Chapter Two

"Go with the Flo" Tutorials

Overview

In the next hour or so you will see how data files are organized, learn to create a data set and generate a simple report.

As you will soon learn, creating a new data set, entering information and obtaining tabular results can all be accomplished quickly and easily using *Flo•Stat*.

The following tutorials assume you have some basic familiarity with the Macintosh computer and its operation. If you are new to the Macintosh, stop now and review the manual which accompanied your Macintosh computer. It is important to the successful completion of these tutorials that you feel comfortable with the basic terms and procedures surrounding the use of the Macintosh before learning to use a new application like *Flo•Stat.*

Tutorial #1: Creating a new data set, entering data, naming variables and obtaining a simple list

We begin our tutorial with data from the fictitious Acme Propulsion Company staff file. The first exercise entails entering these data, saving them to a file, and obtaining a listing.

Acme Propulsion Inc. Staff Information File

First Name	Last Name	Age	Department	Gender	Length of Employment
Steve	Beck	32	Sales	1	2
Gunnar	Valgeirsson	48	Marketing	1	14
Kim	Tyus	23	Sales	2	1
Tom	Good	37	Assembly	1	9
Yunan	Jiang	28	Assembly	1	5
Jessica	Clymer	25	Marketing	2	2
Gina	Donovan	31	Sales	2	4
Lynn	Shuck	28	Marketing	2	6
BoShiu	Wu	22	Assembly	1	1
Regina	Rings	51	Marketing	1	15

Basic terms

The names and characteristics of the fictitious employees of the Acme Propulsion Company represent the kinds of information people work with daily . Employee names could just as well be the names of states in a country, for example. These data could also be daily cash register receipts, student test scores, trees in a forest, automobiles on a dealer's car lot, and so on. People routinely need this kind of information sorted, listed, modified, and summarized. Statistics and *Flo•Stat* help do just that.

Statistics performs a simple and wonderful task. It takes scads of information, reduces it and presents it in a concise fashion - helping to make sense of things. Statistics provides ways to answer such questions as: "What proportion of a company's staff is female," and "What is the average age of the staff?" It also provide answers to more complex questions like, "Is there a difference in the average length of employment among the staff in the company's various departments" and, "What is the degree and nature of association (relationship) between length of employment and age of employee?"

If you stop and think about it, much of the information you work with is organized, or could be organized, in much the same way the Acme Propulsion Staff Information File is laid out.

Each line of information contains the same type of data on each person. When information is organized in this manner, each line is said to represent a **case**.

While in this example, the objects being studied (i.e., **unit of analysis)** are people, the unit of analysis could just as well be rats in a clinical psychology experiment, tires in a road endurance test, or cities in an environmental protection agency study comparing the quality of air in U.S. metropolitan areas.

The pieces of information (i.e., name, age, department, gender, and so on) in the Acme Propulsion Company staff file are called **variables**. Variables describe the characteristics of each case, i.e., staff members at Acme Propulsion. For example, in the first case of the Staff Information File, note that Steve Beck is 32 years of age, works in the sales department, and has worked at Acme Propulsion for 2 years.

Also note that the variables for each case are located in the same relative position in the file. In the Staff Information File each case begins with the variable labeled *fname* and ends with the variable, *tenure*.

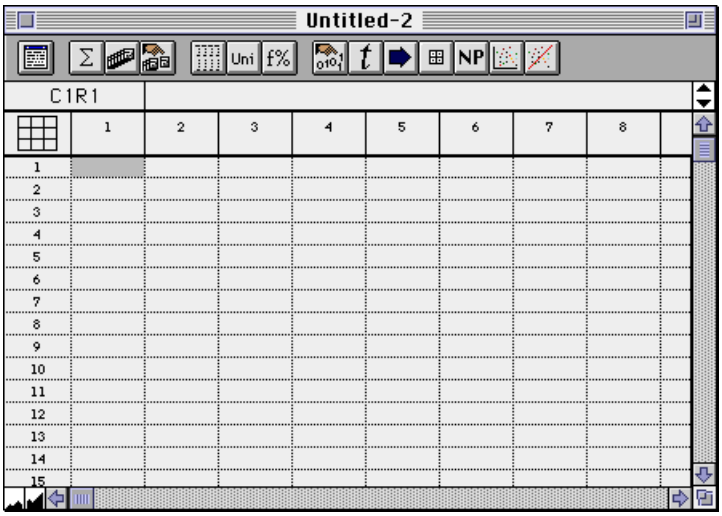
As long as the data you work with are arranged, or can be arranged in this fashion, it is very easy to enter, store, manipulate and analyze those data using *Flo•Stat*.

Data entry

First, enter the Acme Propulsion Staff Information File in to *Flo•Stat*'s data matrix.

Launch *Flo•Stat* by double clicking its application icon.

A window opens containing a set of empty cells - the data matrix. These cells are what holds the information (variables) about each case. Each row of cells represents a case. Each column represents a variable.



The first column and row cell (i.e., C1R1) is selected (darkened) and ready to receive information.

Type the first name (**S t e v e**) of the first case from the Staff Information File data and then hit either the tab key, or the arrow key (->).

Moving among the cells.

- hitting the tab key selects the cell to the right.
- hitting the return or enter key selects the cell below.
- the arrow keys move in their respective direction
- each cell can be directly selected with the mouse

Continue entering data by typing **B e c k** in the second cell, followed by the tab key.

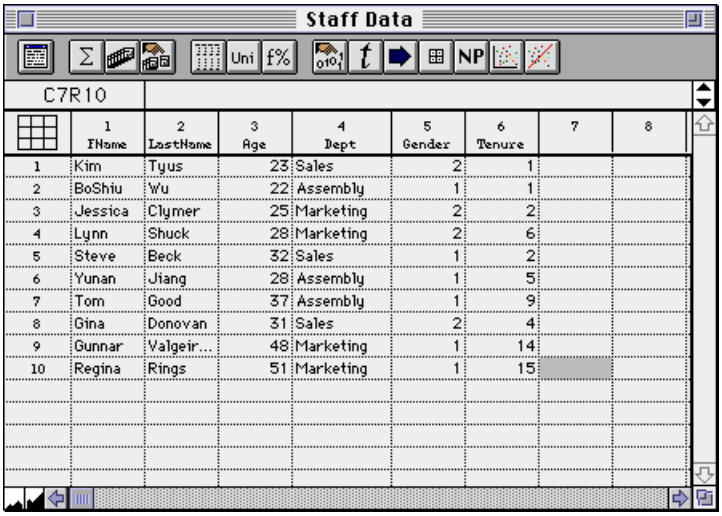
Repeat this process until all of the information for the first case (i.e., **S t e v e B e c k**) has been entered in the first row of cells.
If you enter the data exactly as it appears in the *Staff Information File* table, the cells in row one will contain the staff information on Steve Beck, row two will contain the Gunnar Valgeriersson data, row three the Kim Tyus data, and so on. Column one will contain the first variable, that is each person's first name. Column two will contain each person's last name, column three age, etc.

Data files consisting of one unit of analysis and organized in this fashion are commonly referred as **rectangular data files** .

First Name	Last Name	Age	Department	Gender	Length of Employment
Steve	Beck	32	Sales	1	2
Gunnar	Valgeirsson	48	Marketing	1	14
Kim	Tyus	23	Sales	2	1
Tom	Good	37	Assembly	1	9
Yunan	Jiang	28	Assembly	1	5
Jessica	Clymer	25	Marketing	2	2
Gina	Donovan	31	Sales	2	4
Lynn	Shuck	28	Marketing	2	6
BoShiu	Wu	22	Assembly	1	1
Regina	Rings	51	Marketing	1	15

After having entered the data from all ten cases, your data matrix should be identical to the one below.

Examine your file. If you see a mistake, select that cell and type in the correct information.

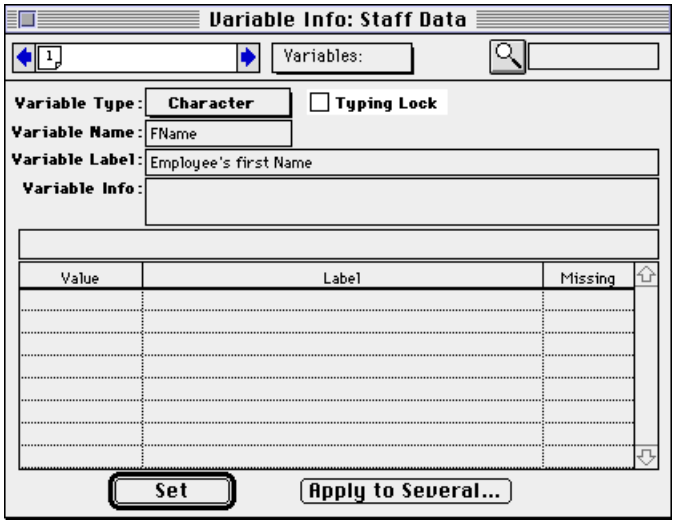


	1 FName	2 LastName	3 Age	4 Dept	5 Gender	6 Tenure	7	8
1	Kim	Tyus	23	Sales	2	1		
2	BoShiu	Wu	22	Assembly	1	1		
3	Jessica	Clymer	25	Marketing	2	2		
4	Lynn	Shuck	28	Marketing	2	6		
5	Steve	Beck	32	Sales	1	2		
6	Yunan	Jiang	28	Assembly	1	5		
7	Tom	Good	37	Assembly	1	9		
8	Gina	Donovan	31	Sales	2	4		
9	Gunnar	Valgeir...	48	Marketing	1	14		
10	Regina	Rings	51	Marketing	1	15		

Once the data are entered, each variable in the data set should be labeled, that is, given a unique name such as **LastName**, **Age**, **Gender**, **Dept**, etc.

To open the window containing the data definition fields, double click on any of the cells in the first column. or select **Var Name, Labels...** from the **Utilities** menu at the top of your screen.

You can enter variable type (numeric or character), variable name, variable label, value labels, information about each variable, and user defined missing values in the variable info window.



Click the button marked **Numeric** next to **Variable Type**. The data in column 1 will now be read as character or string values rather than numeric. (A variable's "type" can be reversed at any time by opening the data definition window and resetting variable type.)

To prevent a column of data from inadvertently being changed, click the **Typing Lock** button on.

Type **FName** in the first edit field. This eight character name is referred to as the **variable name** or **var name.**

Below the var name is a longer edit field reserved for attaching more informative labels, called **variable labels**. Each variable can have a label up to 256 characters in length. Variable names (a maximum of 8 characters) and variable labels (a maximum of 256 characters) are automatically used in all tabular and graphic output.

The variable info edit field is used to attach an additional 256 characters of information about each variable. Unlike the variable label, this information is not displayed in output windows but instead remains only in this window as reference information.

Click the **Set** button, or move the scrolling icon at the top of the window, to save the information just entered.

Columns 1 (first name), 2 (last name) and 4 (department) contain character variables. For *Flo•Stat* to use these data correctly, each must be identified accordingly in the Variable Info window.

Repeat the process of naming the variables and specifying the variable type (character) for the data in column 2 (LastName) and 4 (Dept).

Deleting rows (cases)

By default, new, untitled *Flo•Stat* data sets contain 50 rows and 35 columns. When an analysis is performed using a new data file, *FLO•STAT* assumes the data set contains 50 cases (i.e., rows).

Since there are only 10 cases in the Staff Information File the number of rows in the matrix must be reduced to 10.

Select Rows 11 through 50 by clicking the number 11 in row 11, scrolling to the bottom of the data matrix, holding the shift key and clicking the number 50 in row 50. This action selects rows 11 through 50. Choose **Delete** from the **Edit** menu and these rows will be removed from the matrix.

The data set is now ready to be analyzed.

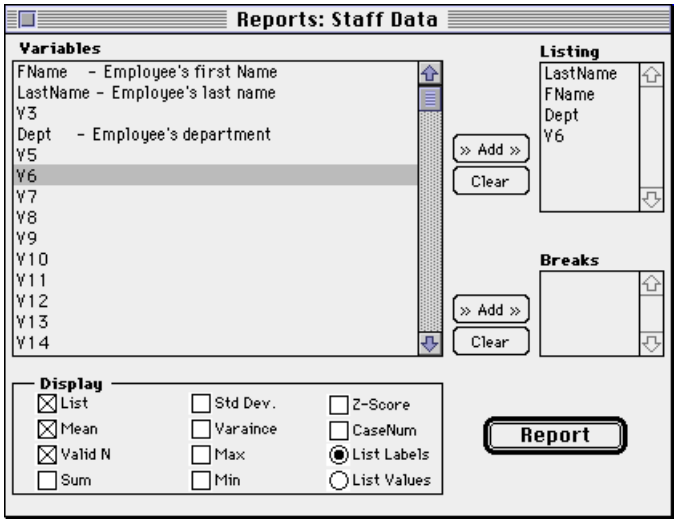
Obtaining a simple list of the data

Select **Reports** from the **Analysis** menu at the top of your screen.

Select **LName** from the variable list. Click the **Add** button**.**

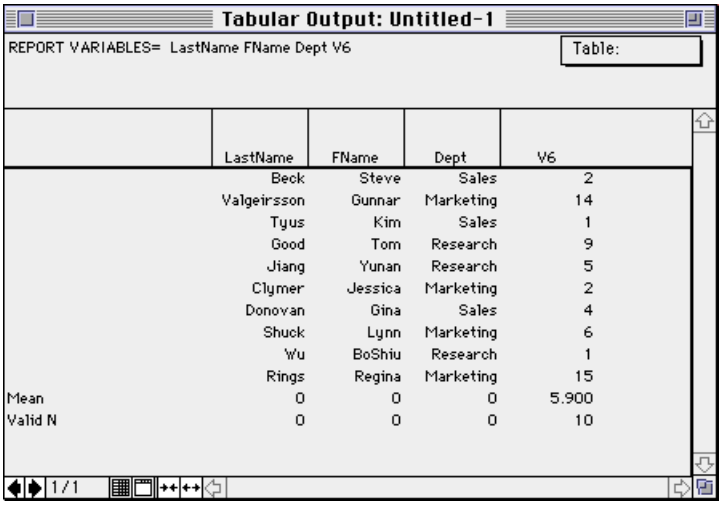
Repeat this for variables **FName**, **Dept**, and **V6**. To alter the order of the variables when displayed, click and drag up or down on a variable in the Listing column.

Select the **Mean** and **Valid N** buttons in the **Display** portion of the window.



Click the **Report** button when ready.

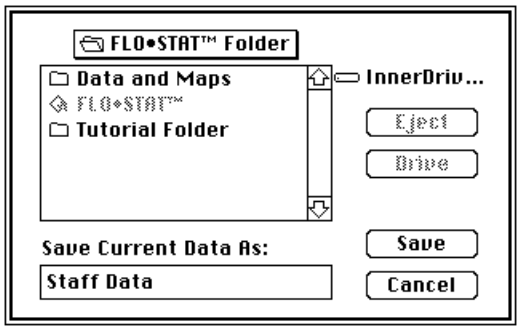
The listing which appears on your screen should be identical to the one below. If any of the first three columns contain a series of dots (i.e., •) return to the variable info window and declare the variable type as character for those variables. (By default, the system assumes numeric data. If it finds character values in a numeric variable it processes those values as if they were **missing**.)



Saving the data set to disk

Select **Save** from the **File** menu.

Name the data set "Staff Data" and click the **Save** button.



To stop and rest before continuing on with the tutorials, select **Quit** from the **File** menu at the top of your screen. The data file has now been saved and will be ready for use when you return.

Tutorial #2: Adding value labels, obtaining frequency distributions, graphing and adding project notes

Open the data file created in Tutorial #1 by double clicking the file's icon. The file can also be opened from within the application by choosing the **Open Data File** command from the **File** menu.

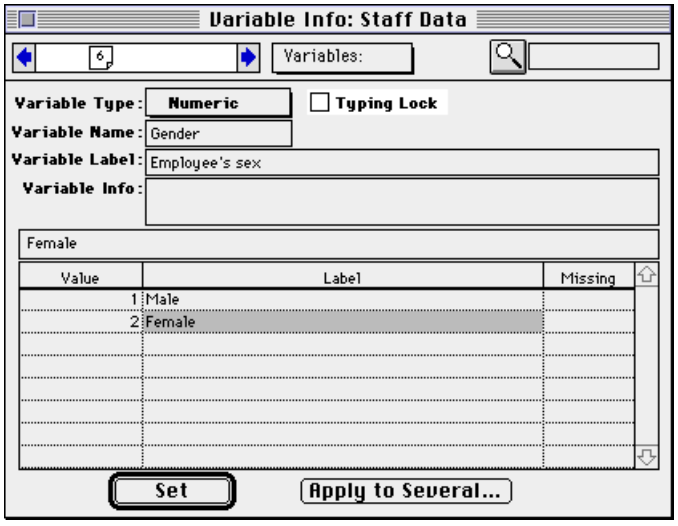
With the data file open, complete the naming and definition of the variables. To name a variable, double click anywhere on that variable's column of data and follow the procedures used in tutorial #1.

As completed in tutorial #1, columns 1, 2 and 4 were named **FName**, **LName** and **Dept**, respectively. Now, name the variables in columns 3, 5 and 6, **Age**, **Gender** and **Tenure**, respectively. Add appropriate variable labels along with these new variables (e.g., Age-Employee's age in years; Gender-Employee's sex; Tenure-Employee's length of employment in years.

Adding value labels

When defining **V5** (**Gender**) the value fields in the variable information window must be used to label the variable's two numerical values (i.e., 1 and 2).

Open the variable info window for **V5** and enter the appropriate values (i.e., 1=male, 2=female). Enter the information exactly as shown on the following page.



When you have finished, click the **Set** button, then close the window.

If the same set of value labels must be applied to more than one variable (as is often the case among questionnaire items using a common set of categories like Strongly Agree, Agree, Disagree, and Strongly Disagree), use the **Apply to**

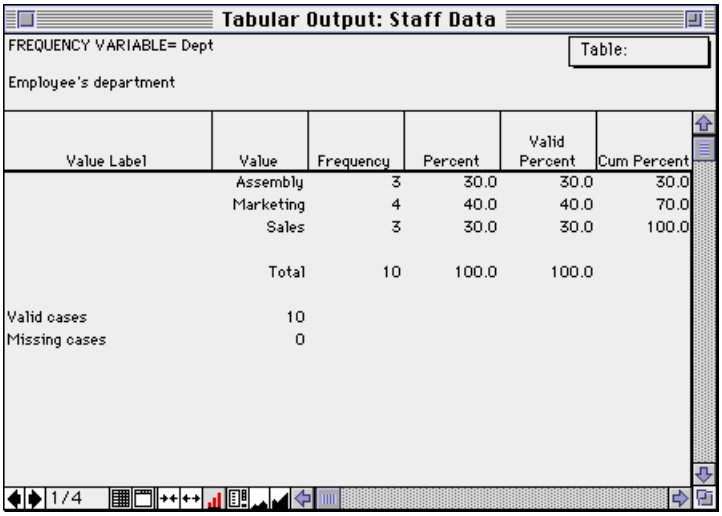
Several... button. Any set of value **labels** can be applied to any other variable or group of variables by clicking the **Apply to Several...** button and selecting the variables from the list presented.

Frequency and percentage distributions

Click the frequencies icon from the menu bar to obtain a variable's frequency and percentage distributions and summary statistics.

Choose **Dept**, **Gender, Age** and **Tenure** from the list of variables.

Click the button marked **Frequencies.**



The tabular output window opens to the first variable in the frequencies list.

To view the second and subsequent variables in the list, you may either click on the right arrow in the lower left corner of the window, select the variable of choice from the **Table** menu in the window, or hit the right arrow key on the keyboard.

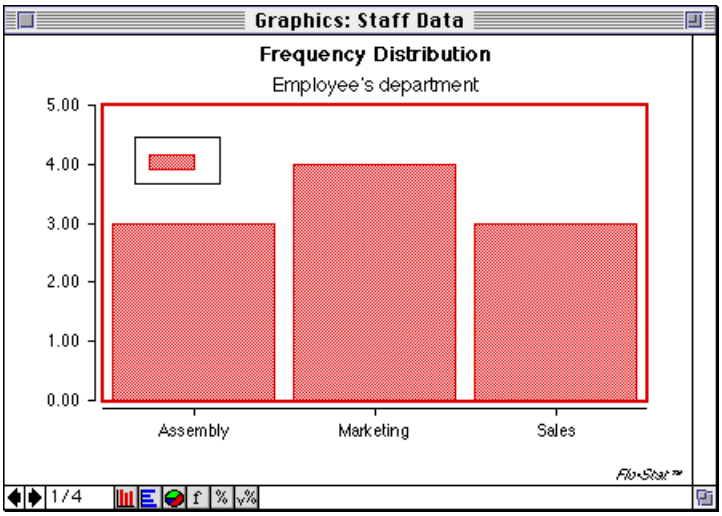
The numbers 1/4 in the lower left corner denotes the first of four output tables. In addition to the frequency distribution, the table contains the value labels, actual values, percent, valid percent, and cumulative percent distributions.

Below these distributions, univariate measures of central tendency and dispersion are displayed. Tables generated on all numerical variables contain this information. These measures are not appropriate for character variables and therefore are not displayed.

To enlarge or shrink the size of the text in the output window click on either side of the view icon.

Graphing the results

To obtain a graph of these distributions, click once on the graph icon at the bottom of the tabular output window or choose **Show Graphics** from the **Output** menu.

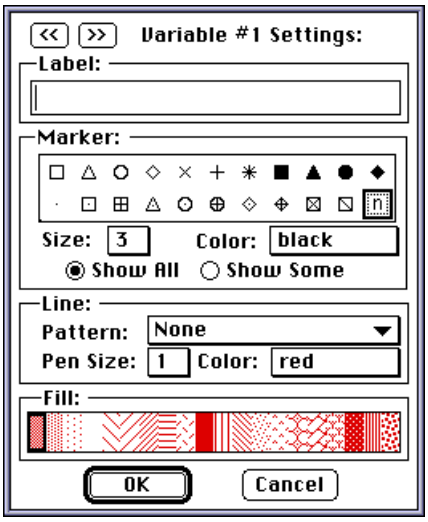


Click the appropriate icon at the bottom of the graphics window to change chart type and distribution. (i.e., frequency, percent and valid percent) graphed.

Double clicking on any chart object opens a dialog box. The dialog box allows you to alter the characteristics of the chart's various objects-color, text, line width, fill pattern, labels, and marker.

Double click near the center of the legend box. The legend box is automatically placed in the upper left corner of each chart. The legend can be moved by clicking and dragging it in the window.

Set the **Fill** pattern to solid and click **OK**.

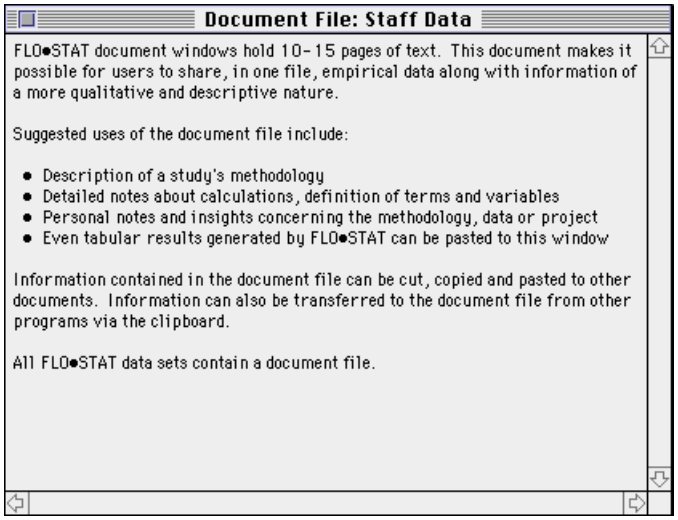


Adding project notes

It is often necessary to maintain project notes with a data set. Similar to mainframe statistical systems, which permit a text document to accompany system files or permanent data sets, it is easy to attach a text file to every *Flo•Stat* data file.

Click the document icon to open the text file which accompanies all *Flo•Stat* data files.
As explained in the document window, the document file is especially handy for keeping study and project notes.

The document window can store and display up to 15 pages of text.



This concludes tutorial #2. If you wish to save the changes made to the data file, select **Save** from the **File** menu.

If you want to continue on with tutorial #3, select **Close Data File** from the **File** menu.

If you wish to stop and continue at some later date, select **Quit**.

Tutorial #3: Adding columns and rows, recoding variables, adding value labels, sorting cases, selecting cases, correlating variables

Add the following annual salary information to the Staff Information data set. Insert these data between **LastName** (column 2) and **Age** (column 3).

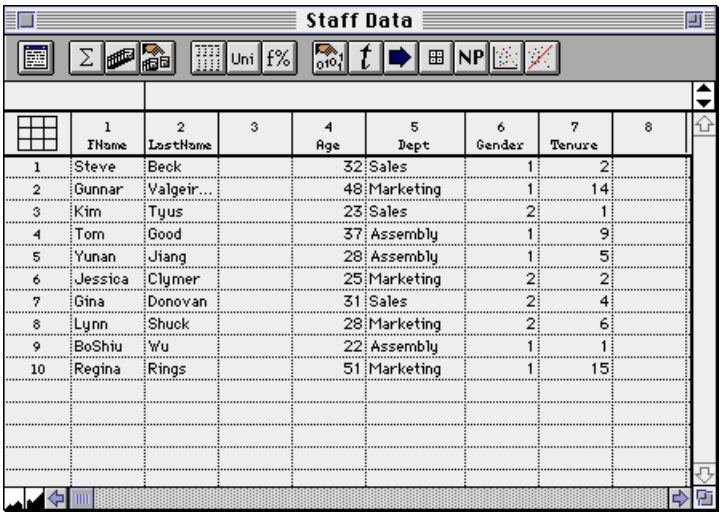
	Last Name	Salary
1	Beck	35640
2	Valgeirsson	79320
3	Tyus	24500
4	Good	38690
5	Jiang	37900
6	Clymer	28350
7	Donovan	42100
8	Shuck	31000
9	Wu	27850
10	Rings	82500

Adding a column

To add an empty column or row to the end of the data matrix, select **Add Column** or **Add Row**, respectively, from the **Edit** menu..

To insert a column or row, select the column or row in front of which you want the new column/row added and select **Insert** from the **Edit** menu. To insert more than one column or row, click and drag across the desired number of column headings, beginning at the point where you want the new columns added, and then select **Insert** from the **Edit** menu.

Insert a new column in front of the variable named **Age** and enter the salary data.



	1 IName	2 LastName	3	4 Age	5 Dept	6 Gender	7 Tenure	8
1	Steve	Beck		32	Sales	1	2	
2	Gunnar	Valgeir...		48	Marketing	1	14	
3	Kim	Tyus		23	Sales	2	1	
4	Tom	Good		37	Assembly	1	9	
5	Yunan	Jiang		28	Assembly	1	5	
6	Jessica	Clymer		25	Marketing	2	2	
7	Gina	Donovan		31	Sales	2	4	
8	Lynn	Shuck		28	Marketing	2	6	
9	BoShiu	Wu		22	Assembly	1	1	
10	Regina	Rings		51	Marketing	1	15	

Name and label the new variable. (e.g., "**Salary**" and "Employees annual salary in dollars.")

Recoding variables

Sometimes it is desirable to recode a variable's information into a new set of values. For example, single year of age data (e.g., 18, 19, 20, 21, etc.) is cumbersome to use in certain analyses and reports. What is often needed, instead, are broad age groupings such as, 18-29, 30-49, and 50+.

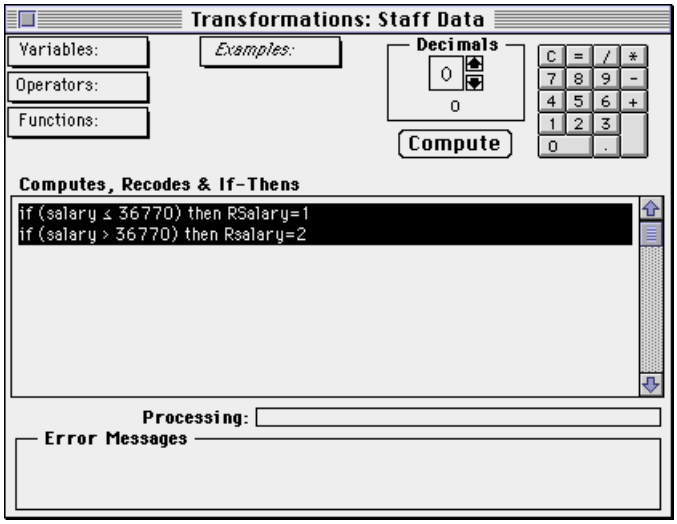
Flo•Stat contains a powerful editor which makes it easy to transform data, and the syntax used to compute and recode variables follows many of the conventions found in the widely used mainframe statistical packages SPSS™ and SAS™. (SPSS and SAS are trademarks of their respective corporations.)

Click the transformation icon.

Type the following IF-THEN statements:

If (Salary ≤ 36770) THEN RSalary=1
If (Salary > 36770) THEN RSalary=2

Select these two statements. Click the **Compute** button.



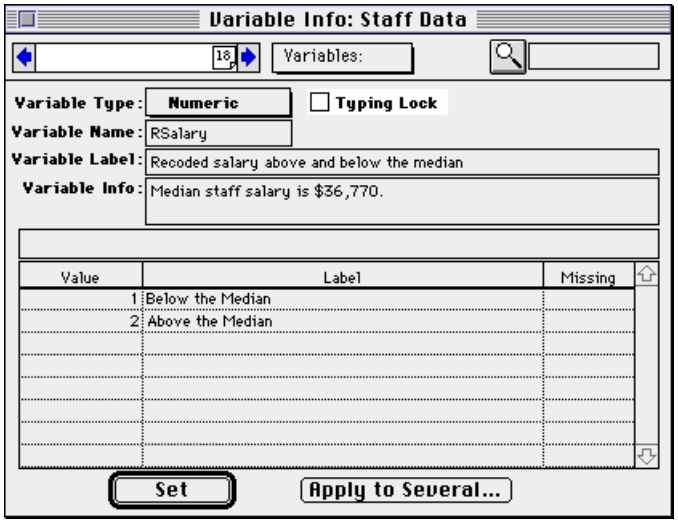
The median employee salary of $36,770 was obtained by running **Frequencies** on the staff's salary variable. Each employee's salary is marked as falling below (1) or above (2) the median. These values are assigned to a new variable named **RSalary**. **RSalary** is automatically added as the last column in the data matrix.

Scroll to the end of the data matrix to see if the new variable and data have been added.

Adding value labels to the new variable

To add value labels to this newly recoded variable, click on the column containing **RSalary**.

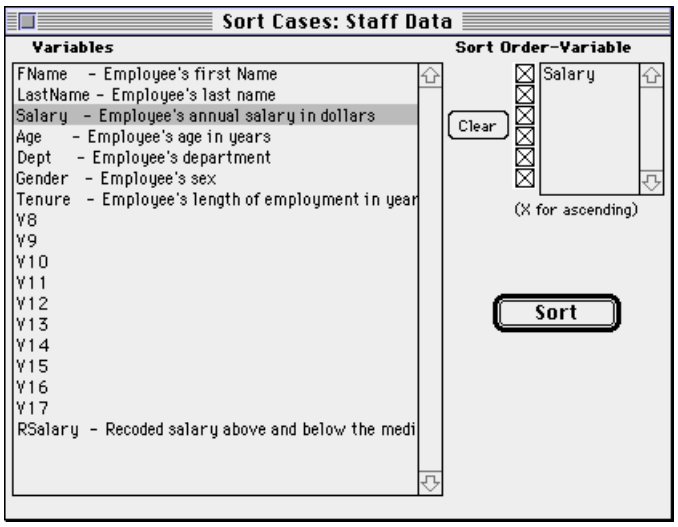
Use the window below for guidance when entering the , variable label, and value labels for **RSalary**.



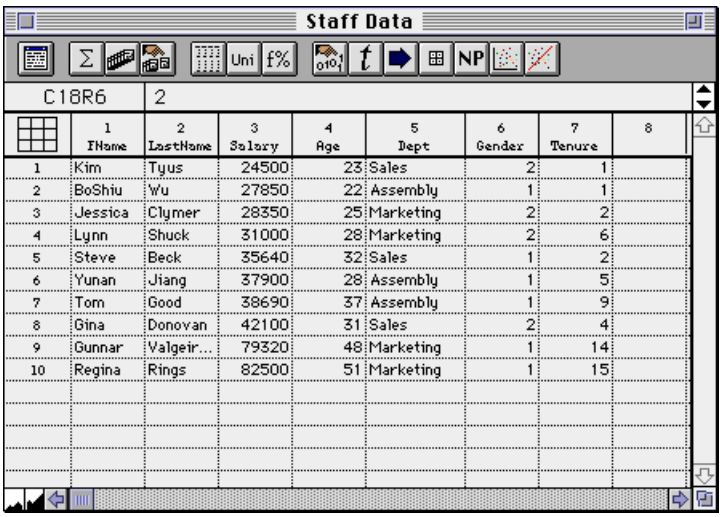
Sorting cases

To prepare a report consisting of a sorted list of employees by their respective salaries, click the sort icon or select **Sort...** from the **Utilities** menu.

Select the variable named **Salary**. Click the **Sort** button when ready.



View the input data window. The cases in the data matrix will be in order from the lowest to highest salary.



	1 FName	2 LastName	3 Salary	4 Age	5 Dept	6 Gender	7 Tenure	8
1	Kim	Tyus	24500	23	Sales	2	1	
2	BoShiu	Wu	27850	22	Assembly	1	1	
3	Jessica	Clymer	28350	25	Marketing	2	2	
4	Lynn	Shuck	31000	28	Marketing	2	6	
5	Steve	Beck	35640	32	Sales	1	2	
6	Yunan	Jiang	37900	28	Assembly	1	5	
7	Tom	Good	38690	37	Assembly	1	9	
8	Gina	Donovan	42100	31	Sales	2	4	
9	Gunnar	Valgeir...	79320	48	Marketing	1	14	
10	Regina	Rings	82500	51	Marketing	1	15	

Salary report

To generate a salary report, click the Reports icon or select **Reports** from the **Analysis** menu at the top of the screen.

Select **Lastname**, **FName**, **Salary** and **RSalary** for the report. Add **Dept** as the break variable. "Breaks" subdivide the data into categories of the break variable(s). To obtain summary statistics for all variables within each category of the break variable, set the **Mean** and **Sum** buttons in the Reports window.

To contract or expand the column widths in the tabular output, click the appropriate icon at the bottom of the output window. To expand or contract the width of any single column, click and drag the vertical line at the top and to the left of the column.

The average salary of employees varies sharply by department. The three employees in the sales department earn an average of just over $34,000, while those in the assembly area earn roughly $34,800. The four employees working in marketing, on the other hand, earn an average of just over $55,000. In short, the annual cost in salaries to the firm is $221,170, $104,440, and $102,240 for the marketing, assembly and sales departments, respectively.

Tabular Output: Staff Data

REPORT VARIABLES= LastName FName RSalary Salary Table:

Break Variable= Dept

Dept	LastName	FName	RSalary	Salary
Assembly	Wu	BoShiu	Below the Median	27850
	Jiang	Yunan	Above the Median	37900
	Good	Tom	Above the Median	38690
Mean	0	0	1.667	34813.333
Sum	0.000	0.000	5.000	104440.000
Marketing	Clymer	Jessica	Below the Median	28350
	Shuck	Lynn	Below the Median	31000
	Valgeirsson	Gunnar	Above the Median	79320
	Rings	Regina	Above the Median	82500
Mean	0	0	1.500	55292.500
Sum	0.000	0.000	6.000	221170.000
Sales	Tyus	Kim	Below the Median	24500
	Beck	Steve	Below the Median	35640
	Donovan	Gina	Above the Median	42100
Mean	0	0	1.333	34080.000
Sum	0.000	0.000	4.000	102240.000

1/1

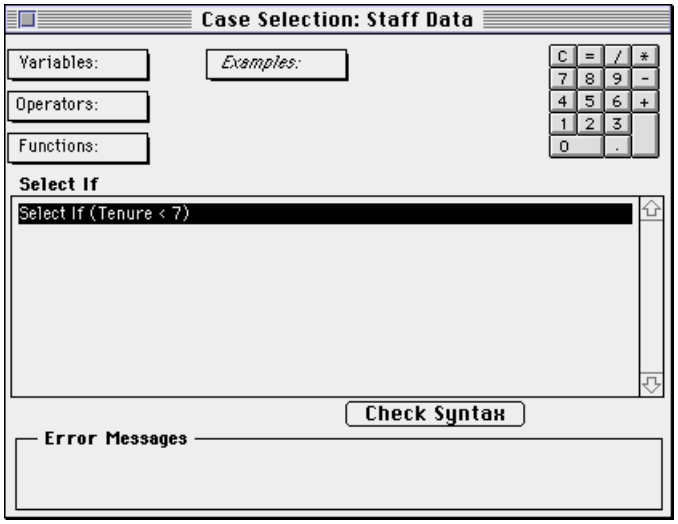
Selecting cases

Are the differences observed in departmental salary averages largely a reflection of differences in pay by occupation, or is there something else causing the sharp differentials? While there are a number of possible explanations, in this case perhaps the variation among departments can be attributed to differences in length of employment.

Click the case selection icon in the icon menu bar to obtain a subset of the ten employees at Acme Propulsion.

Select those employees who have worked at the company for six or fewer years. To accomplish this, type the following statement in the case selection window.

Select If (Tenure < 7)



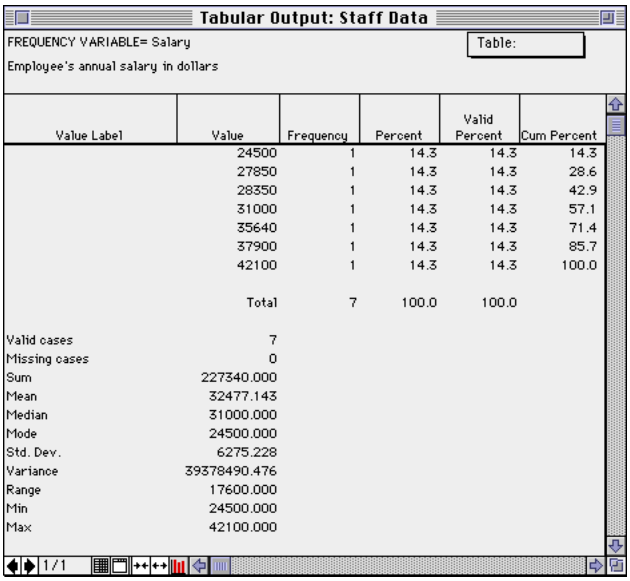
Highlight the statement and click the **Check Syntax** button to test the statement's syntax before running one of the statistical procedures.

It is very important to note that when a **Select If** statement(s) is selected in the Case Selection window, that selection will remain in effect until you return to the Case Selection window and unselect it.

Next, obtain a frequency distribution of the salary variable.

Click the Frequencies icon and choose **Salary** from the list of variables.

As the output table indicates, seven of the ten employees met the selection criteria specified in the Case Selection window (i.e., Tenure < 7) and have worked at Acme Propulsion for six or fewer years. You can also see that their average salaries, as you might have suspected, are substantially lower than what was observed for the firm as a whole. As shown, the mean salary among recent employees is $32,477, roughly $10,000 less than the company as a whole.



Tabular Output: Staff Data

FREQUENCY VARIABLE= Salary

Table:

Employee's annual salary in dollars

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	24500	1	14.3	14.3	14.3
	27850	1	14.3	14.3	28.6
	28350	1	14.3	14.3	42.9
	31000	1	14.3	14.3	57.1
	35640	1	14.3	14.3	71.4
	37900	1	14.3	14.3	85.7
	42100	1	14.3	14.3	100.0
	Total	7	100.0	100.0	

Valid cases	7
Missing cases	0
Sum	227340.000
Mean	32477.143
Median	31000.000
Mode	24500.000
Std. Dev.	6275.228
Variance	39378490.476
Range	17600.000
Min	24500.000
Max	42100.000

1/1

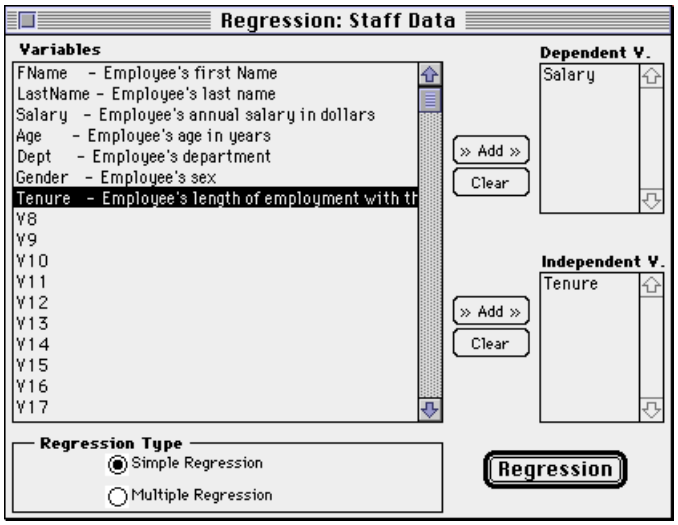
Remember, as long as the **Select If** statement remains highlighted in the Case Selection window, cases failing to match the selection criteria will be filtered out of any further statistical procedures. Return to the case selection window and "turn off" the **Select If** statement by clicking once on the statement.

Measuring the strength of association between variables

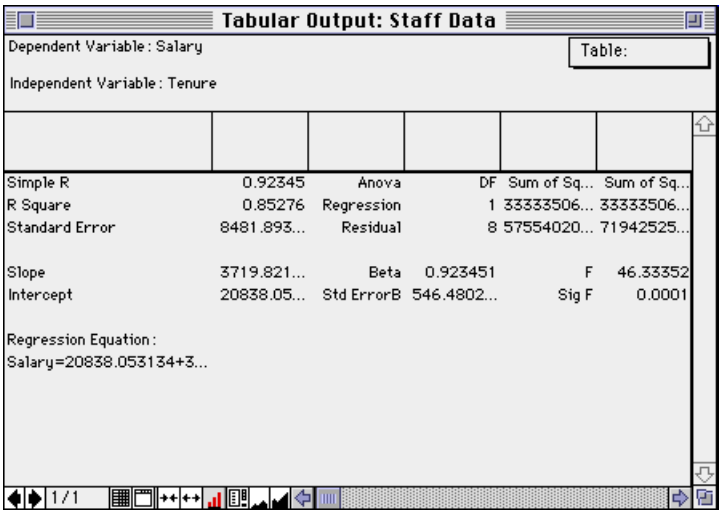
To what extent is salary associated with length of employment (**Tenure**) and can tenure be used to predict salary? Our simple analysis may suggest the two variables are related but the extent of that association is unclear.

Select **the regression icon.**

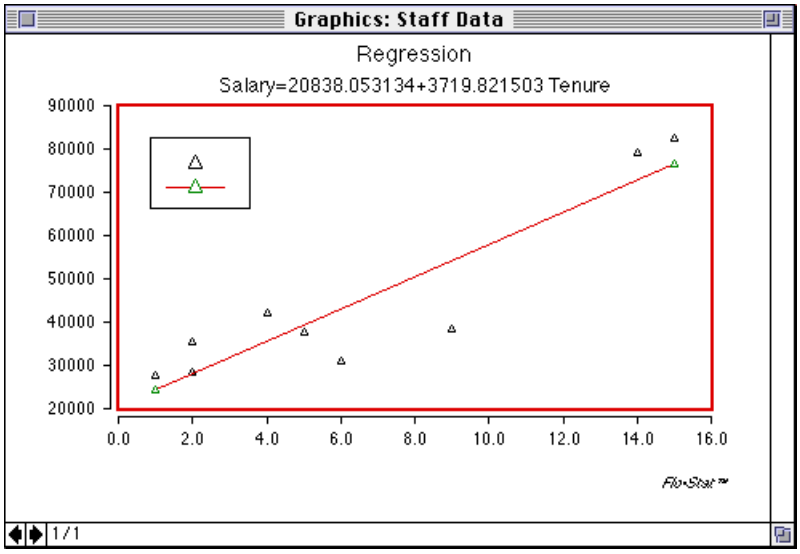
Add **Salary** to the dependent variable list and **Tenure** to the Independent variable list. Set regression type to **Simple Regression.** Click the **Regression** button.



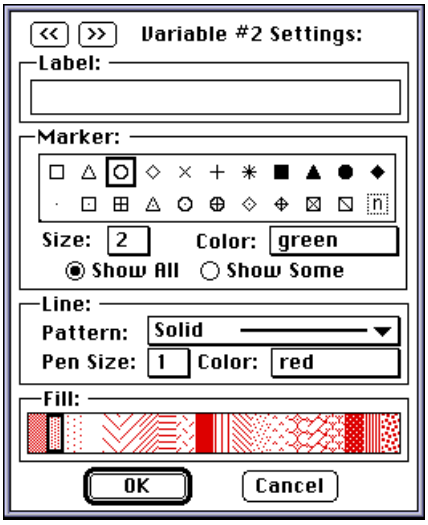
The tabular output indicates that there is indeed a strong, positive association (r=.923) between **Tenure** and **Salary** and that a large portion of the variation in employee salary can be explained by the length of time an employee has worked for Acme Propulsion (R Squared = .852).



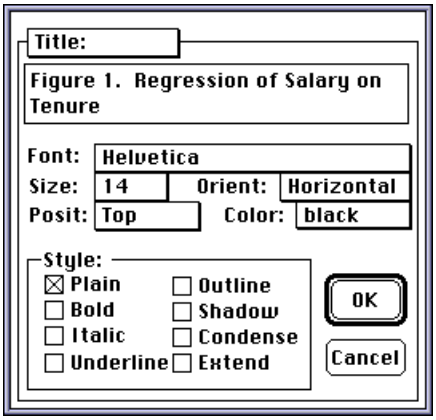
Click the **graphics** icon at the bottom of the tabular output window to examine this relationship graphically.



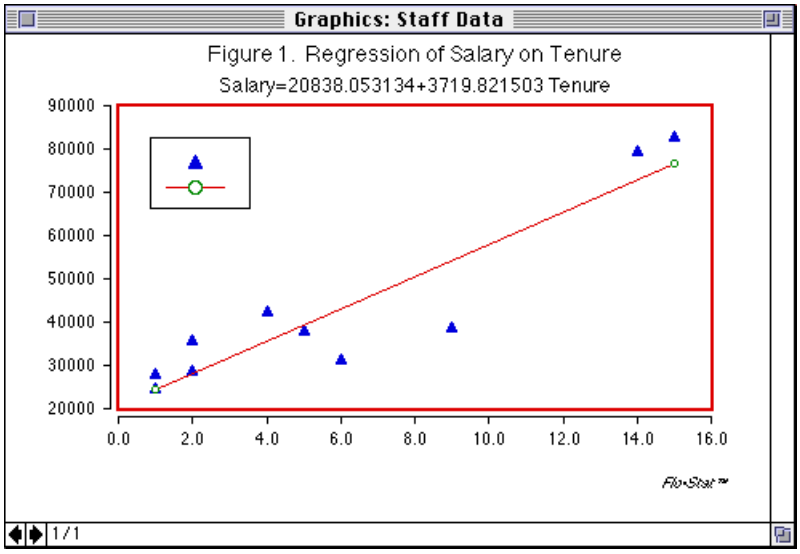
To modify the plot's symbols, double click on the legend's triangle. Choose a new symbol, symbol size and color from the graphic's settings dialog window.



Double click the graph's title. In the title portion of the dialog window type, "Figure 1. Regression of Salary on Tenure."



The scattergram you obtain should look similar to one below.



It is clear from the scattergram and regression line that the relationship between tenure and salary is linear in nature and that the longer a person works at Acme Propulsion their salary increases. (In truth, the relationship between length of employment and salary is far more complex than this, but that's a matter you might want to think about and explore with another data set. Relationships are simple and the work is steady when you're employed at Acme Propulsion.)

To obtain a copy of either the tables or graphs, you can copy the output to the clipboard and paste it into a word document or print directly to a laserprinter.

Conclusion

Congratulations on completing the tutorials! You have reviewed a sufficient number of utilities and statistical procedures to begin entering and analyzing your own data.