

Graphics in SAS

Statistics 135

Autumn 2005



Installing SAS on your PC

I believe it is possible to install **SAS** on your own PC. To do this, follow the following steps

1. Choose Run... from the Start menu
2. File in the Open box with \\dssssoft10\softpc.
3. When prompted for username and password give FAS_DOMAIN\fasusername for the user name and your usual fas password for the password. This will open up a directory window.
4. Copy the folder SAS Server to your PC. Note it may take a few minutes as its big (1.37 Gigabyte)
5. Open this folder followed by folders 9.1 and Disk1. Here you will find the installer program. Install like any other Windows program.

Note that when I tried to install it on one of the Stat Department lab machines, it appeared that the installer got stuck during the “Performing post-processing. This may take a while...” stage with the progress bar at 94%. I eventually had to quit out of the installer program. At one point it appears to have given a working version of **SAS** but I haven’t heavily tested it. However another time I tried it wouldn’t work.

So installer beware ...

Graphics in SAS

As has been implied, there are many approaches to creating graphics in **SAS**. There are the lineprinter based graphics, as created by PROC UNIVARIATE, PROC CHART and PROC PLOT. However, high resolution figures can be created by PROC UNIVARIATE, PROC BOXPLOT, PROC GCHART and GPLOT for example.

While we will focus on high resolutions, lets discuss the low resolution line printer approaches briefly as their approaches are similar to what is needed for the high resolution graphs.

Lineprinter Graphics in SAS

- PROC CHART: Used for creating the following univariate plots of categorical data: vertical bar charts (VBAR), horizontal bar charts (HBAR), pie charts (PIE), and star charts (STAR).

Star charts are circular bar charts, where the bars radiate from a center point. They are useful for cyclical data (such as measurements taken every month or day or hour) or for data that have an inherent ordering.

Here are a couple of examples

```
OPTIONS PAGESIZE = 20; /* for sizing output for LaTeX slides */
PROC CHART DATA = shingles2;
  HBAR potentcat;
  HBAR potentcat / FREQ;
  HBAR potentcat / CFREQ;
  VBAR potentcat / SUMVAR = SALES TYPE = MEAN;
  VBAR potentcat / GROUP = training;
  TITLE 'Breakdown of Region Potential Classes';
  FORMAT potfmt. trainfmt.;
```

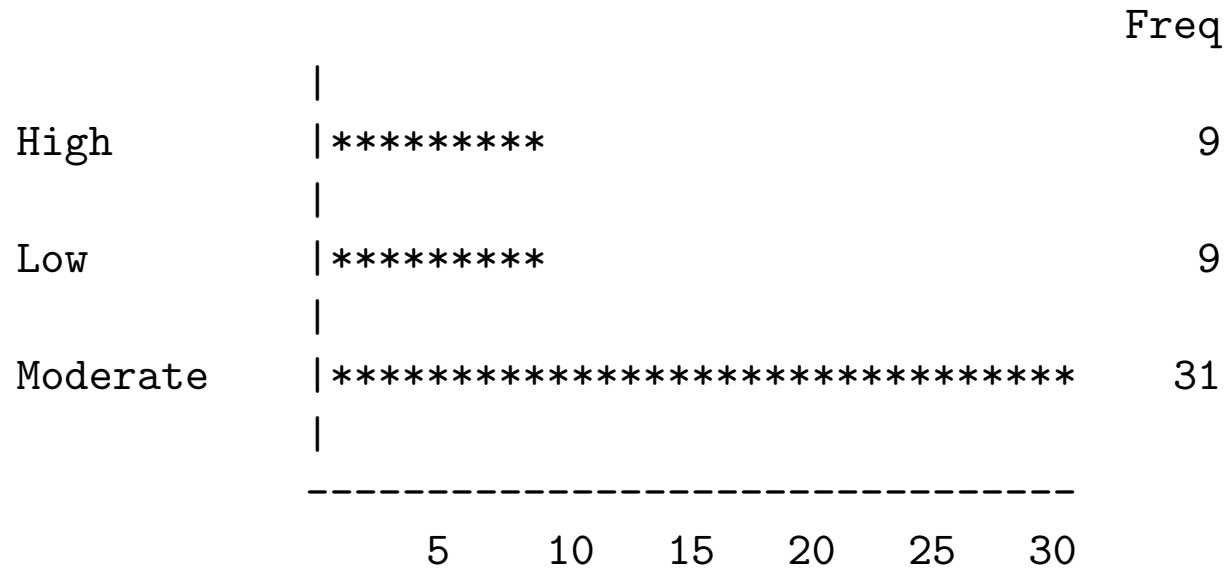
Breakdown of Region Potential Classes

potentcat		Freq	Cum. Freq	Percent	Cum. Percent
High	*****	9	9	18.37	18.37
Low	*****	9	18	18.37	36.73
Moderate	*****	31	49	63.27	100.00

Frequency

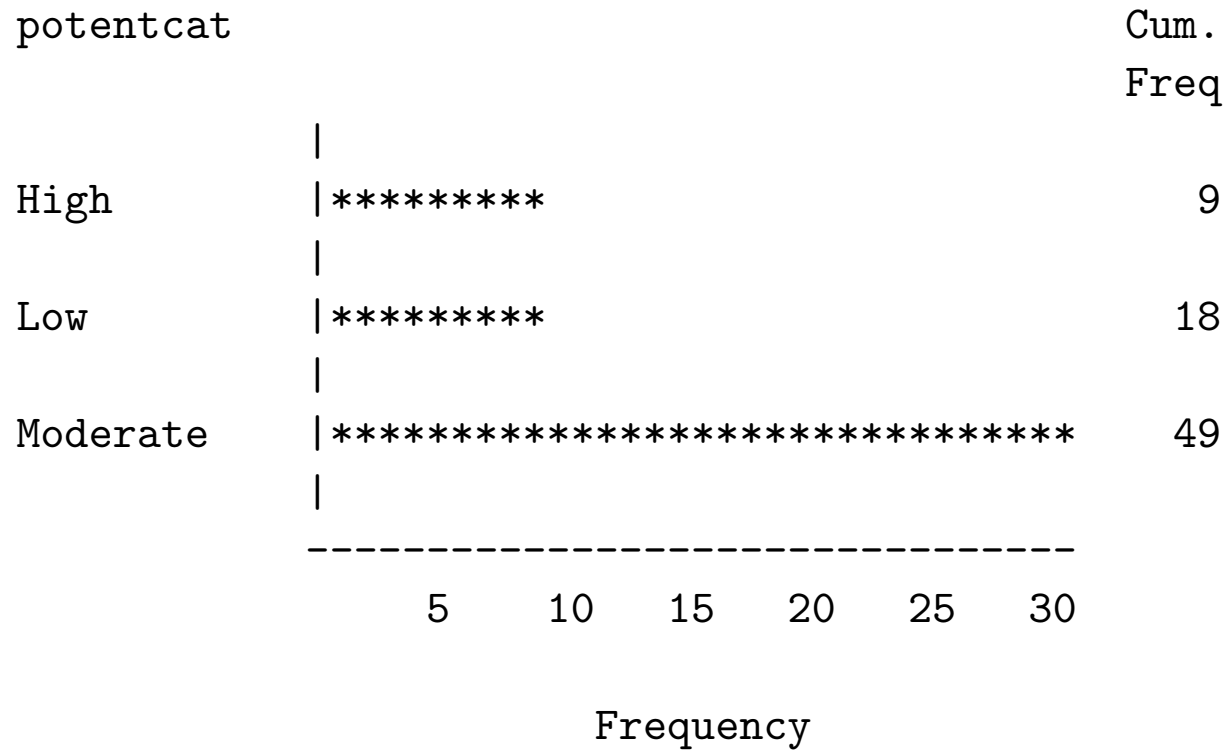
Breakdown of Region Potential Classes

potentcat



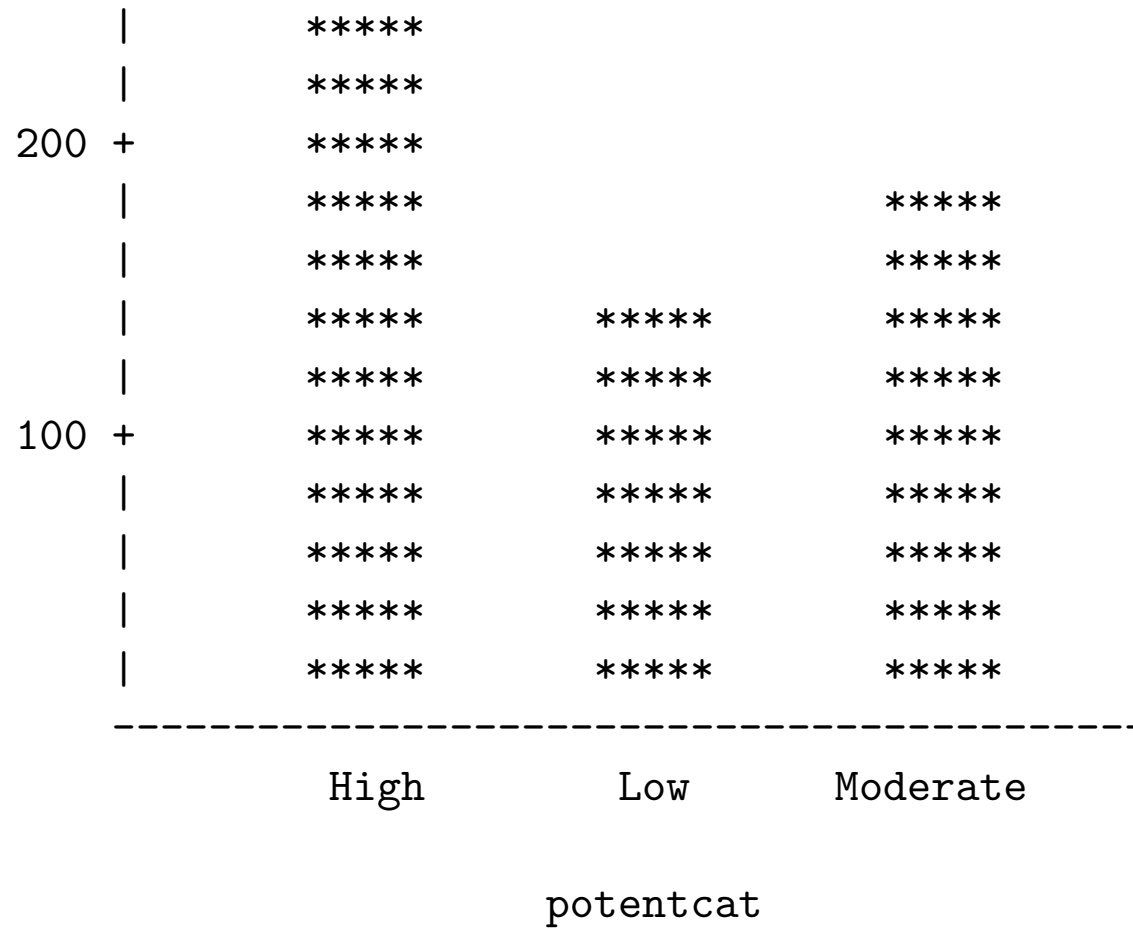
Frequency

Breakdown of Region Potential Classes



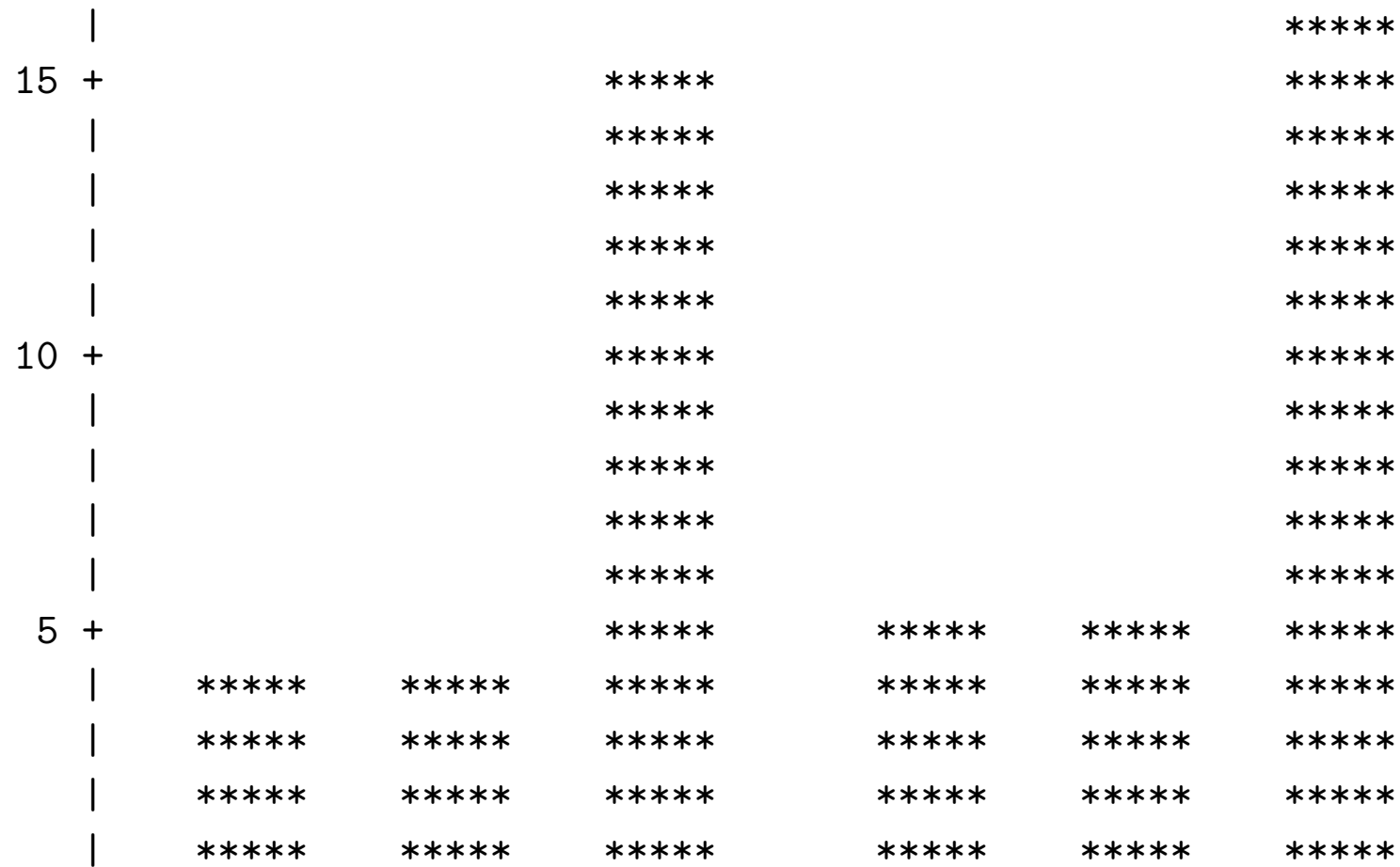
Breakdown of Region Potential Classes

sales Mean



Breakdown of Region Potential Classes

Frequency



High Low Moderate High Low Moderate potentcat
 |----- Test -----| |----- Training -----| training

- PROC GRAPH: Used for creating scatter plots and contour plots.

While PROC PLOT takes many options, it only has two basic statements

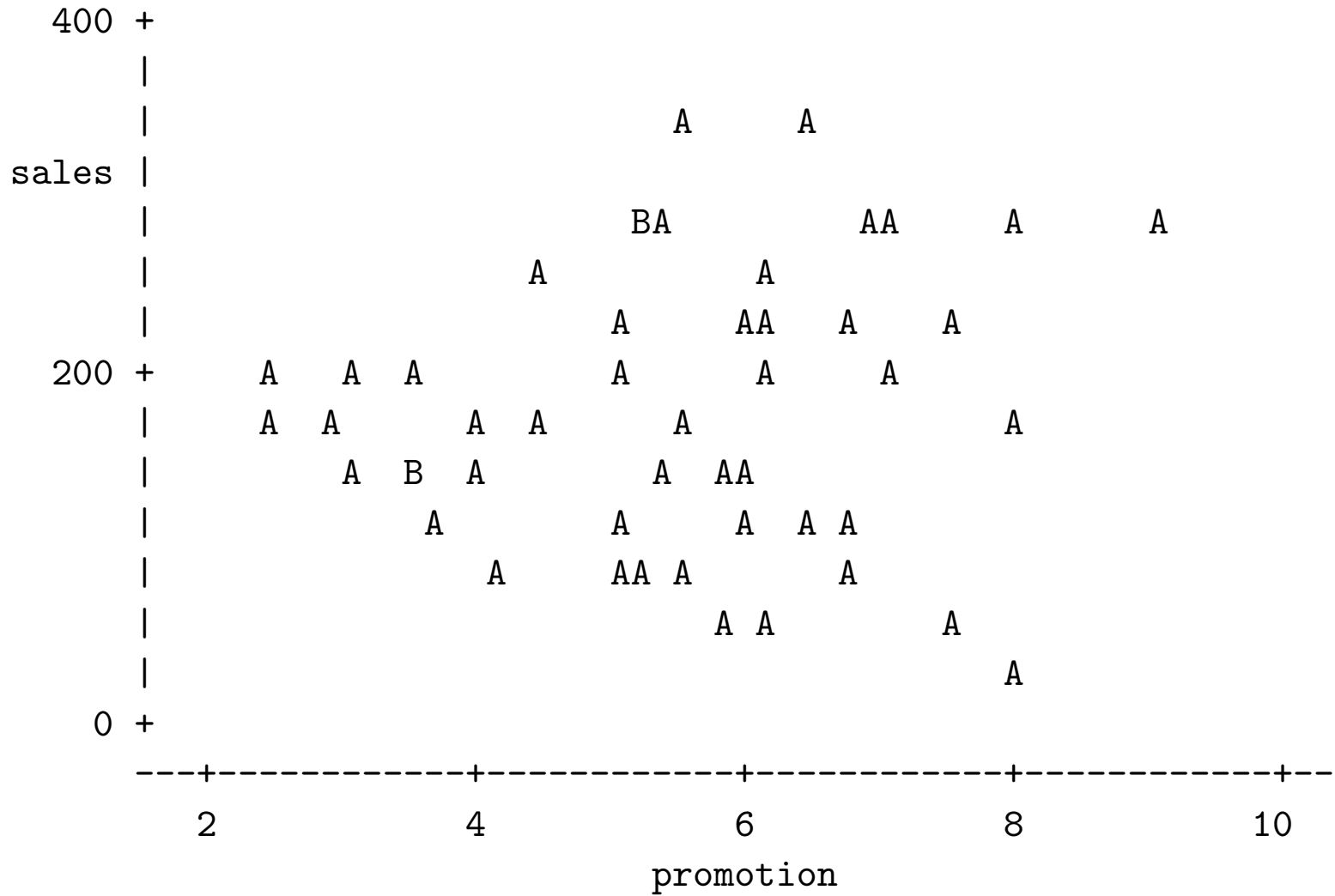
```
PROC PLOT <option(s)>;  
  BY <DESCENDING> variable-1 <...<DESCENDING> variable-n>  
    <NOTSORTED>;  
  PLOT plot-request(s) </ option(s)>;
```

BY works as it does in other PROC. As we saw in the examples, the basic form of the plot-request(s) is

```
PROC PLOT;  
  PLOT sales*promotion; /* y-var * x-var */  
  PLOT sales*brands='+'; /* uses + as plotting symbol */  
  PLOT sales*accounts=pot; /* uses potential levels for  
                             plotting symbol */
```

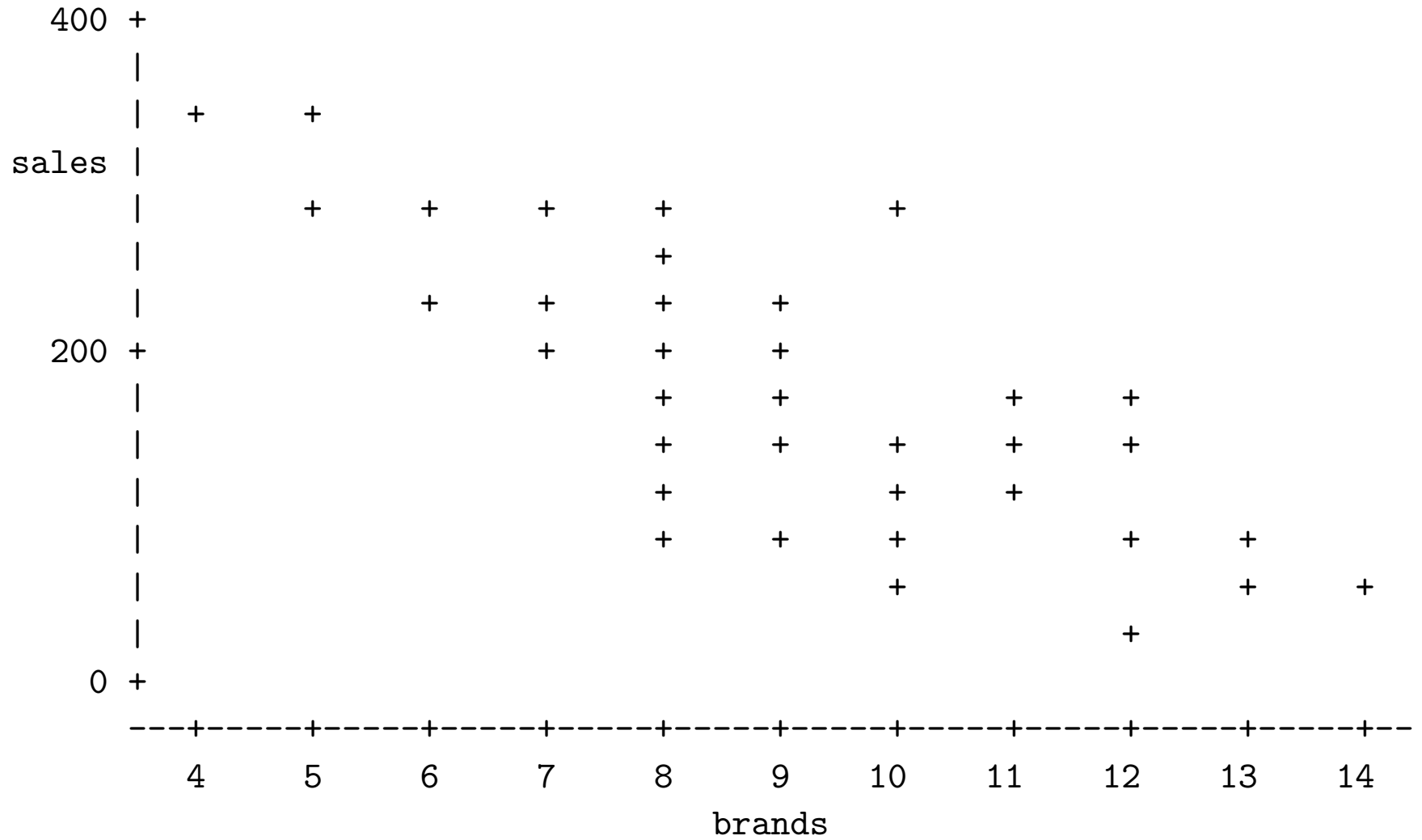
Sales Relationships

Plot of sales*promotion. Legend: A = 1 obs, B = 2 obs, etc.



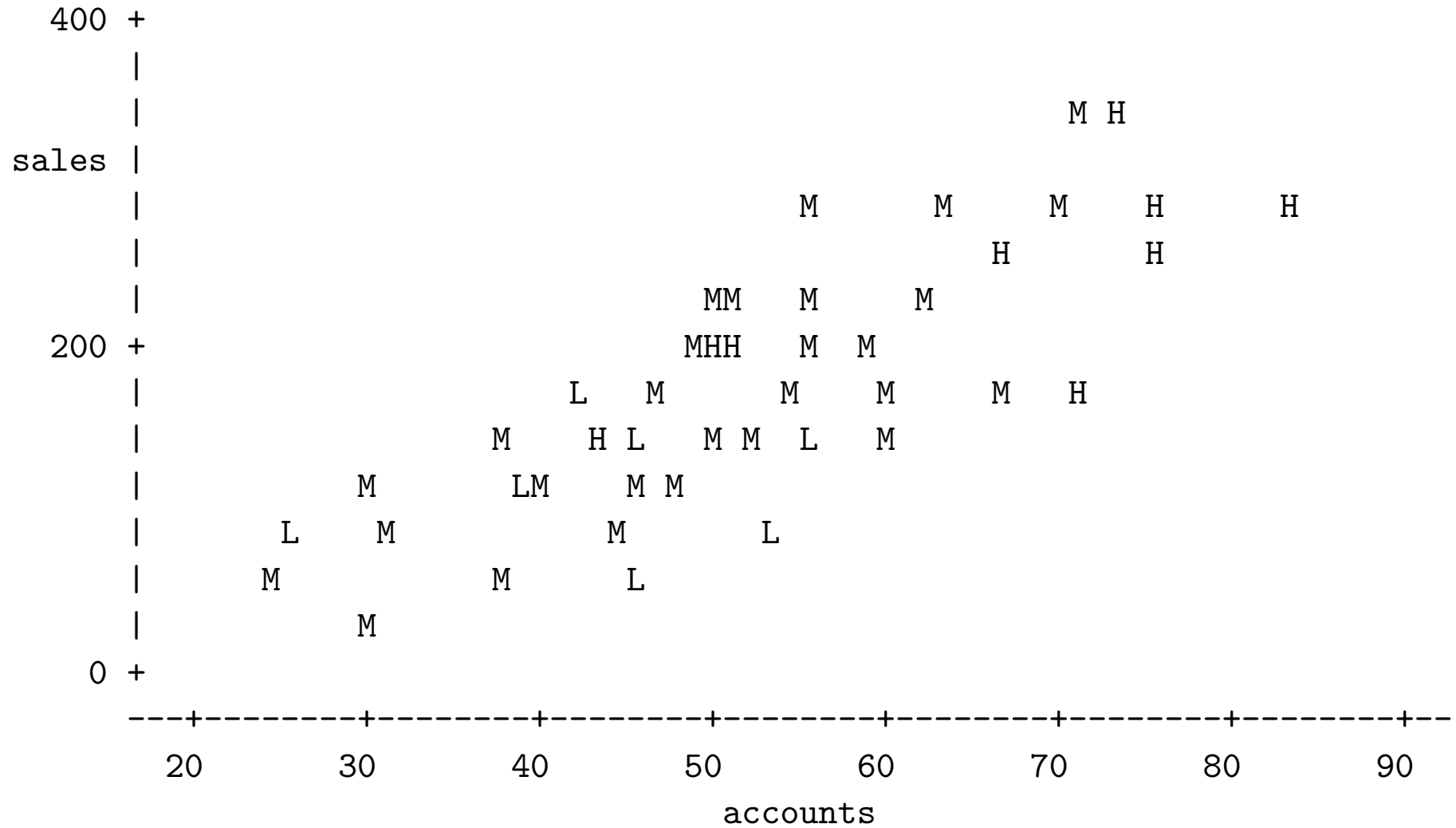
Sales Relationships

Plot of sales*brands. Symbol used is '+'.



Sales Relationships

Plot of sales*accounts. Symbol is value of pot.



High Resolution Graphics in SAS

As mentioned earlier, there are a number of different approaches that result in high resolution graphics in **SAS**. When looking for documentation for them with the built in (or web based) help system, you will need to look in a few places. Some are in BASE SAS (UNIVARIATE), SAS/STAT (BOXPLOT), and SAS/GRAPH (GPLOT and GCHART) [plus probably more].

The basic structure of creating graphs is similar to before. For example, PROC GCHART creates the same types of graphs as PROC CHART, plus a couple more. For example, the earlier code will work by changing PROC CHART to PROC GCHART. Of course PROC GCHART needs many more options, to control colour, fonts, etc. Similarly, PROC GPLOT does everything that PROC PLOT does, but a whole lot more (bubble plots, line plots, regression plots).

It is possible to store your plots in a variety of formats. These include

- Postscript (PS) and Encapsulated Postscript (EPS)
- Portable Document Format (PDF)
- Graphic Interchange Format (GIF)
- JPEG
- Tagged Image Format (TIFF)
- Window bitmap (BMP), Portable bitmap (PBM)
- Computer graphics metafile (CGM). I believe that this is related to Windows metafile. It is recommended if you want to use your graphs with programs such as Excel, Word, or Word Perfect as it is vector based (meaning better quality graphics) and smaller files.

To see what formats are available on your setup, give the command

```
proc gdevice catalog=sashelp.devices;
```

This will bring up an window listing all of the possible graphical devices available. You will probably have to close this window before you can do anything else in **SAS**.

One way to create these files in **SAS** is to generate the graphs with your **SAS** program and then view each of them via the results window. When a graph is the active object, choose Export under the file menu, select the format you wish to save the graph as, and then save it. It is also possible to directly

- create one file that contains one graph
- create one file that contains multiple graphs
- create multiple files that contain one graph

(We'll get back to this later. Lets look at basic graph creation first.)

In addition, it is possible to create web presentations that can be displayed via an ActiveX control, a Java applet, or a static graph (GIF, JPEG, or PNG).

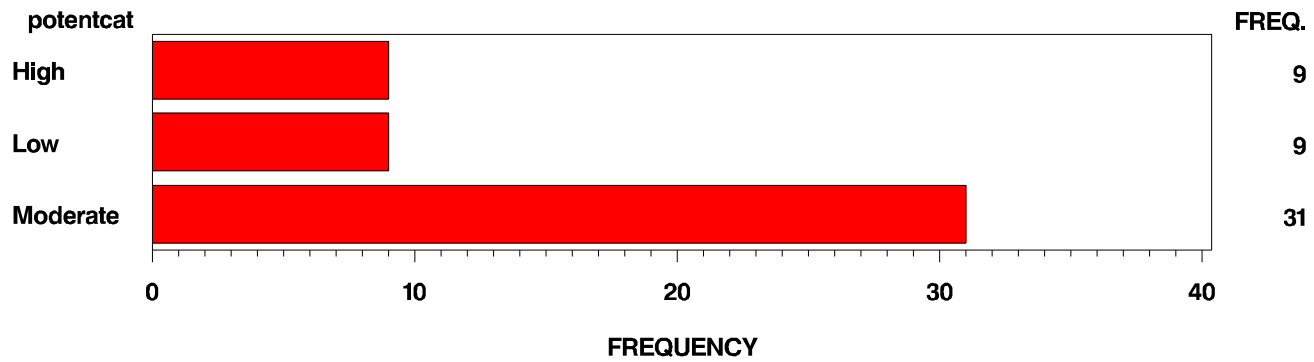
PROC GCHART

Lets start with an example, based on what we did before.

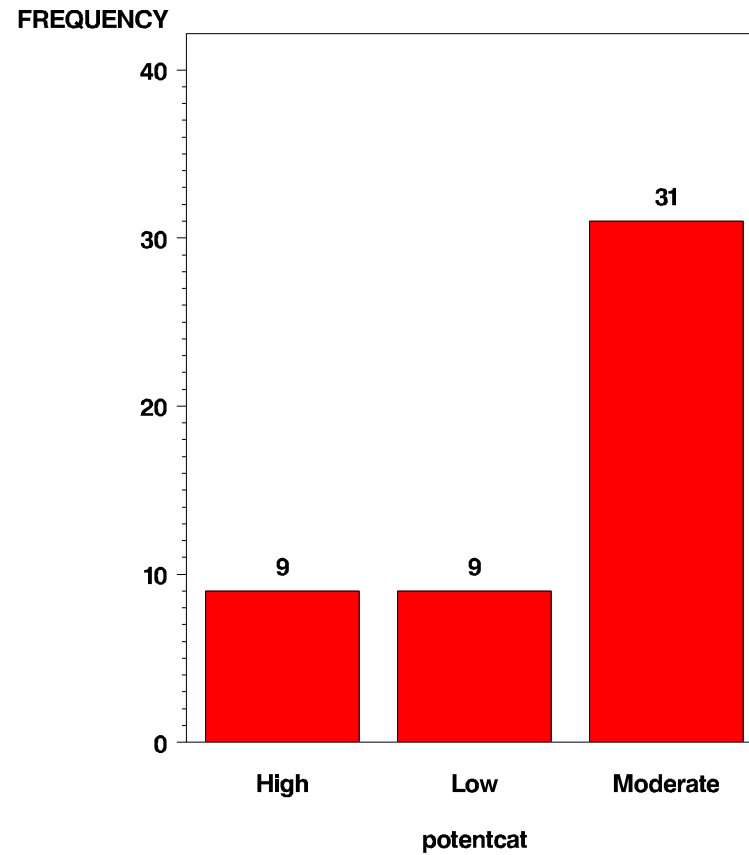
```
PROC GCHART DATA = shingles2;  
  
    TITLE 'Breakdown of Region Potential Classes';  
    FORMAT training $strainfmt. pot potfmt. ;  
  
    HBAR potentcat / FREQ;  
    VBAR potentcat / FREQ;
```

Note I've got a bit of a problem in creating postscript files with valid bounding boxes right now is **SAS**. This is why the graphs not fit into the document nicely. Will figure this out for next class.

Breakdown of Region Potential Classes



Breakdown of Region Potential Classes



The basic structure of PROC GCHART is as follows

```
PROC GCHART <DATA=input-data-set>
    <ANNOTATE=Annotate-data-set>
    <GOUT=<libref.>output-catalog>
    <IMAGEMAP=output-data-set>;
BLOCK chart-variable(s) </ option(s)>;
HBAR | HBAR3D | VBAR | VBAR3D chart-variable(s) </ option(s)>;
PIE | PIE3D | DONUT chart-variable(s) </ option(s)>;
STAR chart-variable(s) </ option(s)>;
```

Block charts are a variation on a 3-D bar chart. However as they don't have use axes, it is claimed that they are more useful when relative magnitude is more significant than exact magnitude of any particular block. Donut charts are like pie charts except they have a hole in the middle where text can be added.

The formats of graphs created in PROC GCHART are highly customizable. We'll talk about those during the discussion of GPLOT.

PROC GPLOT

This PROC will do a wide range of bivariate graphs including

- Scatter plots
- Bubble plots (scatter plots where the size of the plotting points has meaning)
- Line plots
- Regression plots (scatter plots with added regression information - fitted line, confidence & prediction bounds, etc)

Lets look at a couple of examples

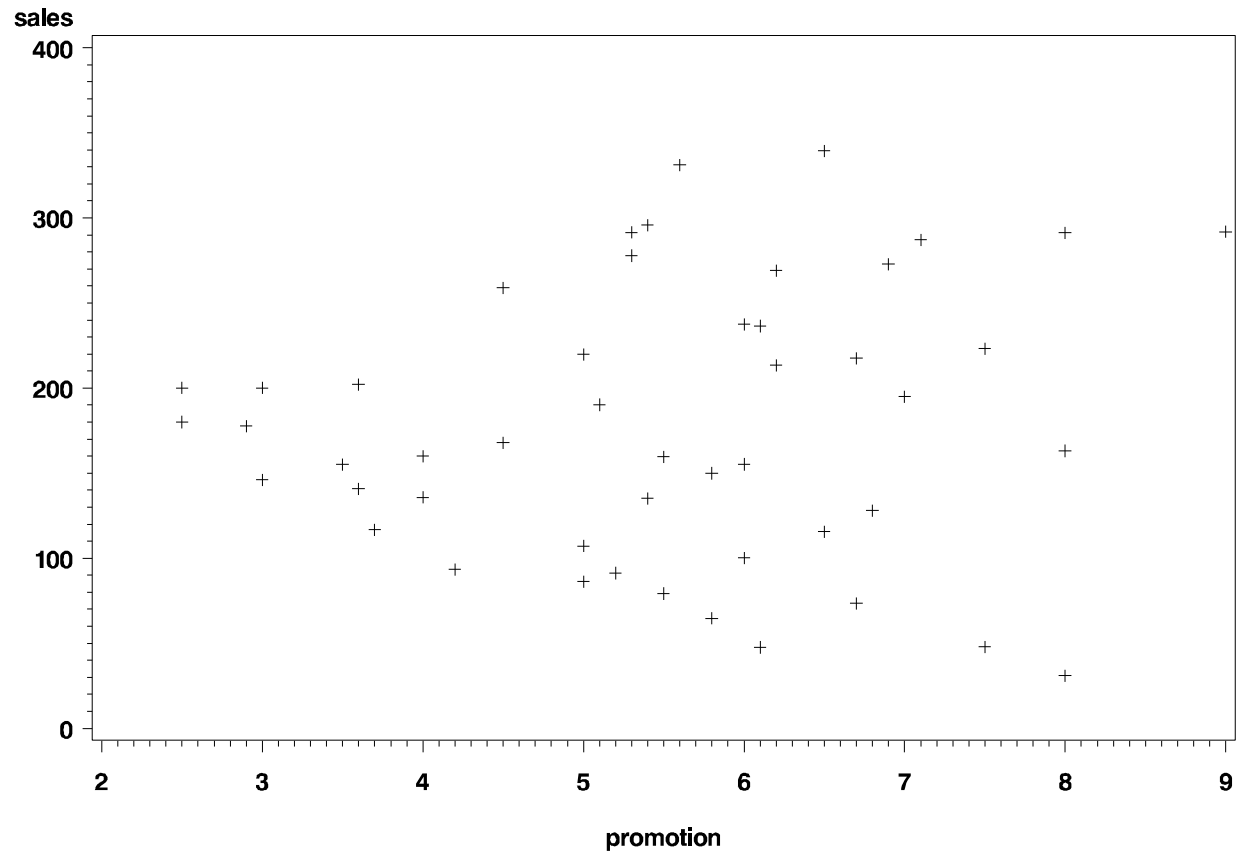
```
PROC GPLOT DATA = shingles2;  
  PLOT sales*promotion;  
  TITLE 'Sales vs Promotion';  
RUN;
```

```
axis1 label = ("Promotion ($1000)");  
axis2 label = ( 'Sales' justify=right '(1000s of squares)' )  
  width = 3;  
footnote j=1 "Shingle Sales Data";
```

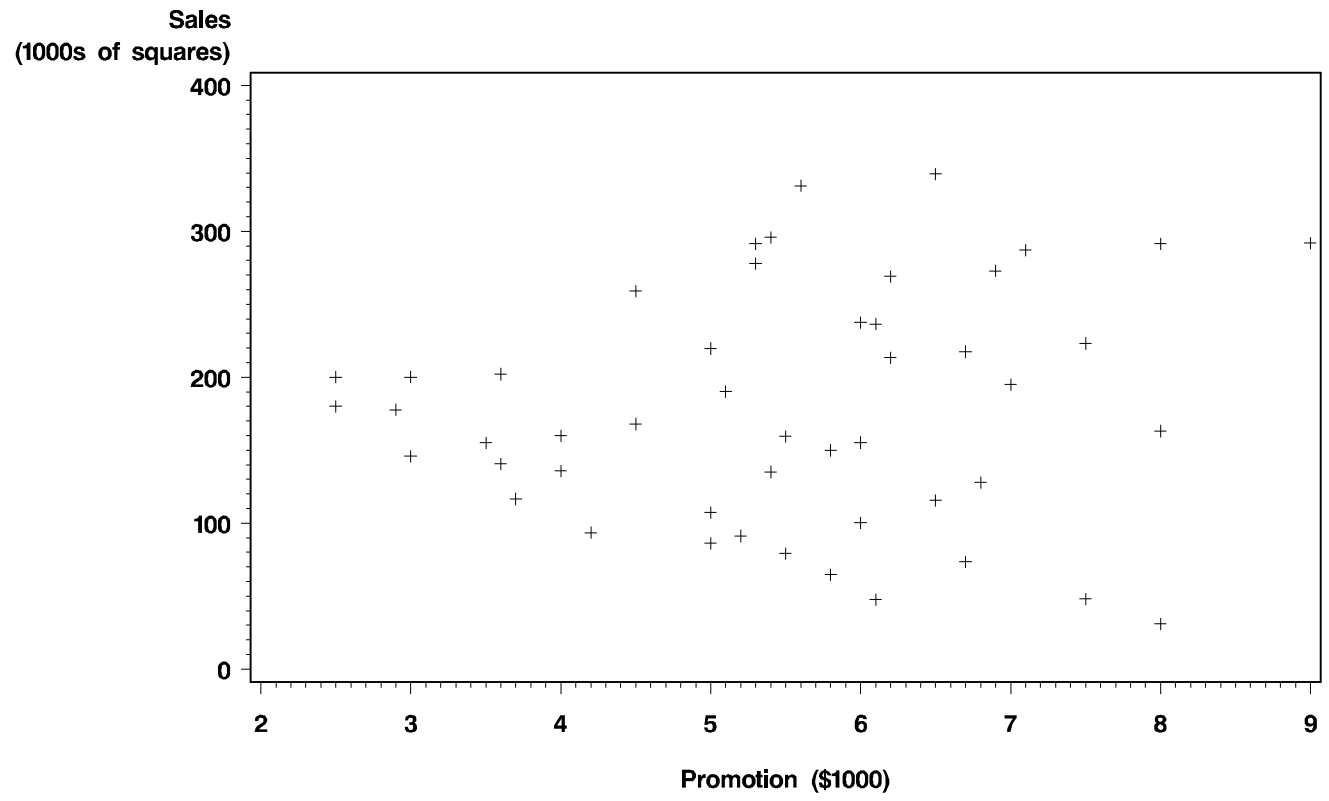
```
PROC GPLOT DATA = shingles2;  
  PLOT sales*promotion / HAXIS=axis1  
                        VAXIS=axis2;  
  TITLE 'Sales vs Promotion';  
RUN;
```

The first version will give the default plot. The second version gives a modified version with better axis labels and a footnote.

Sales vs Promotion



Sales vs Promotion



Shingle Sales Data

```
PROC GPLOT <DATA=input-data-set>  
    <ANNOTATE=Annotate-data-set>  
    <GOUT=<libref.>output-catalog>  
    <IMAGEMAP=output-data-set >  
    <UNIFORM>;
```

```
BUBBLE plot-request(s) </option(s)>;  
BUBBLE2 plot-request(s) </option(s)>;  
PLOT plot-request(s) </option(s)>;  
PLOT2 plot-request(s) </option(s)>;
```

Of the GPLOT options, DATA and UNIFORM are the important ones. UNIFORM specifies that the same axis scaling is used by all plots generated by the PROC, instead of using different values for each plot (based on the min and max values in the plot).

It is possible generate figures with involving two sets of y-axes (with PLOT2 and BUBBLE2) such as plotting the Dow Jones and the S&P 500 against time on the same plot. You wouldn't want to use the same axes for each of them as the Dow and S&P values are taken on quite different ranges.

The general form of the `plot-request` looks like one of

```
PLOT y-var * x-var  
PLOT y-var * x-var = n  
PLOT (y-vars) * (x-vars)
```

The first version will give a default scatter plot. The second version allows for changing the plotting symbol. This could be done via symbol definition or with a third variable. The last version allows for generations of multiple plots where every y-var in the list is plotted against every x-var in the list.

The options for each of these allows for modifications of the axes, adding legends, plotting multiple series on the same plot, setting colours.

It is also possible to change things like fonts and their sizes but that is done differently which we'll get to another time.