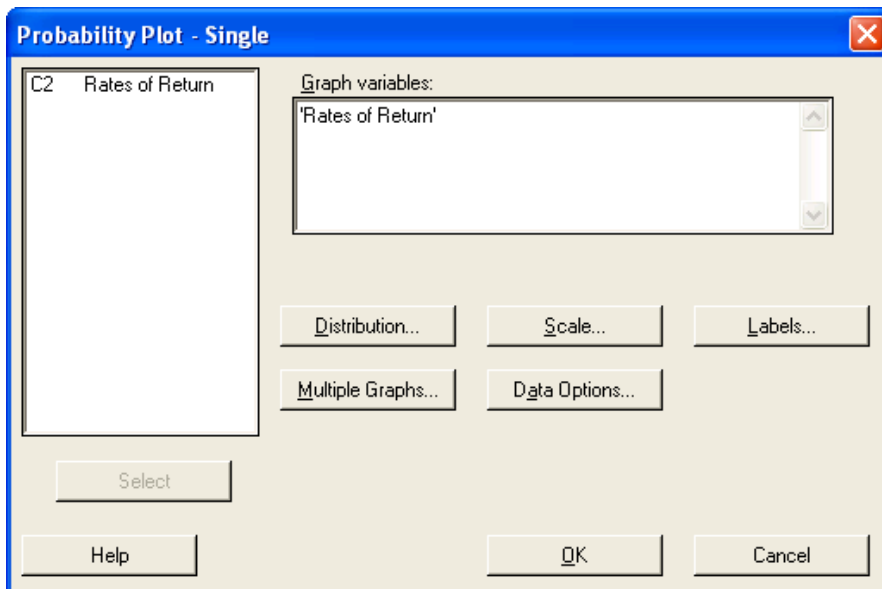This Box-and-Whisker Plot (Boxplot) will display:
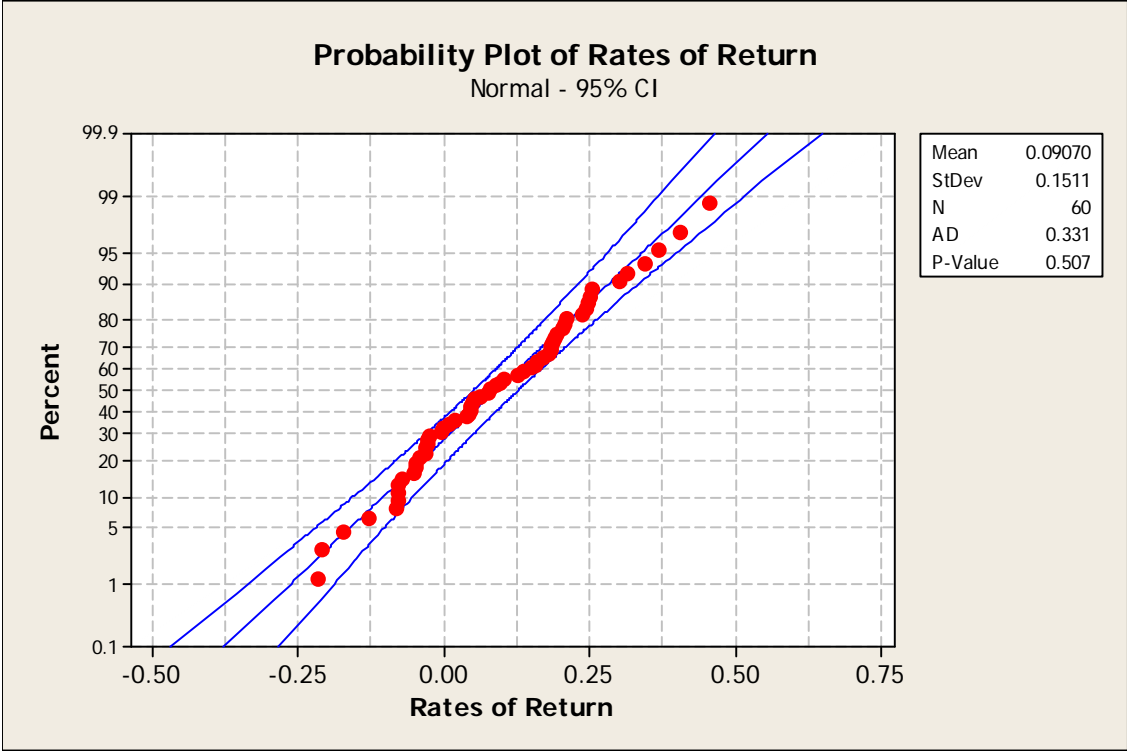
d)   Probability Plot

Graph → Probability Plot → This screen will appear:

Click on appropriate graph and then click **OK.** (For this example, we will display the single probability plot). **→ Double-click on appropriate variable** (For Dell Data, double-click on Rates of Return so that it is displayed under Graph Variables) **→ OK.**
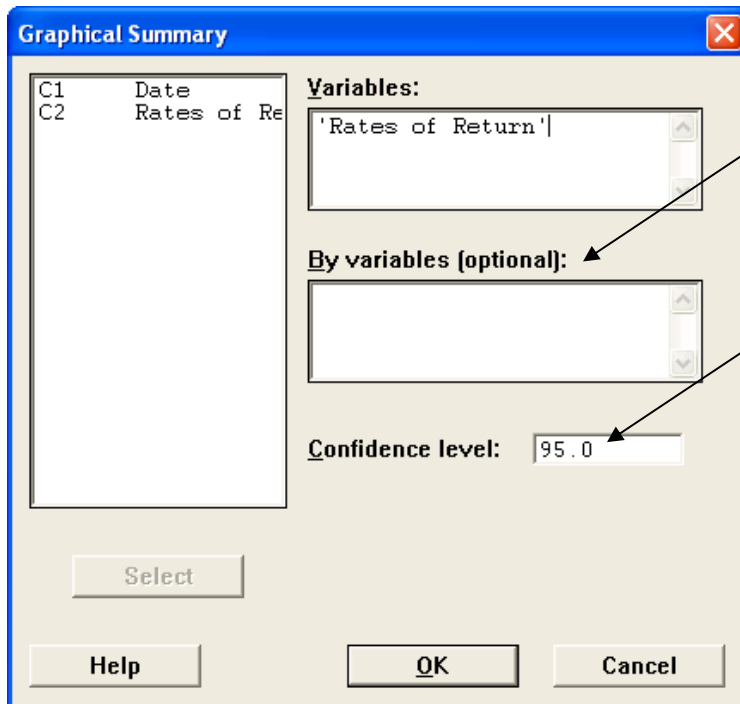
This Probability Plot will display:
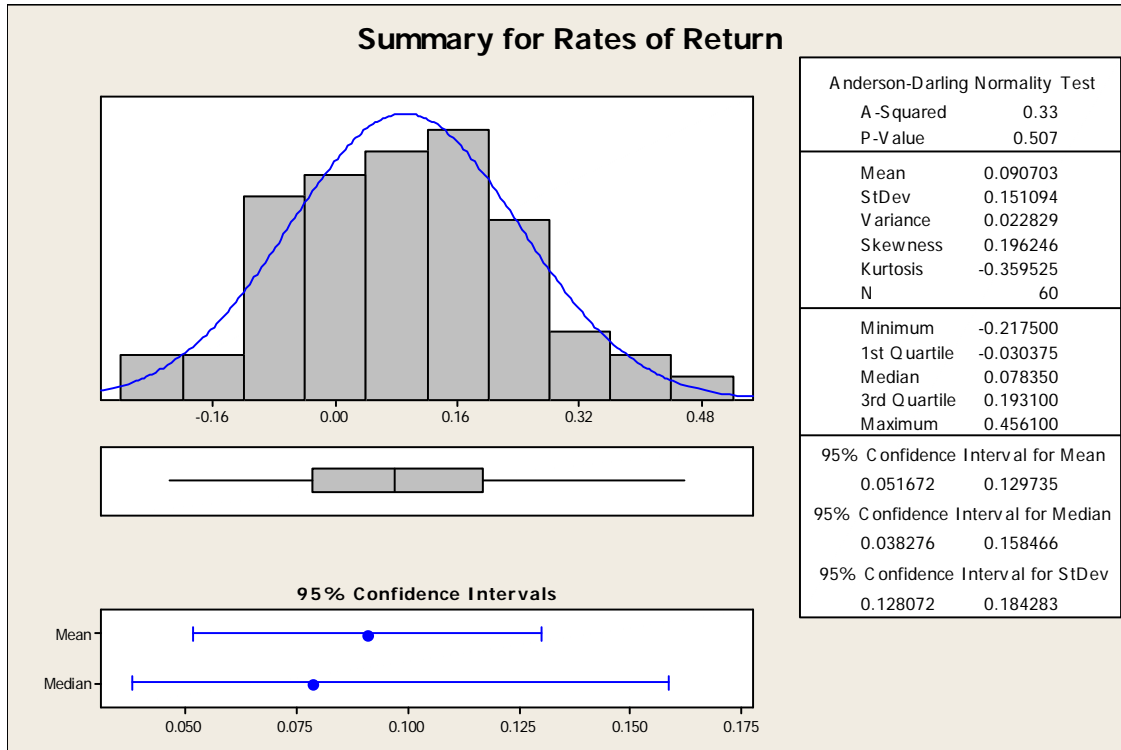
e)   *Graphical Summary*

Stat → Basic Statistics → Graphical Summary → Double-click on appropriate variable
(For Dell Data, double-click on Rates of Return so that it is displayed under Variables) **→ OK.**


See "Note" on the next page for an example using the By variables option.

The default for the confidence level is 95, but you can change it by typing in your desired level.
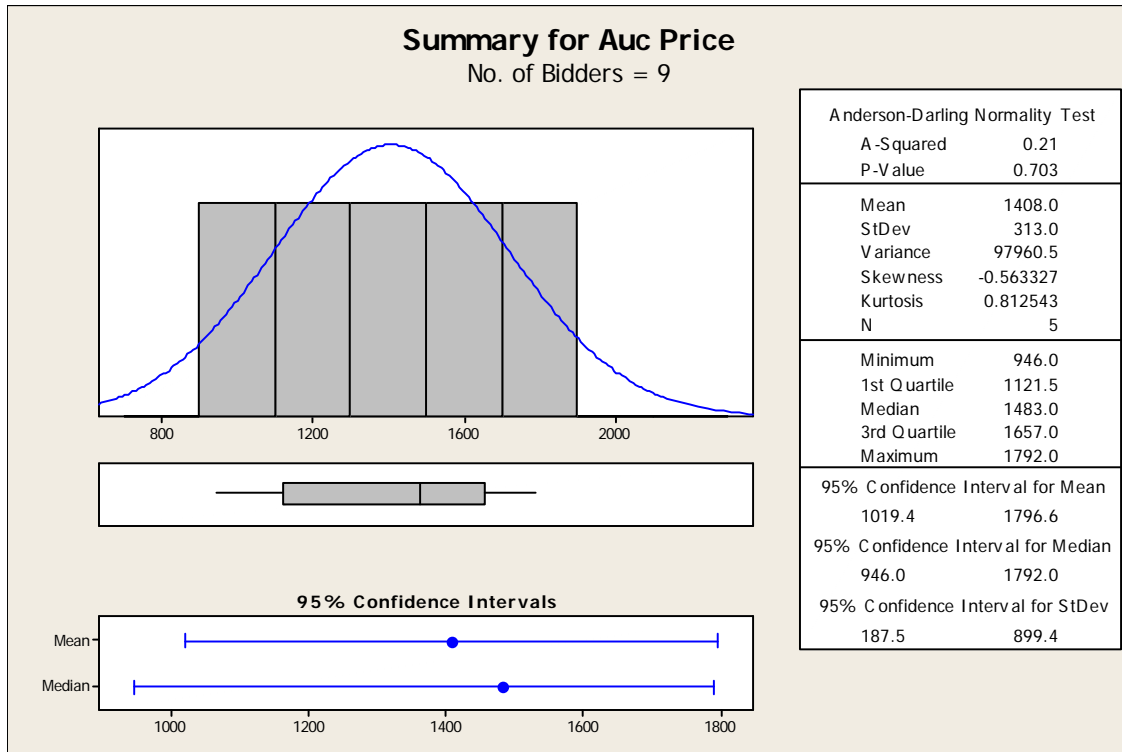
This graphical summary will display:

Summary for Rates of Return

Anderson-Darling Normality Test	
A-Squared	0.33
P-Value	0.507
Mean	0.090703
StDev	0.151094
Variance	0.022829
Skewness	0.196246
Kurtosis	-0.359525
N	60
Minimum	-0.217500
1st Quartile	-0.030375
Median	0.078350
3rd Quartile	0.193100
Maximum	0.456100
95% Confidence Interval for Mean	
0.051672	0.129735
95% Confidence Interval for Median	
0.038276	0.158466
95% Confidence Interval for StDev	
0.128072	0.184283

Note: The By variables option is used to create multiple graphical summaries based on a type of grouping variables, called a by variable. For an example using the Auction Data, if use Auc Price as the Variable and No. of Bidders as the By variable,
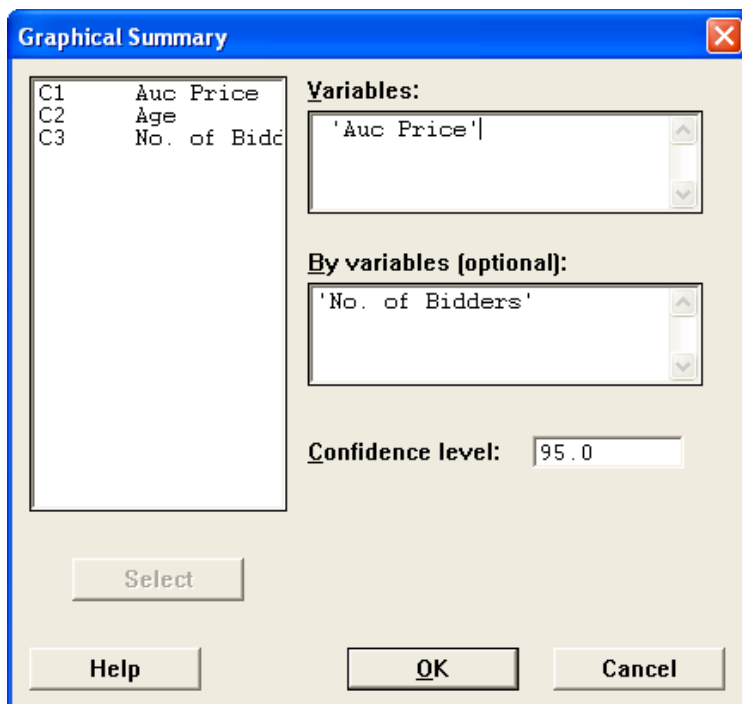
the output will display a graphical summary for every group of number of bidders. Here is one of the graphical summaries that is displayed:
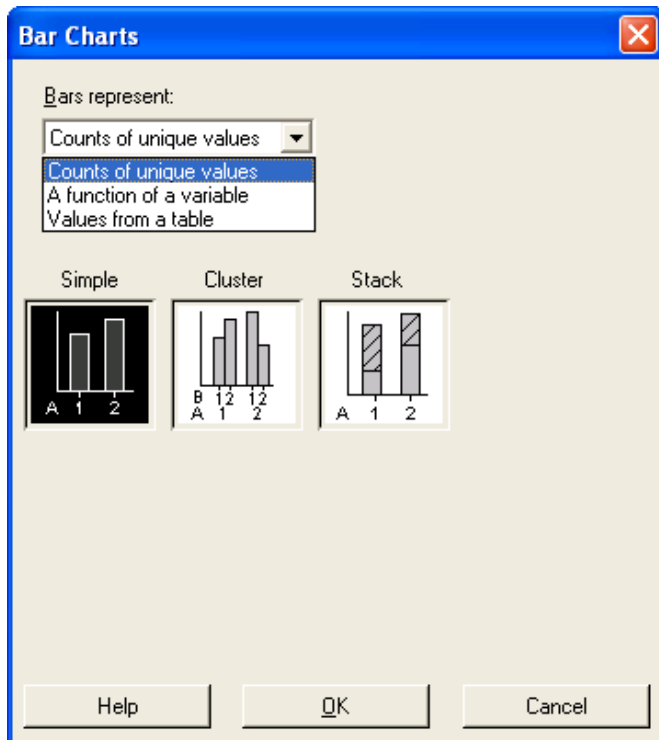

Summary for Auc Price
No. of Bidders = 9

Anderson-Darling Normality Test	
A-Squared	0.21
P-Value	0.703
Mean	1408.0
StDev	313.0
Variance	97960.5
Skewness	-0.563327
Kurtosis	0.812543
N	5
Minimum	946.0
1st Quartile	1121.5
Median	1483.0
3rd Quartile	1657.0
Maximum	1792.0

95% Confidence Interval for Mean
1019.4   1796.6
95% Confidence Interval for Median
946.0   1792.0
95% Confidence Interval for StDev
187.5   899.4

95% Confidence Intervals
Mean
Median

Thus, only the auction prices for when the number of bidders = 9 is shown.
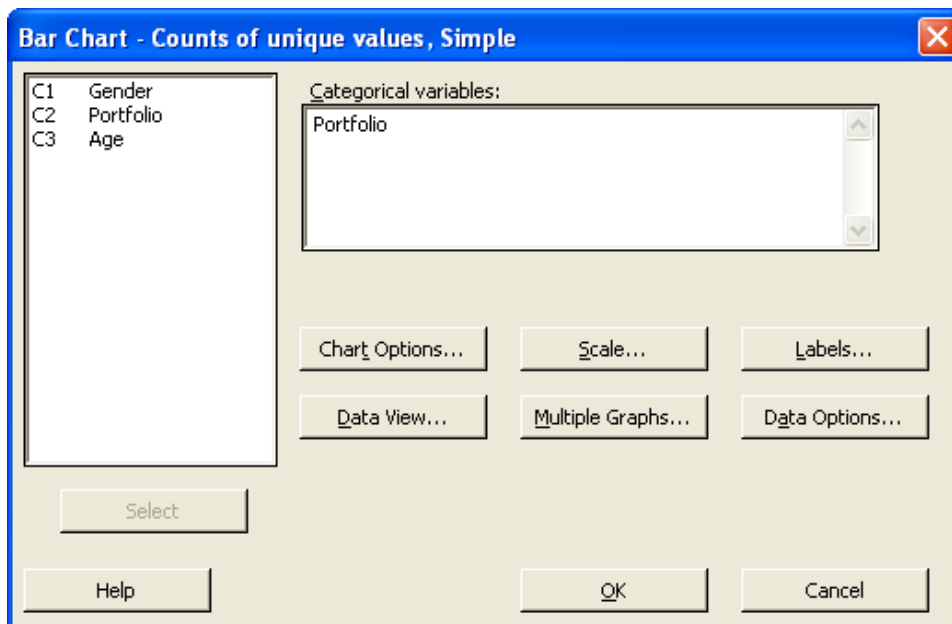
f)   Bar Chart

i)   <u>Bars representing counts of unique values</u>

Choose this graphical format if you have one or more columns of categorical data and you want to chart the frequency of each category.

Graph → Bar Chart → Choose **Counts of unique values** from the drop box and Click **OK.**
(For this example, we will use the Student Data and show a simple Bar Chart.)

Bar Charts

Bars represent:

```
Counts of unique values        ▼
```
```
Counts of unique values
A function of a variable
Values from a table
```

Simple	Cluster	Stack

Help   OK   Cancel

Double-click on the appropriate variable (Here we'll choose Portfolio).

Bar Chart - Counts of unique values, Simple

```
C1    Gender
C2    Portfolio
C3    Age
```

Categorical variables:

```
Portfolio
```

Chart Options...   Scale...   Labels...

Data View...   Multiple Graphs...   Data Options...

Select

Help   OK   Cancel

As you can see from the screen above, you are given the option to alter the output by clicking on the buttons. If you click on the *Chart Options* button, this screen will appear:
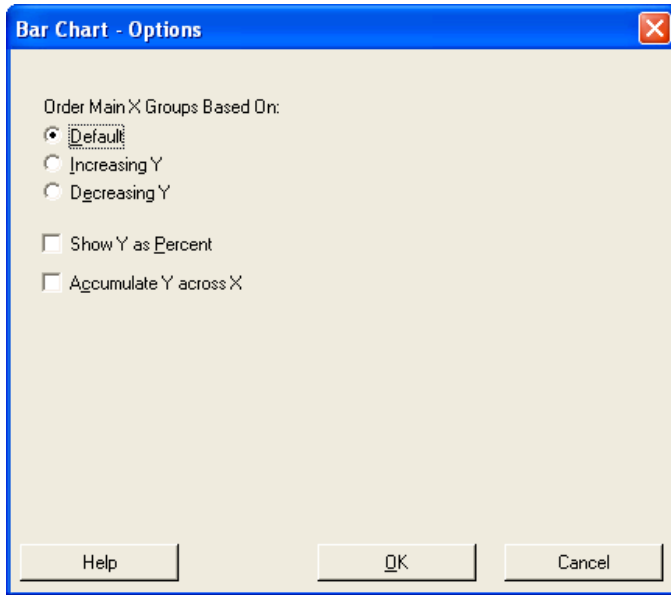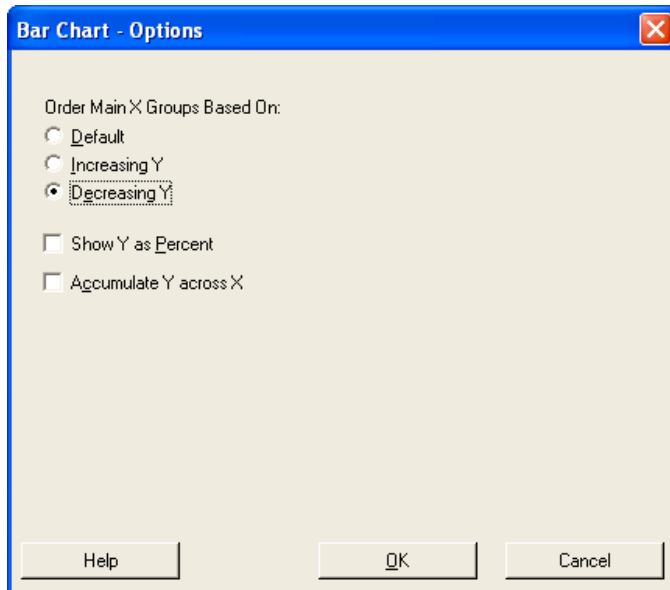
Use to specify increasing or decreasing bar order, choose a cumulative y-scale, or choose a percent scale for the y-axis.
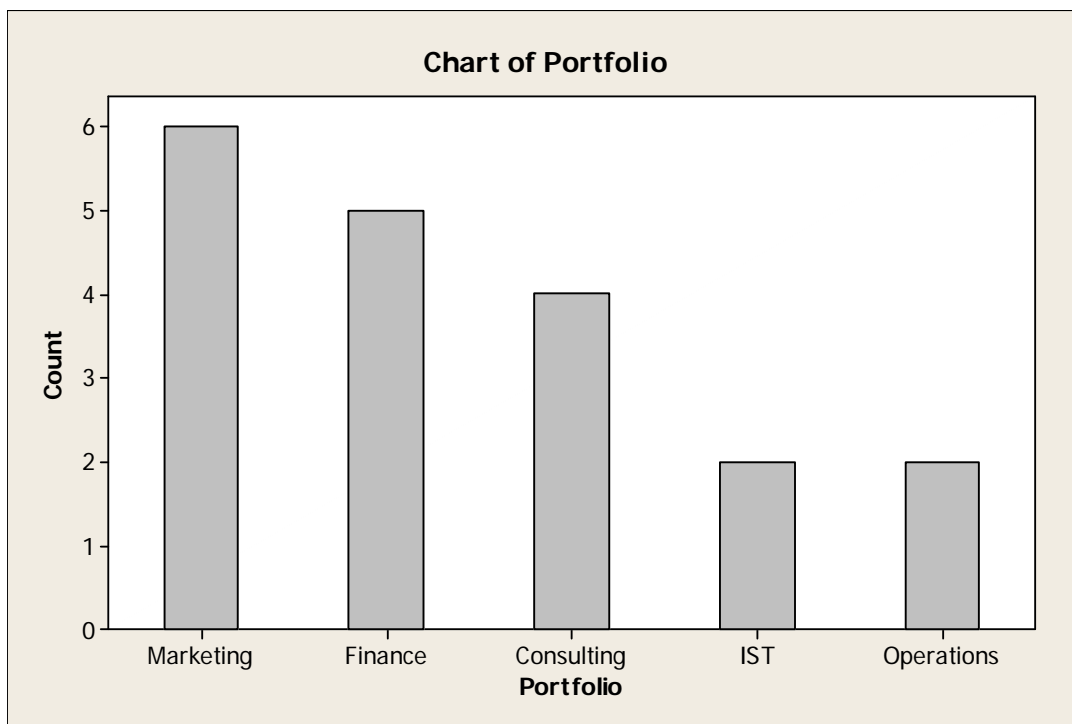
To display the bar chart, click on **OK.**
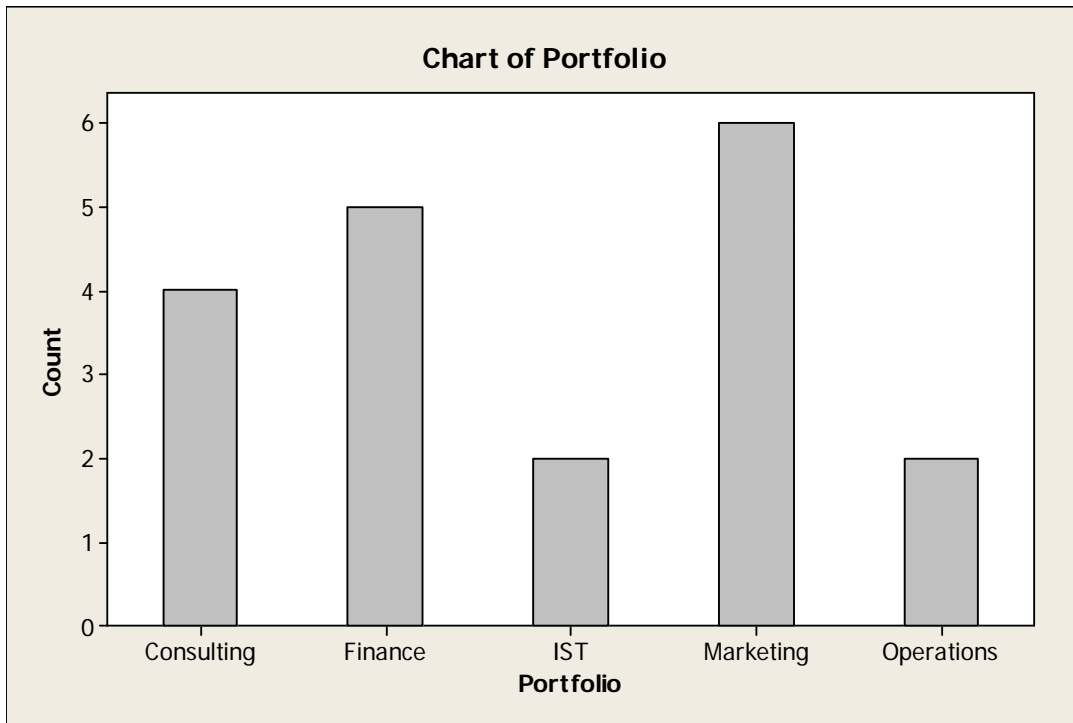
If we would have chosen Decreasing Y instead of Default after clicking on the *Bar Chart Options* button,

Then the bar chart would have looked like this:

ii)   Bars representing a function of a variable

Choose if you have one or more columns of data and you want to chart a function of the data. Quite a few of these functions are summary statistics.

Graph → Bar Chart → Choose **A function of a variable** from the drop box (Then, for this example, we will click on Cluster under Multiple Y's) and Click **OK.**
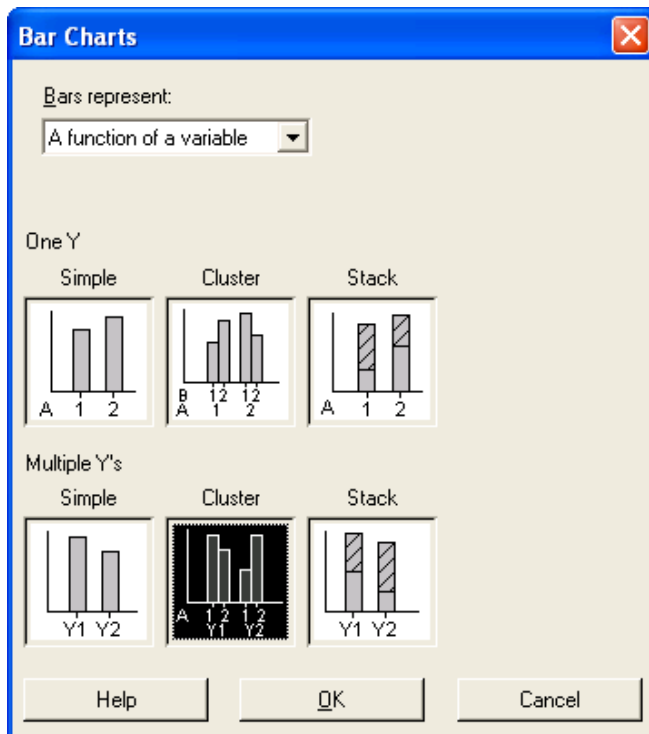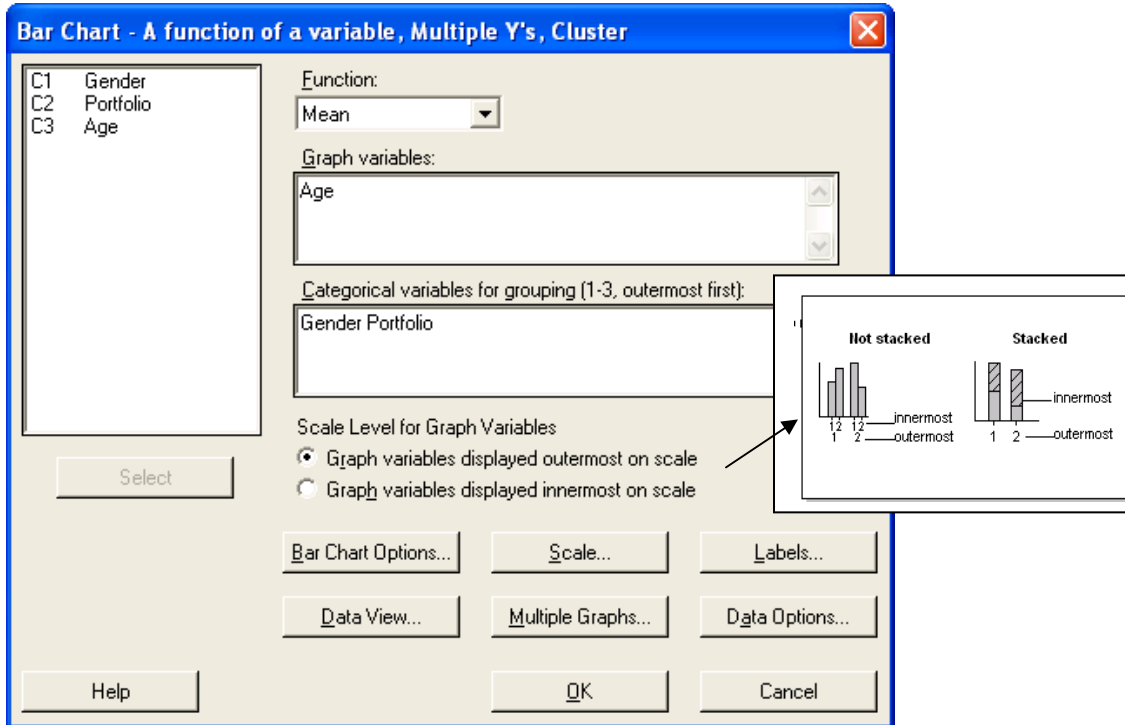
Click on appropriate variable from the drop box to choose a function. (Here we'll choose mean) →
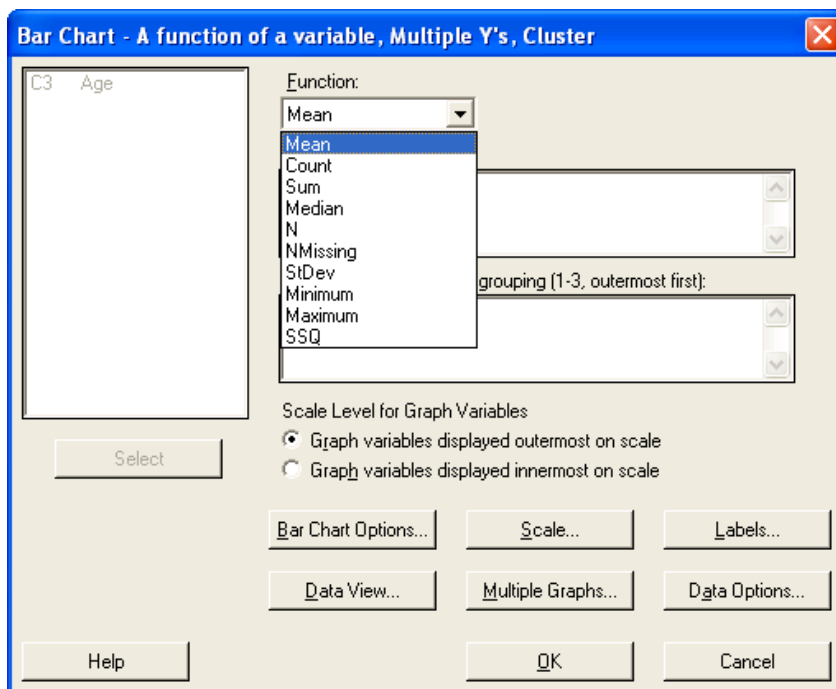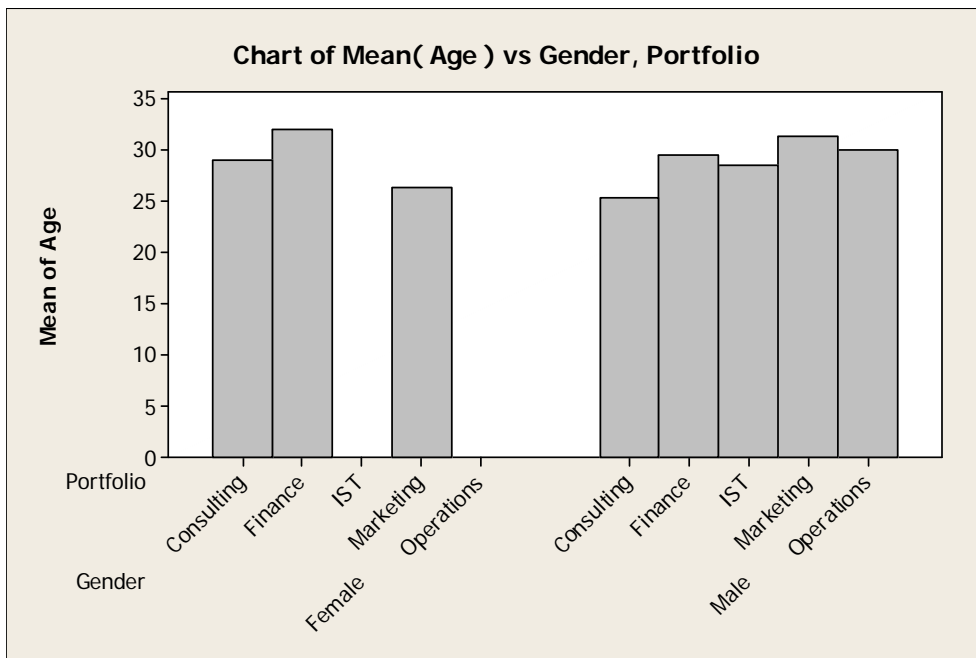

Double-click on appropriate variable in Graph variables box and then **double-click on appropriate variable in the Categorical variables for grouping box.** (For Student Data, age

was put under the graph variable and gender and portfolio was put under the categorical variables.) → **OK.**
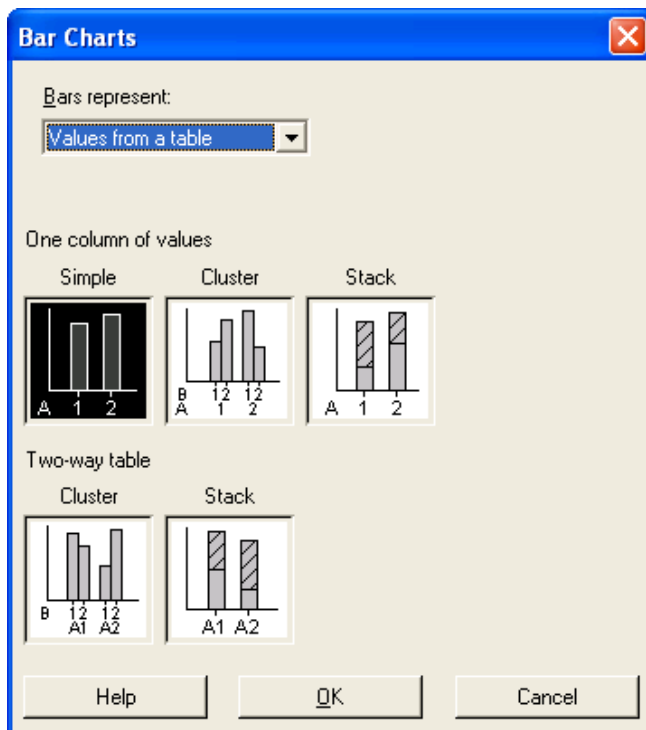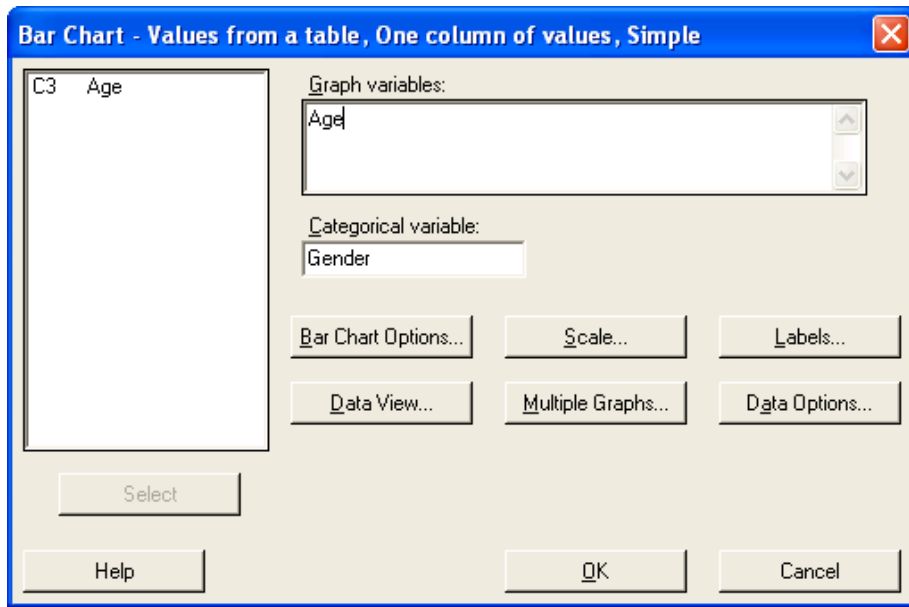

The bar chart will display:

iii)   Bars representing values from a table

Choose if you have one or more columns of summary data and you want to chart the summary value for each category.

Graph → Bar Chart → Choose **Values from a table** from the drop box (Then, for this example, we will click on Simple under One column of values) and Click **OK.**

Double-click on appropriate variable in Graph variables box and then **double-click on appropriate variable in the Categorical variable.** (For Student Data, age was put under the graph variable and gender was put under the categorical variable.) **→ OK.**

The bar chart will display:
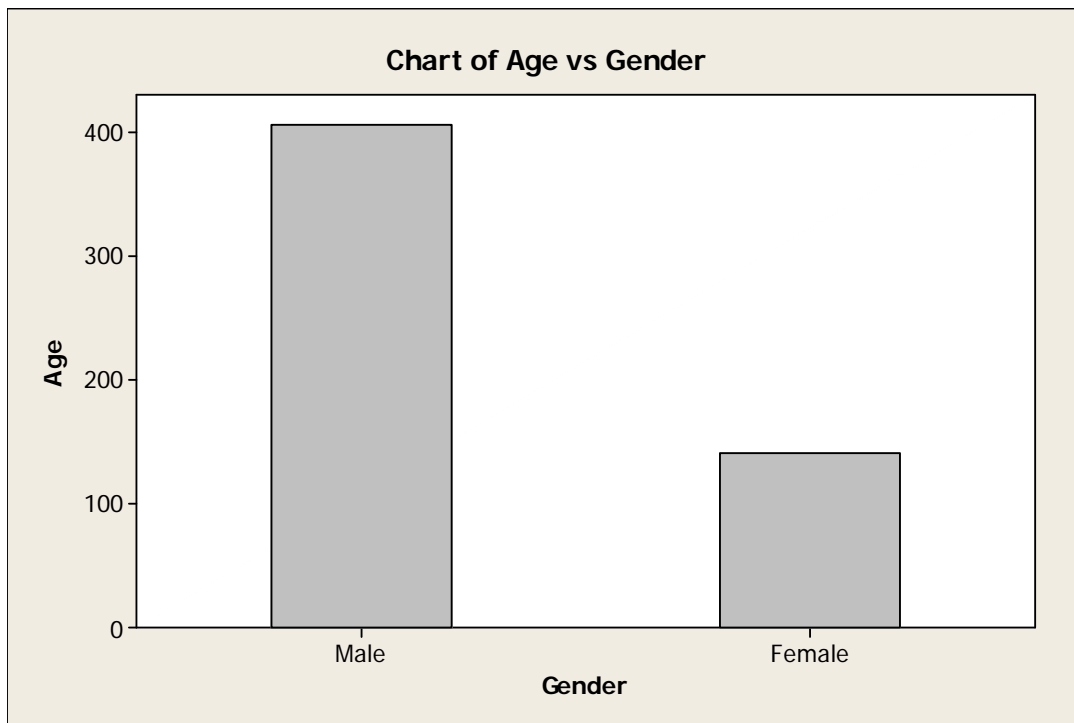
Although it does not provide much use in context to sum the ages of males versus females, this example was completed to showcase the use of this function.
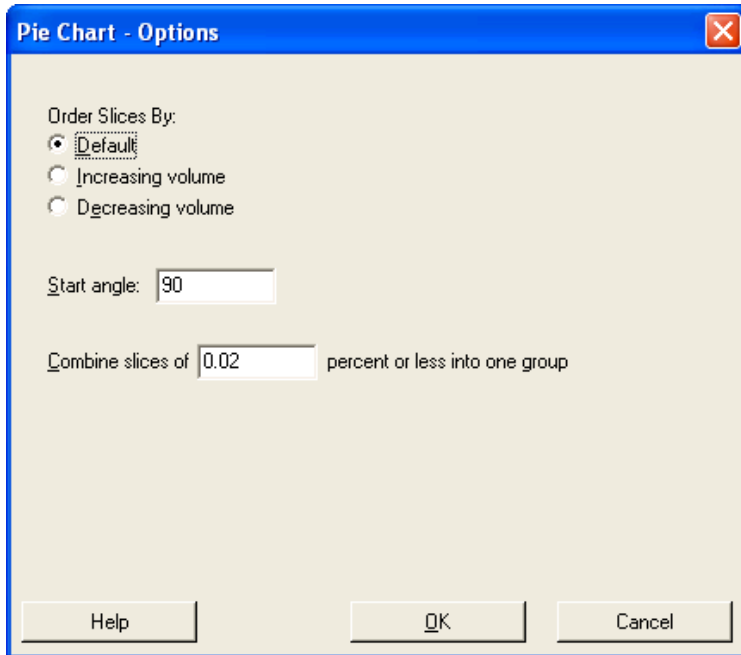
g)   *Pie Chart*

i)   <u>Chart raw data</u>

Choose when each row in a column represents a single observation. Each slice in the pie is proportional to the number of occurrences of a value in the column.

Graph → Pie Chart → Click on Chart raw data → Double-click on appropriate variable in Categorical variables box (For Student Data, double-click on Portfolio).
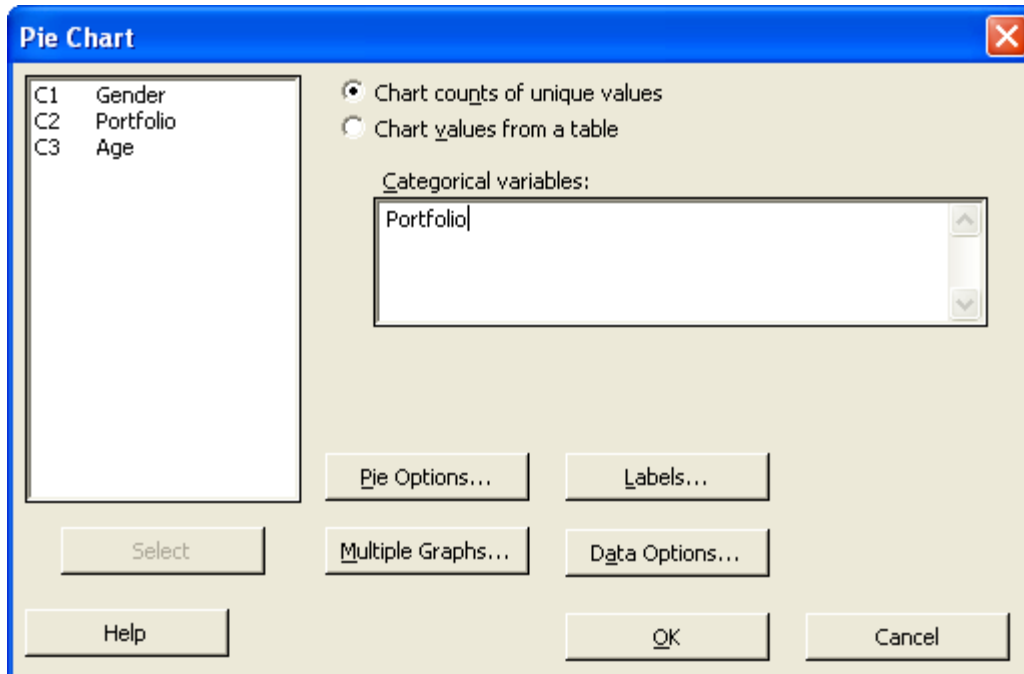
As you can see from the screen above, you are given the option to alter the output by clicking on the buttons. If you click on the *Pie Options* button, this screen will appear:

To display the pie chart, click on **OK.**
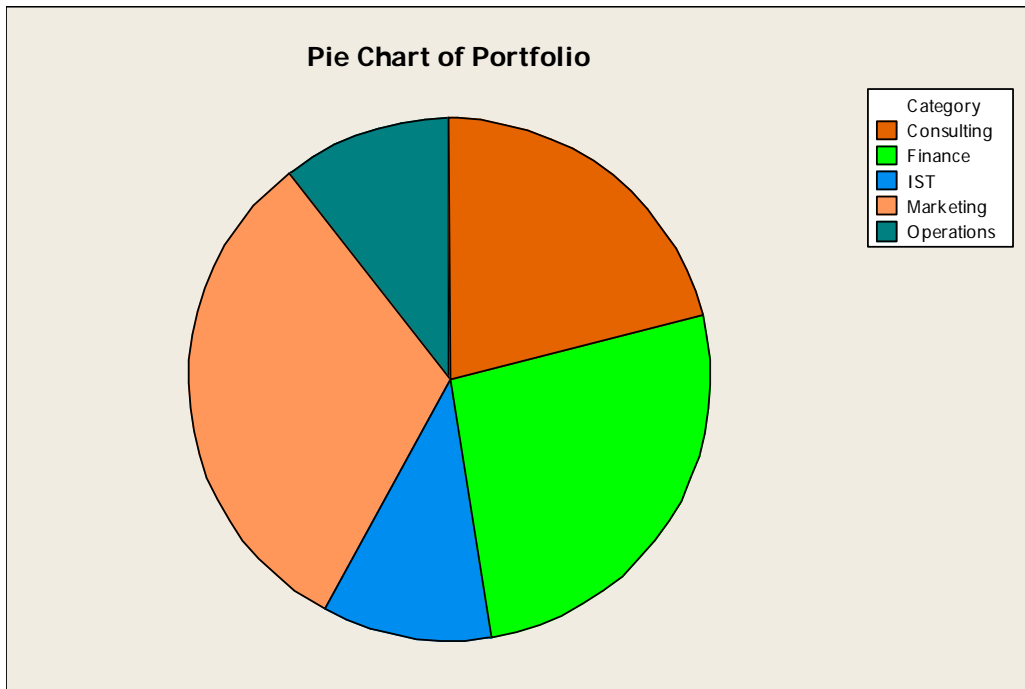
ii)   Chart values from a table

Choose when the category names are in one column and summary data are in another column.

Let's look at how to use a pie chart if our data was organized differently. (Look at Student (2) Data)

Graph → Pie Chart → Click on Chart values from a table → Double-click on appropriate variable in Categorical variable box and **double-click on appropriate variable in the Summary variables box.** (For Student (2) Data, double-click Gender for Categorical variable and Count for Summary variables.)
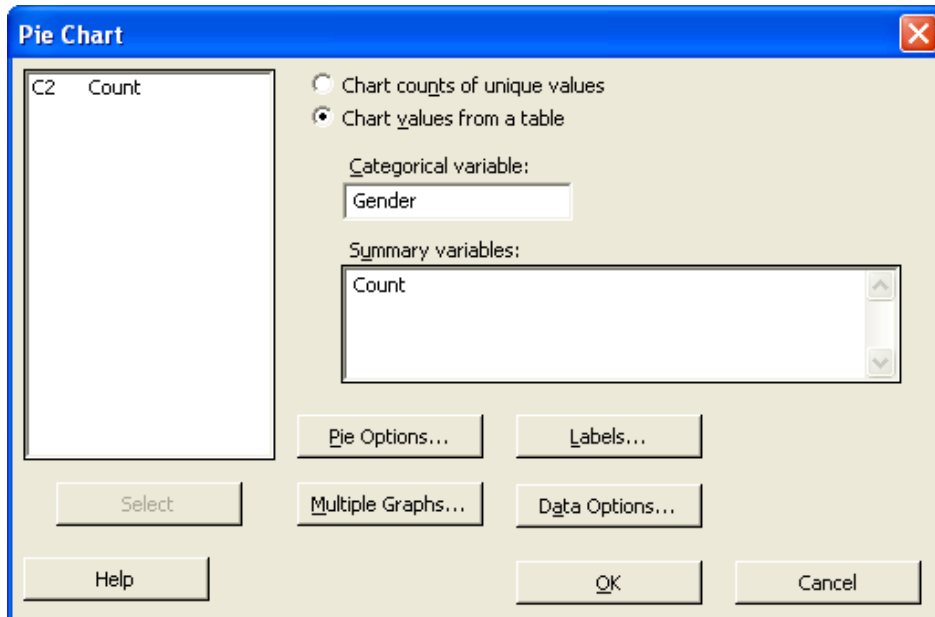
To display the pie chart, click on **OK.**
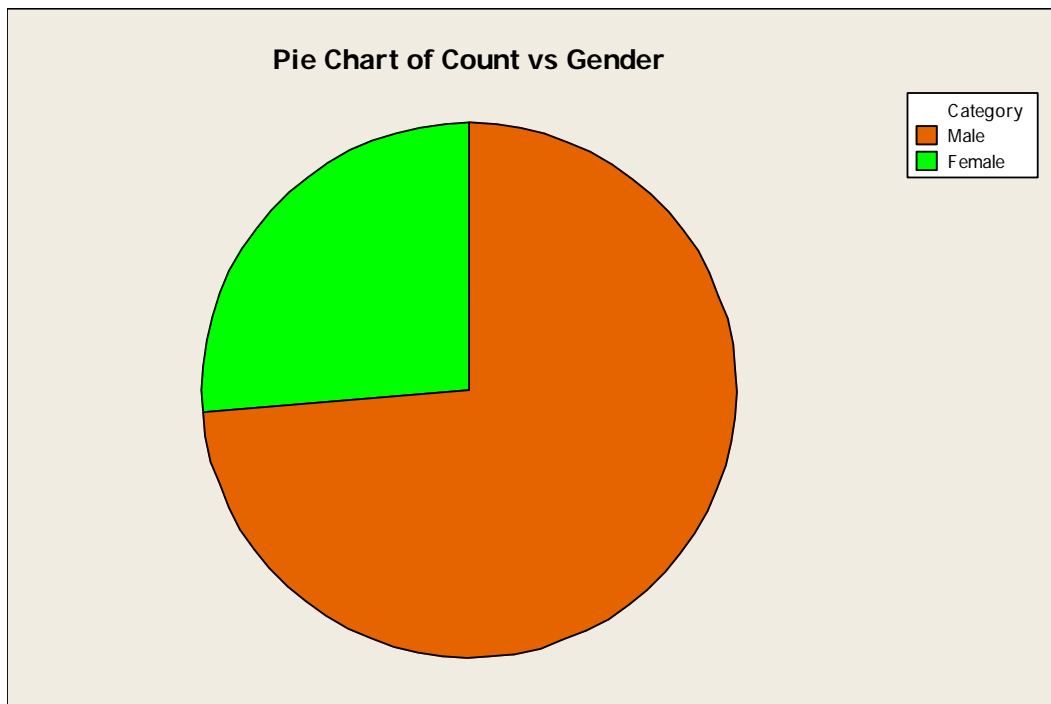
V.   Estimation and Testing

a)   1Z

Use 1-Sample Z to compute a confidence interval or perform a hypothesis test of the mean when σ is known. For a two-tailed one-sample Z:

$H_0: \mu = \mu_0$ versus $H_1: \mu \neq \mu_0$

where μ is the population mean and μ_0 is the hypothesized population mean.

Auction Price Example: Let's say that our null hypothesis is that the Mean Age equals 130 and that the alternative hypothesis is that the Mean Age does not equal 130.

Stat → Basic Statistics → 1-Sample Z → Double-click on appropriate variable (For this example, double-click on Age) **→ Type in Standard deviation in appropriate box** (Type 27.3955—Note: I used the Column Statistics function to get the standard deviation) **→ Check Perform Hypothesis test → Type in Hypothesized mean** (Type 130).

Choose if you have entered raw data in columns. Enter the columns containing the sample data.

Choose if you have summary values for the sample size, mean, and standard deviation.

If you click on the *Graphs* button, this screen will appear:
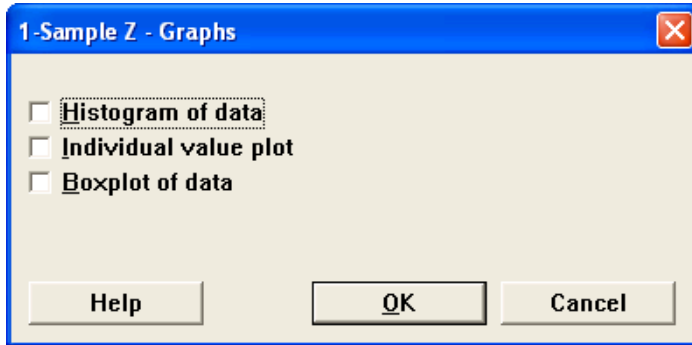
You can display the listed graphs. The graphs show the sample mean and a confidence interval for the mean. When you do a hypothesis test, the graphs also show the null hypothesis test value.

If you click on the *Options* button, this screen will appear:
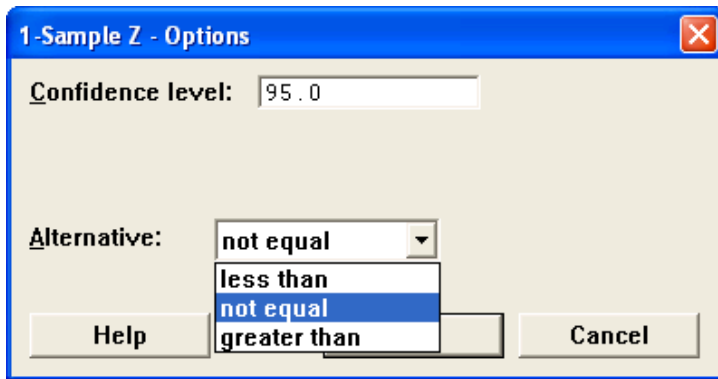
Here you can choose the confidence level and whether you want to do a one-tailed and two-tailed test.

Click **OK** to see this output display in the session window:

One-Sample Z: Age

```
Test of mu = 130 vs not = 130
The assumed standard deviation = 27.3955


Variable   N     Mean   StDev  SE Mean       95% CI          Z      P
Age        32  144.938  27.395   4.843  (135.446, 154.429)  3.08  0.002
```

Looking at the Confidence Interval, Z value, or P value, we know that we can reject the null hypothesis for our example.

b)   1t

Performs a one sample t-test or t-confidence interval for the mean.

Use 1-Sample t to compute a confidence interval and perform a hypothesis test of the mean when the population standard deviation, σ, is unknown. For a two-tailed one-sample t:

   H_0: $\mu = \mu_0$ versus H_1: $\mu \neq \mu_0$
         where μ is the population mean and μ_0 is the hypothesized population mean.

Another example using Auction Data: Let's say that our null hypothesis is that the Number of Bidders equals 9 and the alternative hypothesis is that the Number of Bidders does not equal 9.

Stat → Basic Statistics → 1-Sample t → Double-click on appropriate variable (For this example, double-click on No. of Bidders) **→ Check Perform hypothesis test → Type in Hypothesized mean** (Type 9) **→ OK.**
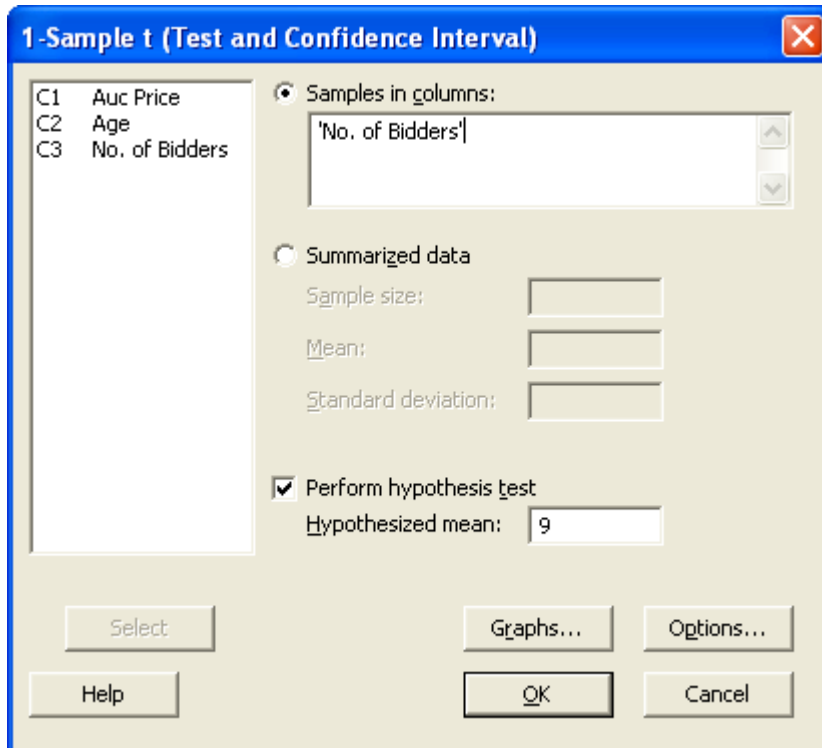
This output is displayed in the session window:

One-Sample T: No. of Bidders

```
Test of mu = 9 vs not = 9


Variable          N      Mean     StDev   SE Mean         95% CI          T      P
No. of Bidders   32   9.53125   2.83963   0.50198  (8.50745, 10.55505)   1.06   0.298
```

Looking at the Confidence Interval, Z value, or P value, we know that we cannot reject the null hypothesis in this example.

c)   *2t*

You can perform an independent two-sample t-test and generate a confidence interval.

Use 2-Sample t to perform a hypothesis test and compute a confidence interval of the difference between two population means when the population standard deviations, σ's, are unknown. For a two-tailed two-sample t:

$H_0: \mu_1 - \mu_2 = \delta_0$ versus $H_1: \mu_1 - \mu_2 \neq \delta_0$

where μ_1 and μ_2 are the population means and δ_0 is the hypothesized difference between the two population means.

Student Data Example: Let's say that our null hypothesis is that the Mean Age of Females minus the Mean Age of Males is zero and the alternative hypothesis is that it doesn't equal zero.

Stat → Basic Statistics → 2-Sample t → Click by appropriate button (For Student Data, click by Samples in one column) **→ Double-click on appropriate variables** (Double-click on Age so it is displayed by Samples, and double-click on Gender so it is displayed by Subscripts).

2-Sample t (Test and Confidence Interval)

C1   Gender
C2   Portfolio
C3   Age

Samples in one column
Samples: Age
Subscripts: Gender

Samples in different columns
First:
Second:

Summarized data
    Sample size: Mean: Standard deviation:
First:
Second:

Assume equal variances

Select   Check to assume that the populations have equal variances.

Help

Graphs...   Options...

OK   Cancel

Choose if the sample data are in a single column, differentiated by subscript values (group codes) in a second column.

Choose if the data of the two samples are in separate columns.

Choose if you have summary values for the sample size, mean, and standard deviation for each sample.

If you click on the *Graphs* button, this screen will appear:
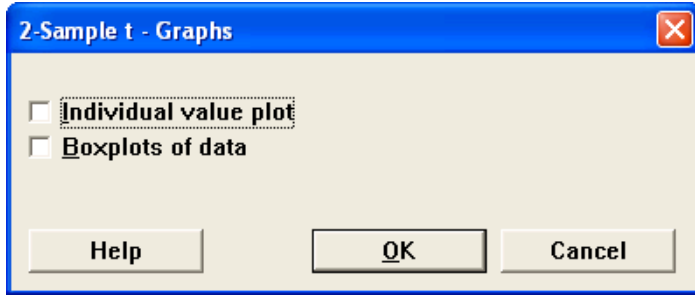
-Displays an individual value plot and a boxplot of the variables. The graphs also display the sample means.

If you click on the *Options* button, this screen will appear:
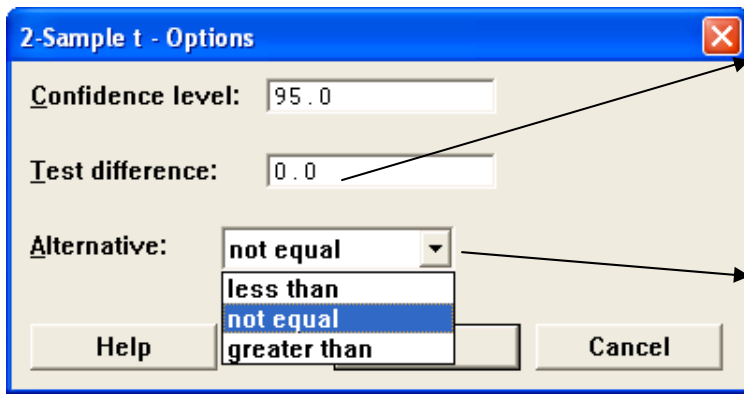
Choose desired confidence level.

Enter the null hypothesis value, which is the hypothesized difference in population means.

Choose whether you wan to do a one or two-tailed test.

Click **OK** to see this output display in the session window:

Two-Sample T-Test and CI: Age, Gender

```
Two-sample T for Age

Gender    N    Mean   StDev   SE Mean
Female    5   28.00    3.00       1.3
Male     14   28.93    4.30       1.2


Difference = mu (Female) – mu (Male)
Estimate for difference:   -0.928571
95% CI for difference:   (-4.866635, 3.009492)
T-Test of difference = 0 (vs not =): T-Value = -0.53   P-Value = 0.611   DF = 10
```

Looking at the Confidence Interval, T value, or P value, we know that we cannot reject the null hypothesis.

d)   *1P*

Use to perform a test of one binomial (two categories) proportion.

Use 1 Proportion to compute a confidence interval and perform a hypothesis test of the proportion.:

H_o: $p = p_o$ versus H_a: $p \neq p_o$

where p is the population proportion and p_o is the hypothesized value.

Another Student Data Example: Let's say that our null hypothesis is that the proportion of Male students equals .5 and the alternative hypothesis is that the proportion of Male students does not equal .5.

Stat → Basic Statistics → 1 Proportion → Double click on appropriate variable (For Student Data, double-click on Gender.) **→ Check Perform hypothesis test → Type in Hypothesized proportion** (Type 0.5) **→ OK**

Note: Each cell of these columns must be one of two possible values and correspond to one item or subject.

Choose if you have summary values for the number of trials and successes.

If you click on the *Options* button, this screen will appear:
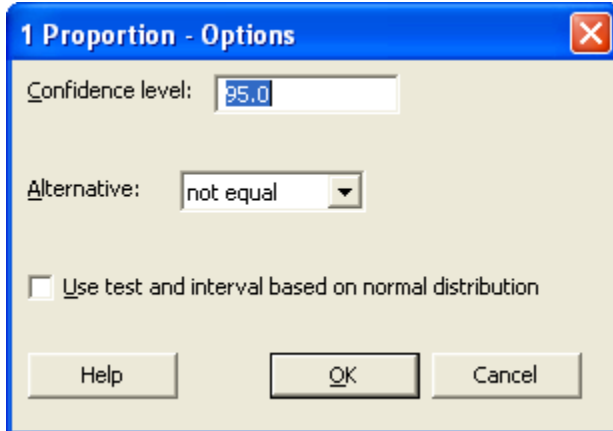
Click **OK** to see this output display in the session window:

Test and CI for One Proportion: Gender

```
Test of p = 0.5 vs p not = 0.5

Event = Male

                                       Exact
Variable   X   N  Sample p       95% CI        P-Value
Gender    14  19  0.736842  (0.487971, 0.908534)   0.064
```

Looking at the Confidence Interval or P value, we know that we cannot reject the null hypothesis. (Note: The range of the confidence interval is relatively large because the sample size is small.)

e)   2P

Use to perform a test of two binomial proportions.

Use the 2 Proportions command to compute a confidence interval and perform a hypothesis test of the difference between two proportions. For a two-tailed test of two proportions:

H_0: p_1 - p_2 = p_0 versus H_1: p_1 - p_2 ≠ p_0

where p_1 and p_2 are the proportions of success in populations 1 and 2, respectively,

and p_0 is the hypothesized difference between the two proportions.

Another Student Data Example: Let's say that our null hypothesis is that the proportion of Male students with a portfolio in Consulting minus the proportion of Female students with a portfolio in Consulting equals zero, and the alternative hypothesis is that it doesn't equal zero.

Stat → Basic Statistics → 2 Proportions → Click by appropriate button (For Student Data, click by Summarized data) **→ Double-click on appropriate variables or type in appropriate numbers** (Type in 14,3,5,1) **→ OK.**
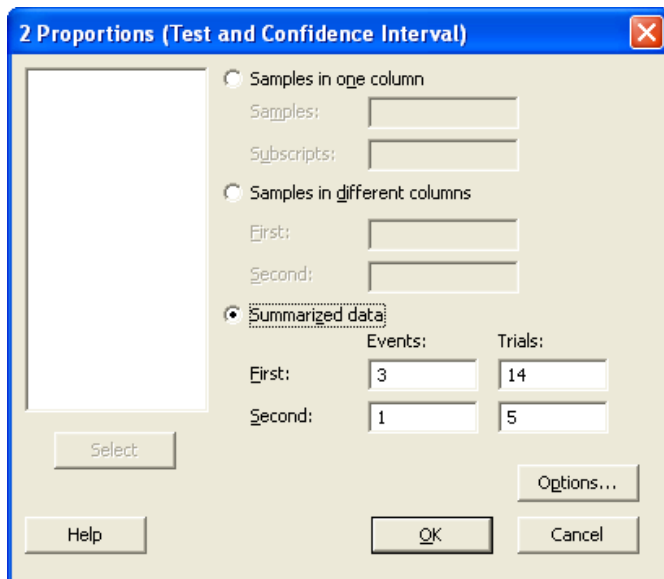
Since we don't have two columns where each cell of these columns is one of two possible values, we need to enter raw data into single columns for each sample. You can look at the sample data to get the information. Our data shows up that 3 out of 14 Males have a Consulting portfolio, and 1 out of 5 Females have a Consulting portfolio.

This output will be displayed in the session window:

Test and CI for Two Proportions

```
Sample   X    N   Sample p
1        3   14   0.214286
2        1    5   0.200000


Difference = p (1) - p (2)
Estimate for difference:   0.0142857
95% CI for difference:  (-0.396962, 0.425534)
Test for difference = 0 (vs not = 0):   Z = 0.07   P-Value = 0.946

Fisher's exact test: P-Value = 1.000

* NOTE * The normal approximation may be inaccurate for small samples.
```

Looking at the Confidence Interval or P value, we know that we cannot reject the null hypothesis.

f)   Normality Test

This function generates a normal probability plot and performs a hypothesis test to examine whether or not the observations follow a normal distribution. For the normality test, the hypotheses are:

H_0: data follow a normal distribution vs. H_a: data do not follow a normal distribution

Stat → Basic Statistics → Normality Test → Double-click on appropriate variable (For Auction Data, double-click on Auc Price) **→ OK.**

Normality Test

C1	Auc Price
C2	Age
C3	No. of Bidd

Variable: 'Auc Price'

Percentile Lines
- ● **None**
- ○ At **Y** values:
- ○ At **d**ata values:

Tests for Normality
- ● **Anderson-Darling**
- ○ **R**yan-Joiner   [Similar to Shapiro-Wilk]
- ○ **K**olmogorov-Smirnov

Select

Title:

Help   **OK**   **Cancel**

Click on *Help* button for info on Percentile Lines

Choose type of Normality Test

Can replace default title with your own

This graph displays:

Probability Plot of Auc Price
Normal

Mean	1327
StDev	393.5
N	32
AD	0.376
P-Value	0.391

Percent (y-axis: 99, 95, 90, 80, 70, 60, 50, 40, 30, 20, 10, 5, 1)

Auc Price (x-axis: 500, 1000, 1500, 2000, 2500)
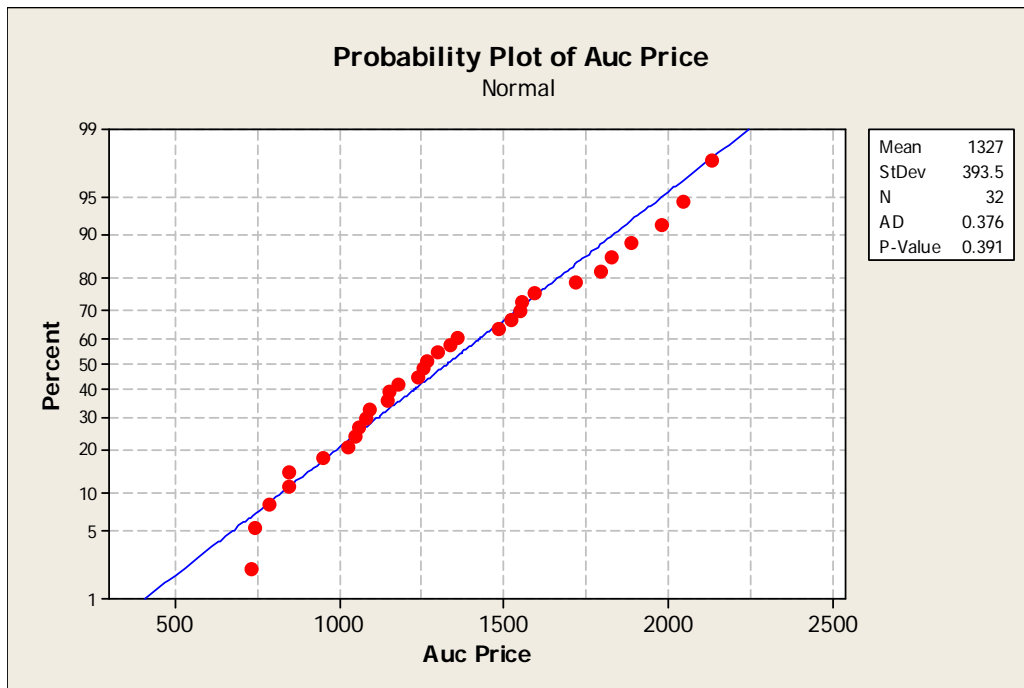
Interpreting the Results

The graphical output is a plot of normal probabilities versus the data. The Anderson-Darling test's p-value indicates that, at α levels greater than 0.391, there is evidence that the data do not follow a normal distribution. Thus, if α = .01, .05, or .10 as it typically is, we could not reject the null hypothesis in this example.
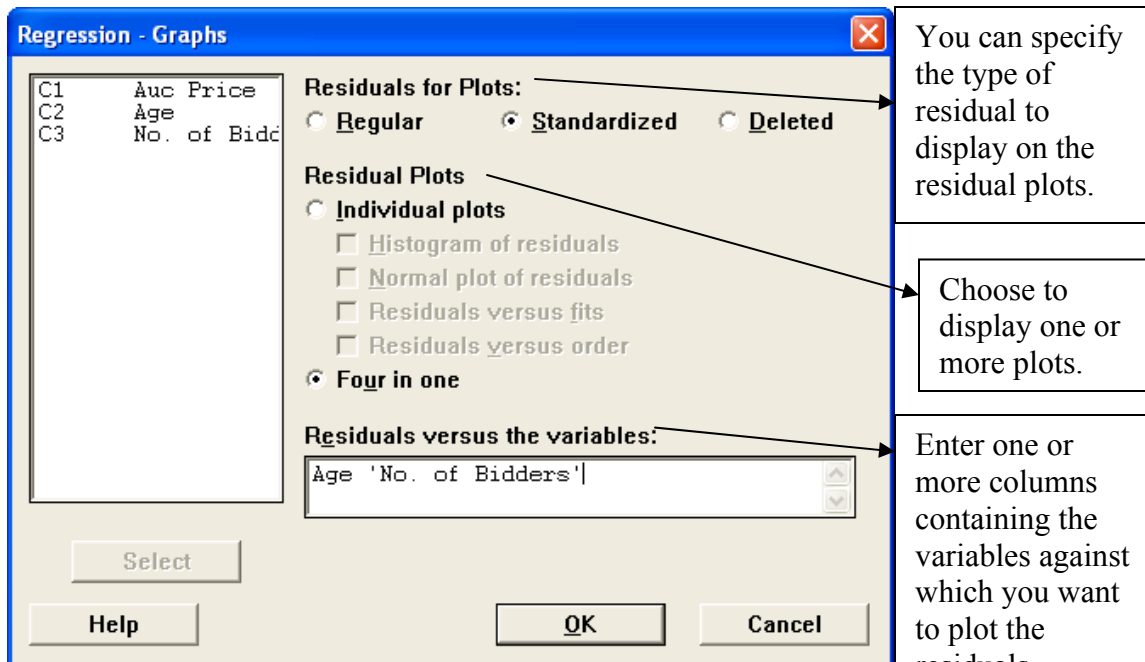
VI. Regression

a) *Finding the Best Regression Model Using All Independent Variables*

Stat → Regression → Regression → Double-click on dependent variable, then double-click on all independent variables (For The Auction Data, double-click on Auc Price so that it is displayed under Response, and then double-click on Age and No. of Bidders so that they are displayed under Predictors).
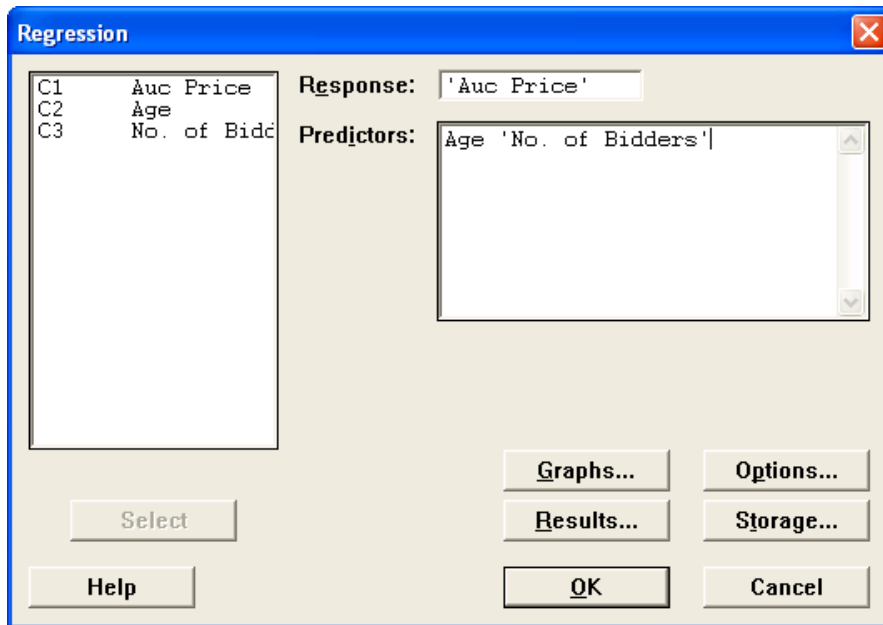
As you can see from the screen above, you are given the option to alter the output by clicking on the buttons. If you click on the *Graphs* button, this screen will appear:

You can specify the type of residual to display on the residual plots.

Choose to display one or more plots.

Enter one or more columns containing the variables against which you want to plot the residuals. Minitab displays a separate graph for each column.

Click on appropriate choice for Residuals for Plots and Residual Plots, and double-click on appropriate variables under Residuals versus the variables (See above for choices made for Auction Data in this example) → **OK.**

The output in the session window is shown:

Regression Analysis: Auc Price versus Age, No. of Bidders

```
The regression equation is
Auc Price = - 1339 + 12.7 Age + 86.0 No. of Bidders

Predictor          Coef   SE Coef       T      P
Constant        -1339.0     173.8   -7.70  0.000
Age             12.7406    0.9047   14.08  0.000
No. of Bidders   85.953     8.729    9.85  0.000

S = 133.485   R-Sq = 89.2%   R-Sq(adj) = 88.5%

Analysis of Variance
Source          DF        SS       MS       F      P
Regression       2   4283063  2141531  120.19  0.000
Residual Error  29    516727    17818
Total           31   4799790

Source          DF   Seq SS
Age              1  2555224
No. of Bidders   1  1727838
```
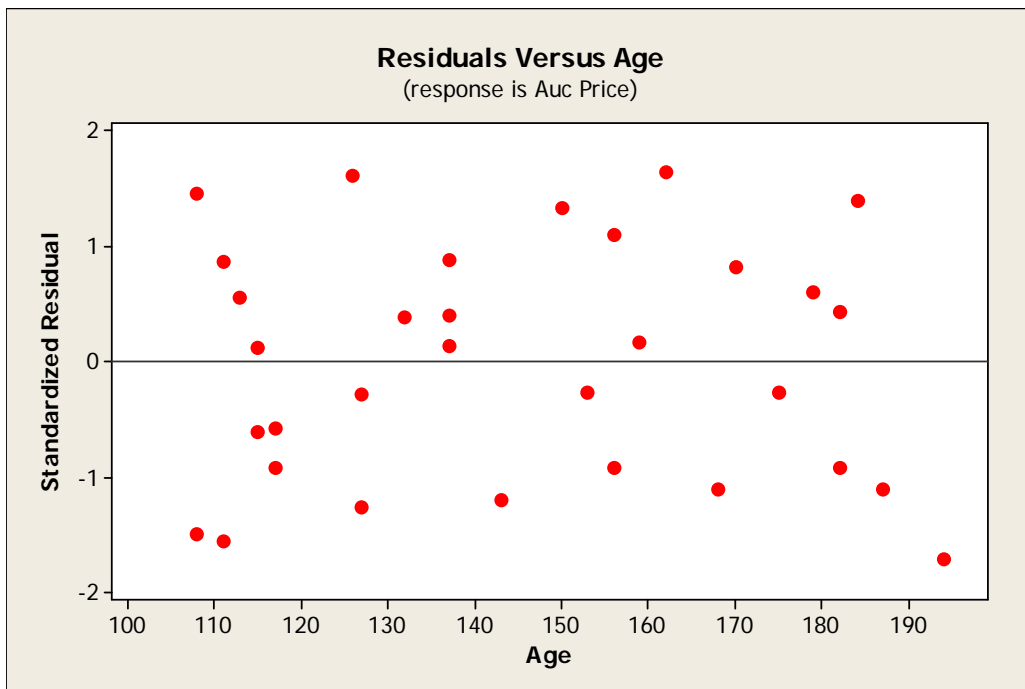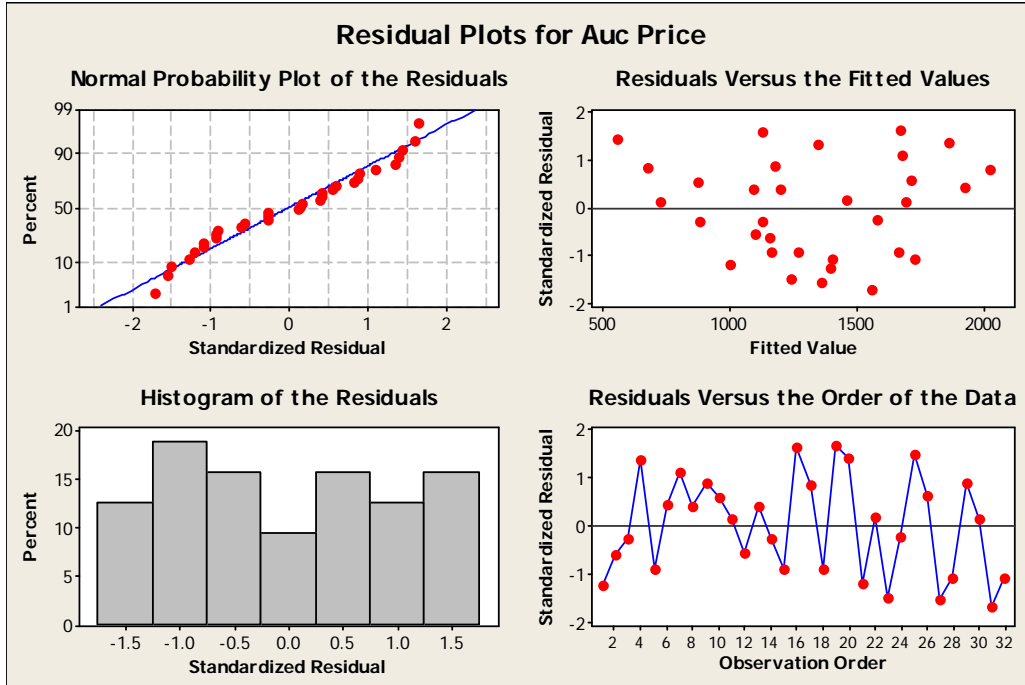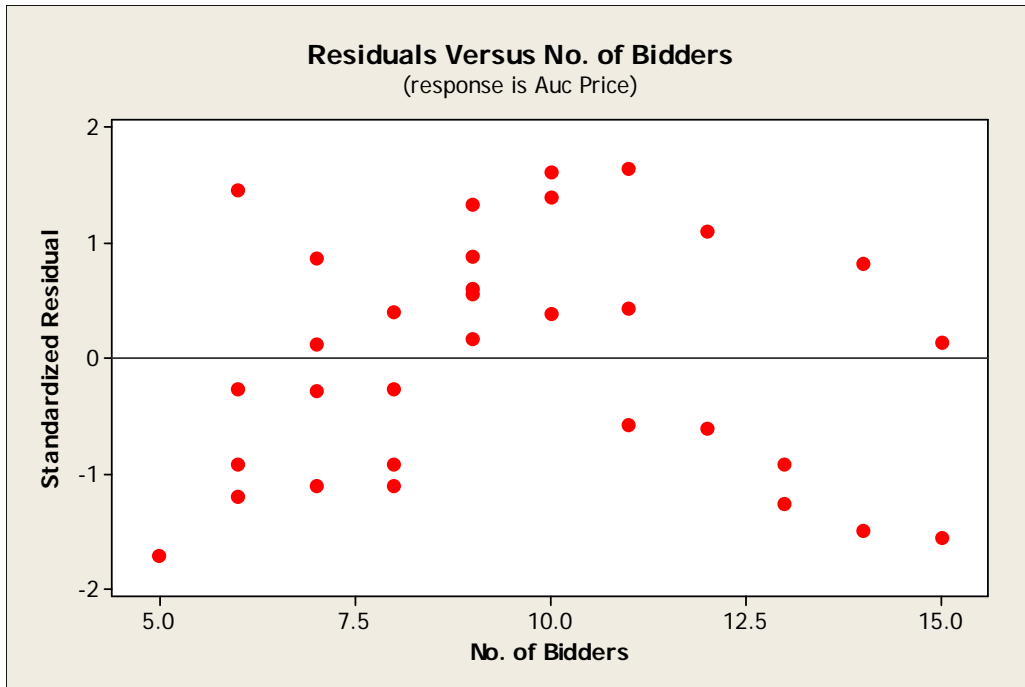
In addition, the following graphs will display:

Residuals Versus No. of Bidders
(response is Auc Price)

If you click on the **Options** button, this screen will appear:

Enter a column of weights to perform weighted regression.

Check to fit a constant term (the y-intercept of the regression line). Uncheck to fit the model without a constant term. Minitab does not display R^2 for this model.

Check to display

Check to perform tests

Type in values for independent variables to get prediction interval

Type the desired confidence level

Click appropriate boxes and input appropriate data (as instructed above) → (For Auction data, 150 10 was entered under Prediction intervals for new observations in order to see what the

predicted auction price would be for a clock that is 150 years old, when there are 10 bidders) →
OK.

Because of the data entered in the Prediction intervals for new observations box, this additional output will be displayed in the session window.

```
Predicted Values for New Observations

New
Obs     Fit   SE Fit       95% CI            95% PI
  1  1431.7     24.6  (1381.4, 1481.9)  (1154.1, 1709.3)


Values of Predictors for New Observations

New        No. of
Obs  Age  Bidders
  1  150     10.0
```

If you click on the **Results** button, this screen will appear:
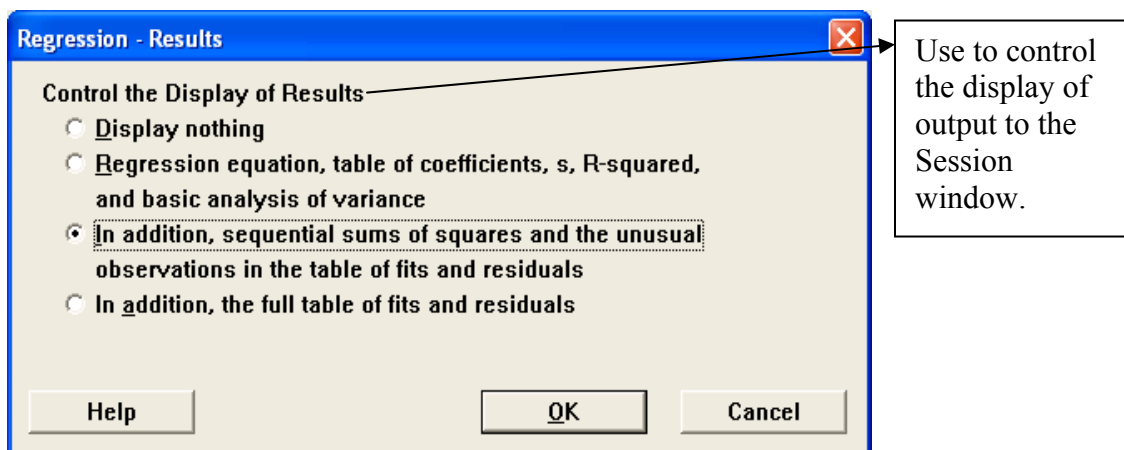
Click **OK** to see the output in the session window:

b)   Identifying Relevant Independent Variables for Best Regression Model

Follow these procedures in order to determine which independent variables are included in the best regression model:

Stat → Regression → Best Subsets → Double-click on dependent variable, then double-click on all independent variables (For The Auction Data, double-click on Auc Price so that it is displayed under Response, and then double-click on Age and No. of Bidders so that they are displayed under Free Predictors).

If you click on the *Options* button, this screen will appear:

You can enter the minimum and maximum number of independent variables you want for the regression equation.

Specify a number from 1 to 5. For example if you choose 3, Minitab will print information from the "best" 3 models of each size (if there are that many).

Uncheck to exclude the intercept term from the regression models.

This output appears in the session window:

Best Subsets Regression: Auc Price versus Age, No. of Bidders

```
Response is Auc Price

                                        N
                                        o
                                        .

                                        o
                                        f

                                        B
                                        i
                                        d
                                        d
                                     A  e
                       Mallows      g  r
Vars  R-Sq  R-Sq(adj)     C-p      S  e  s
   1  53.2       51.7     98.0  273.53  X
   1  15.6       12.8    199.3  367.43     X
   2  89.2       88.5      3.0  133.48  X  X
```
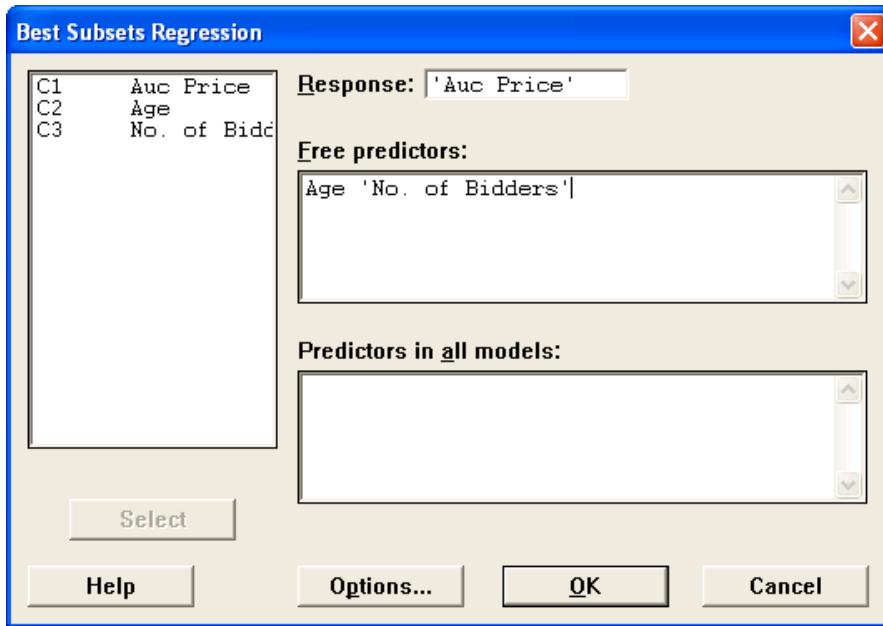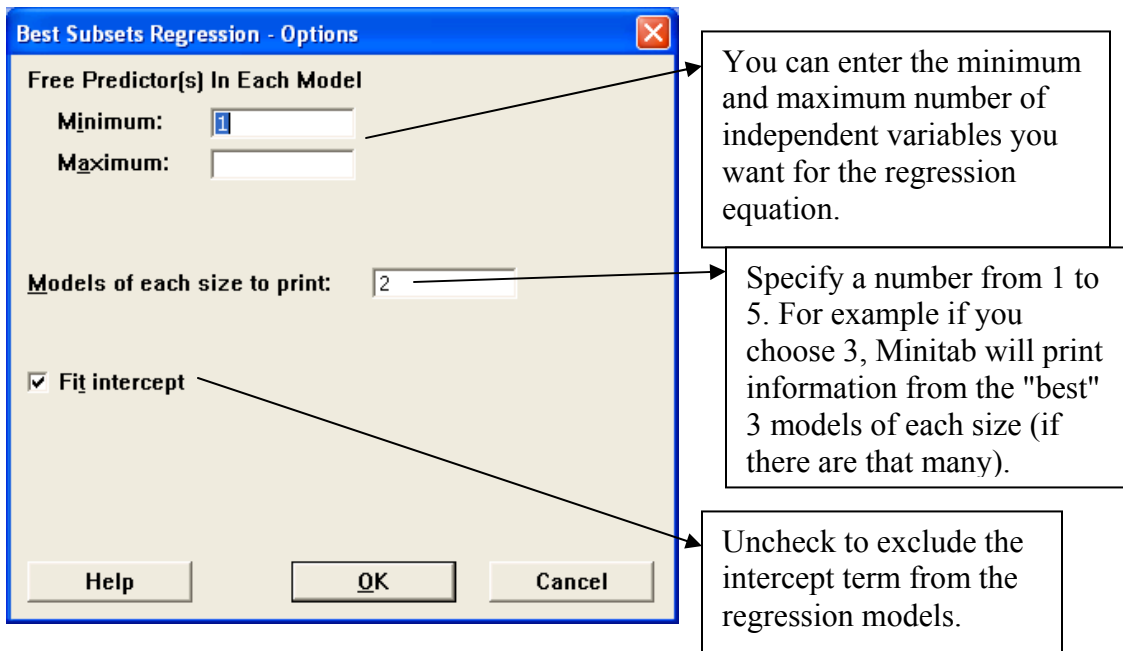
Interpreting the Output: To identify which independent variables are included in the best regression model, first find the highest number under the R-Sq(adj) column and follow the row across. There will be an 'X' in an independent variable column if that variable is included in the best regression model.
(Thus for this example, the best regression model has an adjusted R-squared of 88.5% and both independent variables (Age and No. of Bidders) are included in the model.)

c)   Correlation

To get the coefficient of correlation, go to **Stat → Basic Statistics → Correlation → Double-click on appropriate variables** (For the Auction Data, click on all variables to get every correlation coefficient) **→ OK.**
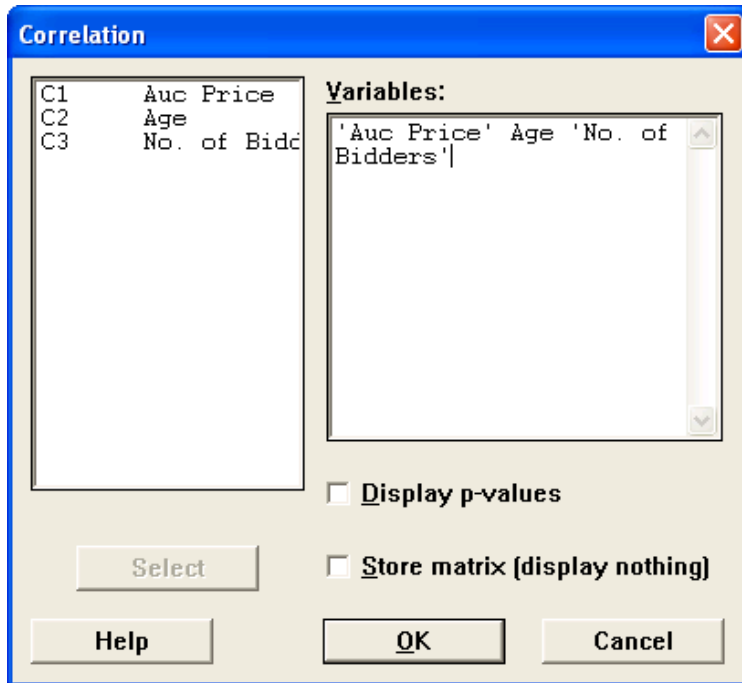
If **Display p-values is unchecked**, this output will be displayed:

Correlations: Auc Price, Age, No. of Bidders

```
                Auc Price           Age
Age                 0.730
No. of Bidde        0.395        -0.254

Cell Contents: Pearson correlation
```

If **Display p-values is checked**, this output will be displayed:

Correlations: Auc Price, Age, No. of Bidders

```
                Auc Price           Age
Age                 0.730
                    0.000

No. of Bidde        0.395        -0.254
                    0.025         0.161


Cell Contents: Pearson correlation
               P-Value
```
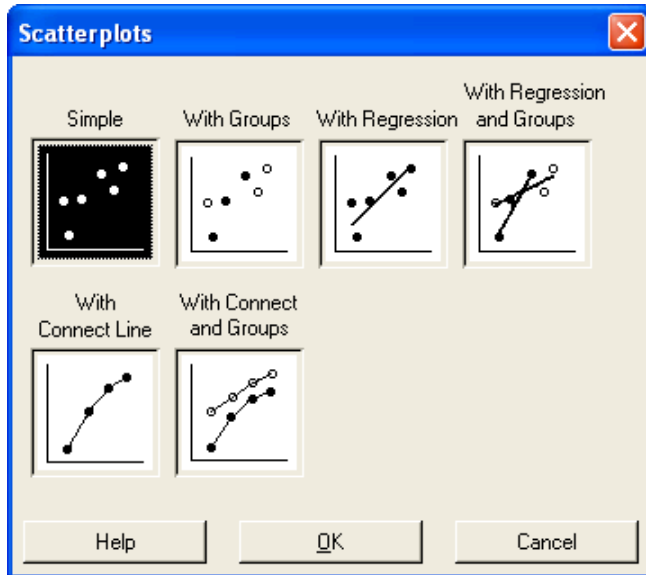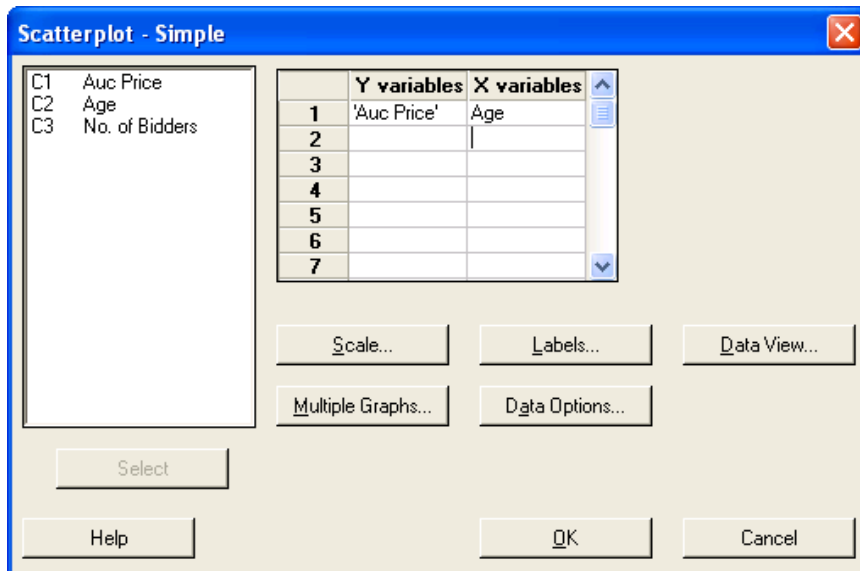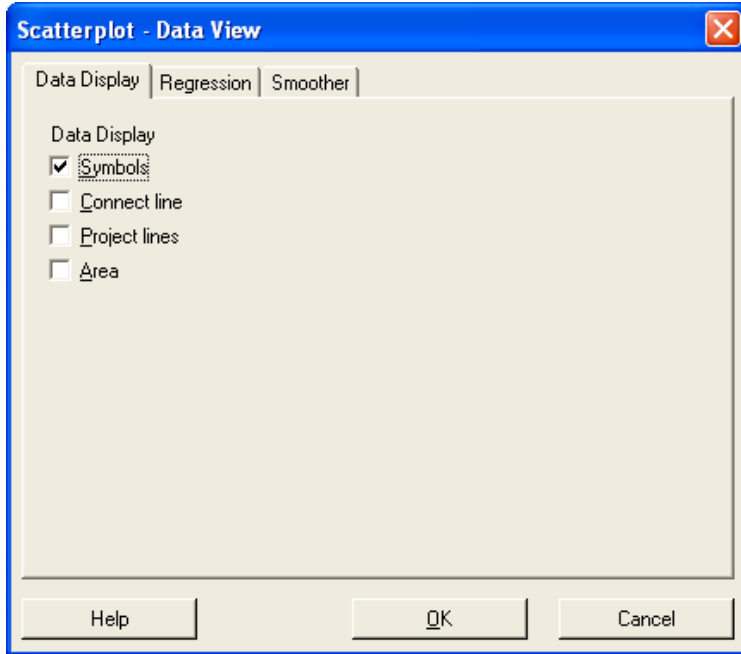
d)   *Scatterplot*

Graph → Scatterplot → This screen will appear:

Double-click on appropriate variables for the x- and y- axes (For Auction Data, double-click on Auction Price for the y variable and Age for the x variable) **→ OK.**

If you click on the **Data View** button, this screen will appear:

Use to represent the data with one or more data display types, including symbols, connect lines, project lines, and areas.

This screen shows the *Data Display* tab. If you click on the *Regression* tab, this screen will appear:
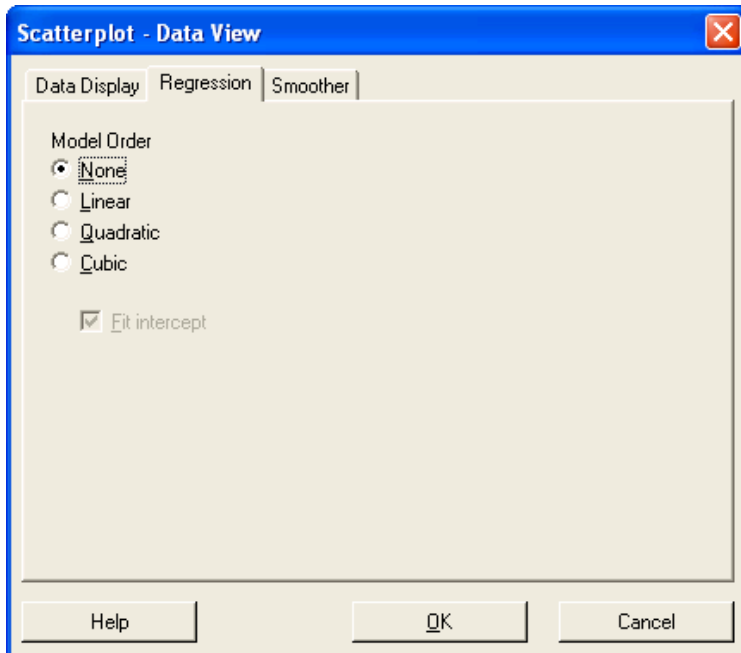
Use to fit a least squares regression line to the scatterplot. Use the regression fit to examine the relationship between the dependent variable (y) and the independent variable (x). If you hover the mouse pointer over the fitted line, Minitab displays the fitted regression equation.

This scatterplot will display:
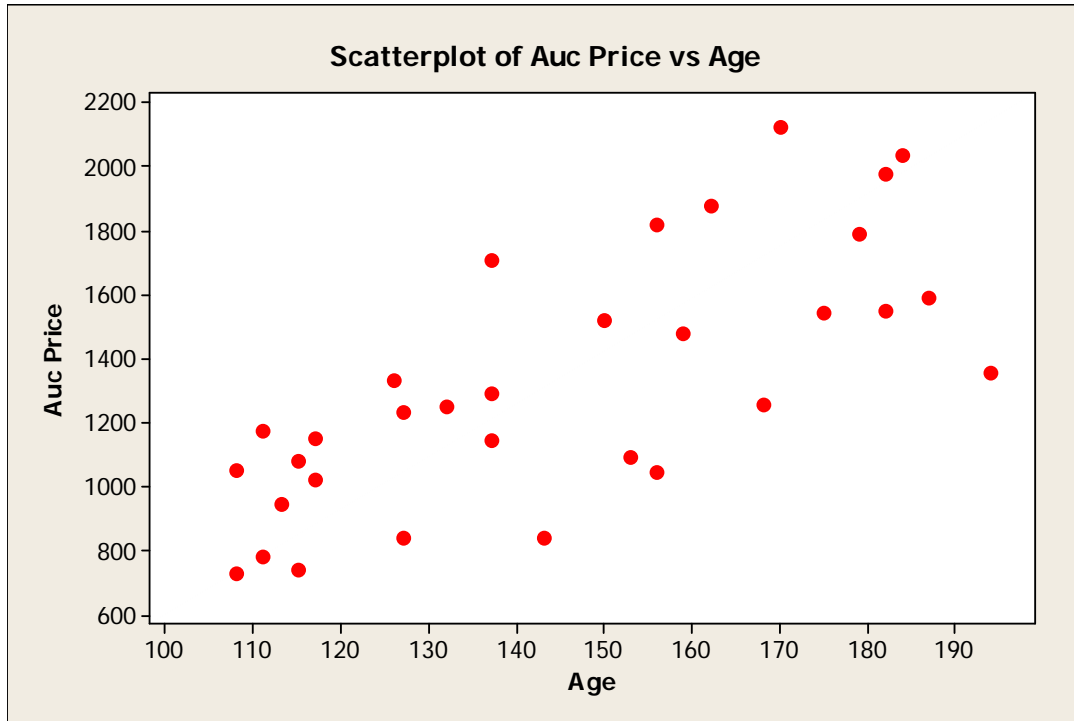
If we would have chosen With Regression instead of Simple,
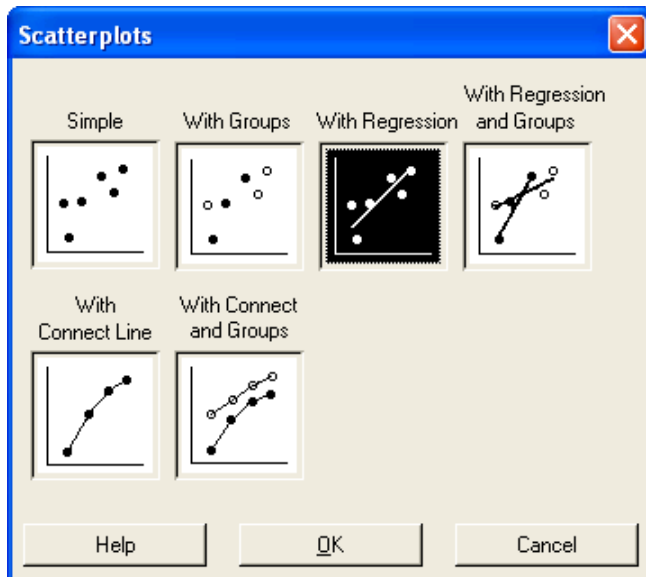
This scatterplot would have displayed:

If we would have chosen With Connect Line,

This scatterplot would have displayed:

Scatterplot of Auc Price vs Age

VII. Appendix

Note: You can copy and paste the data from the Appendix into Minitab if desired.

a)   Dell Data

Here are the monthly returns for a share of stock for Dell during the period January 1994 through December 1998.

Monthly Rate of Return = {[Beginning of Month Price] + [Dividend Per Share]}/ [Beginning of Month Price]

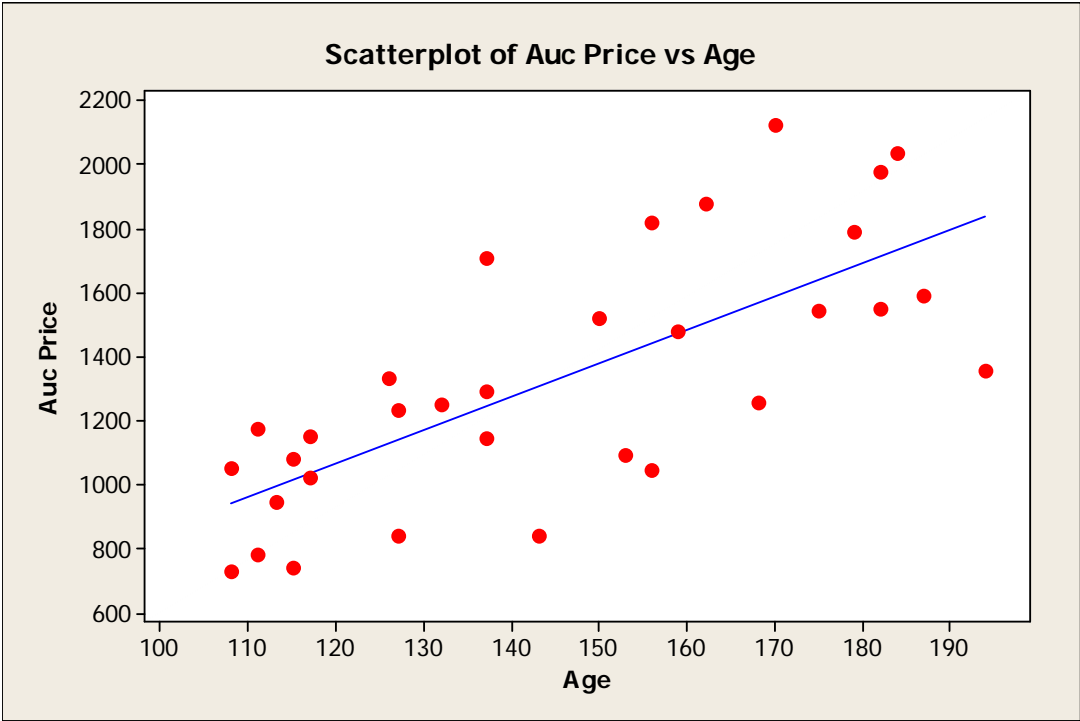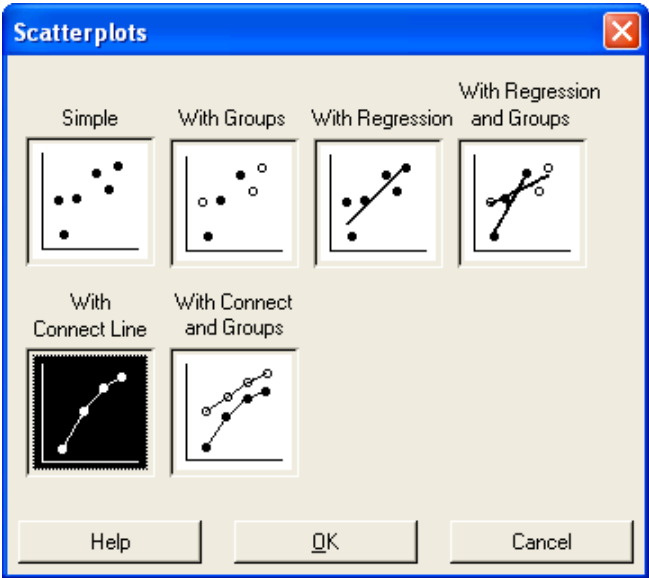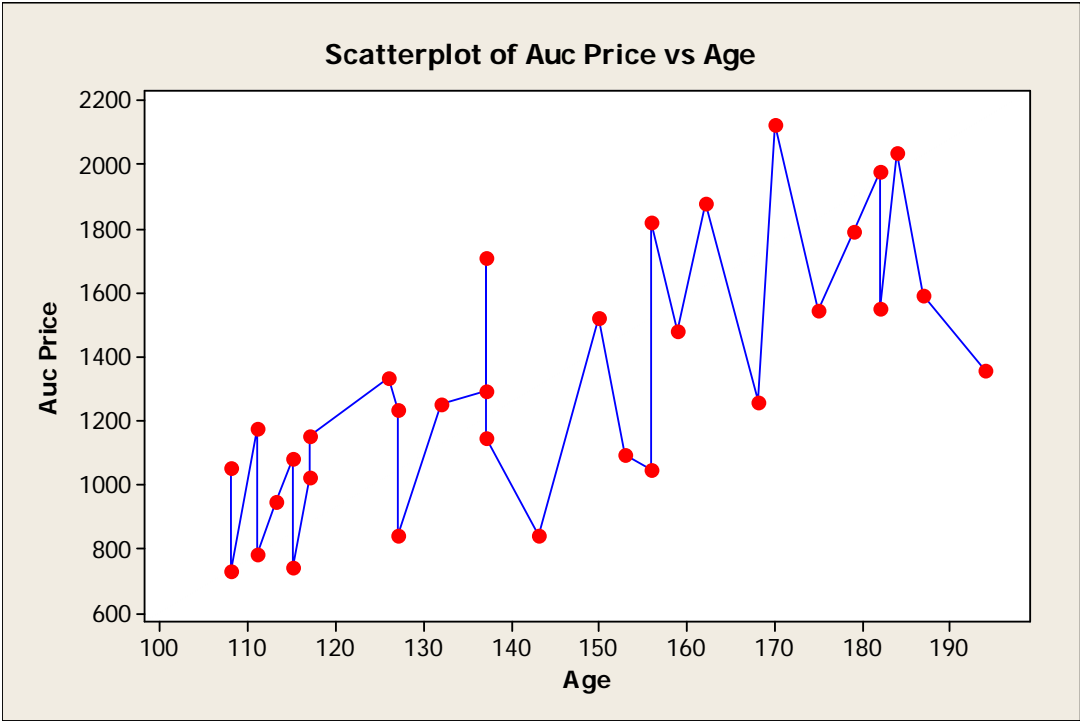All share prices and dividends are adjusted for stock splits.

Date	Rates of Return		
31-Jan-94	-0.0276	31-Jul-96	0.0909
28-Feb-94	0.1364	30-Aug-96	0.2095
31-Mar-94	0.01	30-Sep-96	0.1583
29-Apr-94	-0.1287	31-Oct-96	0.0466
31-May-94	0.3011	29-Nov-96	0.2488
30-Jun-94	-0.0786	31-Dec-96	0.0455
29-Jul-94	0.0616	31-Jan-97	0.2447
31-Aug-94	0.1607	28-Feb-97	0.0756
30-Sep-94	0.1519	31-Mar-97	-0.0492
31-Oct-94	0.1886	30-Apr-97	0.2375
30-Nov-94	-0.0323	30-May-97	0.3443
30-Dec-94	-0.0479	30-Jun-97	0.0439
31-Jan-95	0.0396	31-Jul-97	0.4561
28-Feb-95	-0.0264	29-Aug-97	-0.0402
31-Mar-95	0.0542	30-Sep-97	0.1805
28-Apr-95	0.2514	31-Oct-97	-0.1729
31-May-95	-0.0799	28-Nov-97	0.0507
30-Jun-95	0.1935	31-Dec-97	-0.0022
31-Jul-95	0.0811	30-Jan-98	0.1838
31-Aug-95	0.1846	27-Feb-98	0.4067
29-Sep-95	0.1039	31-Mar-98	-0.0313
31-Oct-95	0.0971	30-Apr-98	0.1919
30-Nov-95	-0.0509	29-May-98	0.0205
29-Dec-95	-0.2175	30-Jun-98	0.1263
31-Jan-96	-0.2094	31-Jul-98	0.17
29-Feb-96	0.2557	31-Aug-98	-0.0791
29-Mar-96	-0.0255	30-Sep-98	0.315
30-Apr-96	0.3694	30-Oct-98	-0.0038
31-May-96	0.2071	30-Nov-98	-0.0716
28-Jun-96	-0.0813	31-Dec-98	0.2035

b)   *Auction Data*

A collector of antique grandfather clocks believes that the winning price of a clock at an auction is a positive function of the age of the clock. In addition, the collector also believes that the auction price is a positive function of the number of bidders. To test this theory, the collector collected a sample of data on the auction prices, ages, and number of bidders at the auction at which the grandfather clocks were sold.

Auc Price	Age	No. of Bidders
1235	127	13
1080	115	12
845	127	7
1522	150	9
1047	156	6
1979	182	11
1822	156	12
1253	132	10
1297	137	9
946	113	9
1713	137	15
1024	117	11
1147	137	8
1092	153	6
1152	117	13
1336	126	10
2131	170	14
1550	182	8
1884	162	11
2041	184	10
845	143	6
1483	159	9
1055	108	14
1545	175	8
729	108	6
1792	179	9
1175	111	15
1593	187	8
785	111	7
744	115	7
1356	194	5
1262	168	7

c) *Student Data*

In order to have example data with categorical variables, this hypothetical sample of Carlson MBA students was used:

Gender	Portfolio	Age
Male	Finance	29
Male	IST	33
Male	Finance	29
Female	Marketing	25
Male	Finance	35
Male	Consulting	25
Male	Consulting	27
Female	Marketing	25
Male	Operations	27
Male	Operations	33
Male	Marketing	38
Male	Marketing	28
Female	Consulting	29
Male	Marketing	28
Male	IST	24
Male	Finance	25
Female	Marketing	29
Female	Finance	32
Male	Consulting	24

d) *Student (2) Data*

Gender	Count
Male	14
Female	5