Introduction to Statistical Graphics Procedures

Selvaratnam Sridharma

U.S. Census Bureau

Overview

Statistical graphic procedures

SGPLOT

SGPANEL

SGSCATTER

SGRENDER and

Graph Template Language (GTL) Templates

SGPLOT

Procedure to create single cell graphs

creates one plot, or more plots overlaid on a single set of axes

scatter plots, line plots, histograms, bar charts, regression plots

SGPANEL

A panel of graph cells

for values of one or more classification variables

Variety of plot types

like SGPLOT Procedure

SGSCATTER

Paneled graph of scatter plots for different combination of variables

Overlay fit plots and ellipses on scatter plots

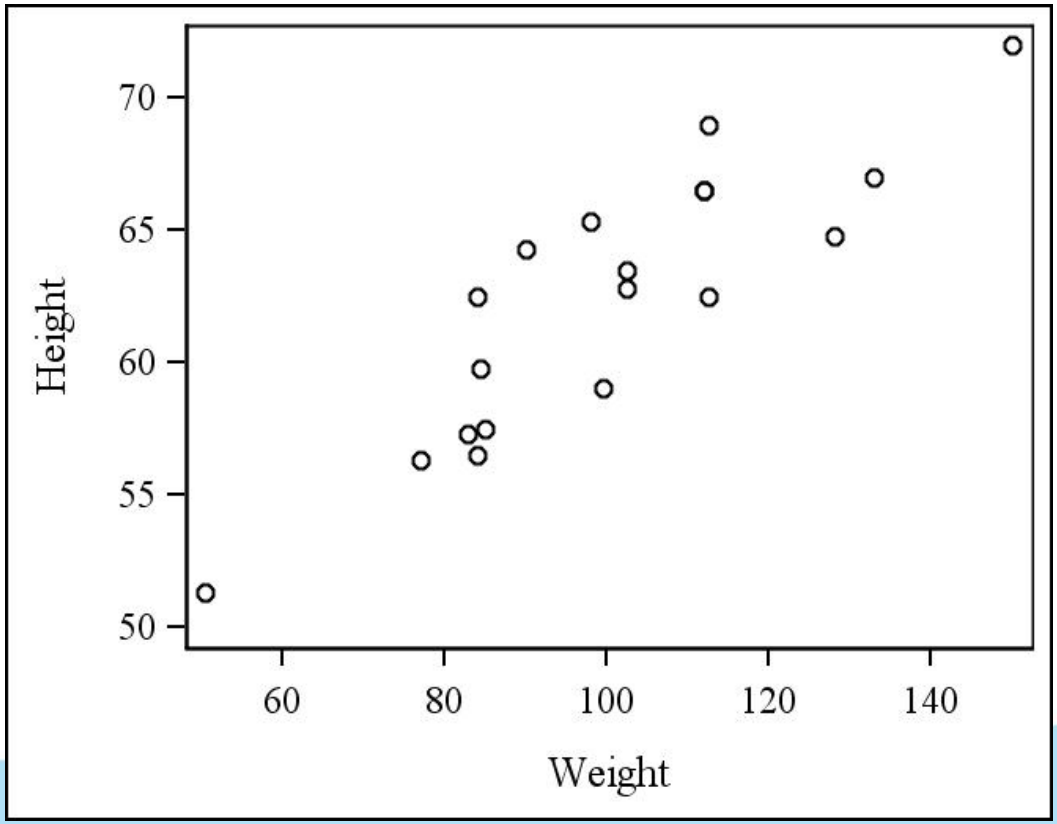
SGRENDER

Uses GTL templates

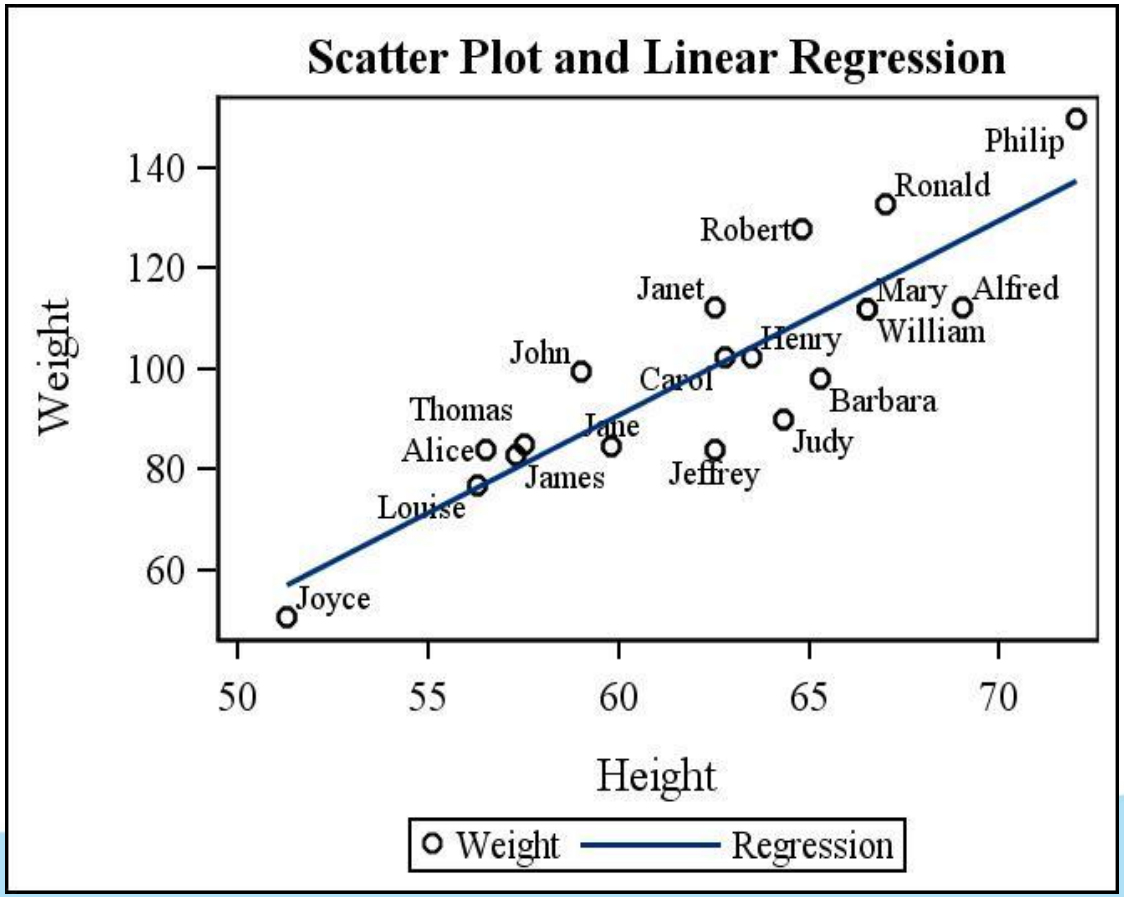
create customized layouts and graphs

SGPLOT

```
PROC SGPLOT DATA = Sashelp.Class;
   SCATTER X = Height  Y = Weight;
RUN;
```

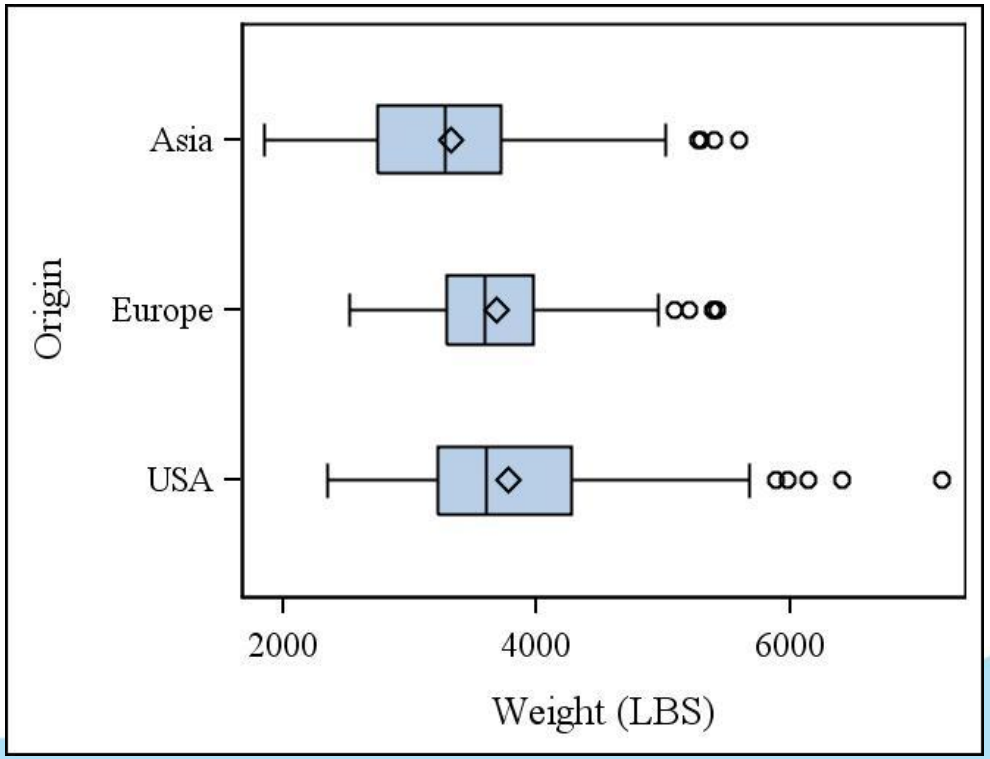


```
PROC SGPLOT DATA = Sashelp.Class;
    TITLE 'Scatter Plot and Linear Regression';
    SCATTER  X = Height  Y = Weight;
    REG X = Height  Y= Weight / DATALABEL = Name;
RUN;
```

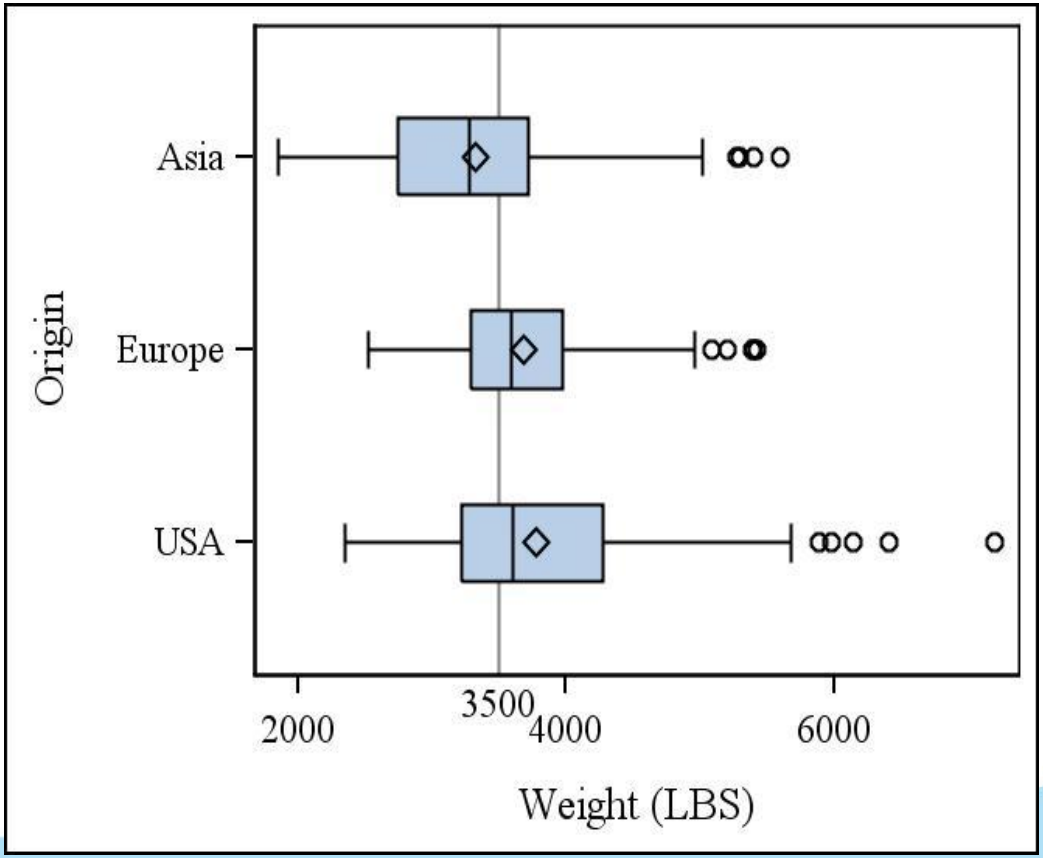


Box Plot

```
PROC SGPLOT DATA = Sashelp.Cars;
    HBOX Weight / CATEGORY = Origin;
RUN;
```

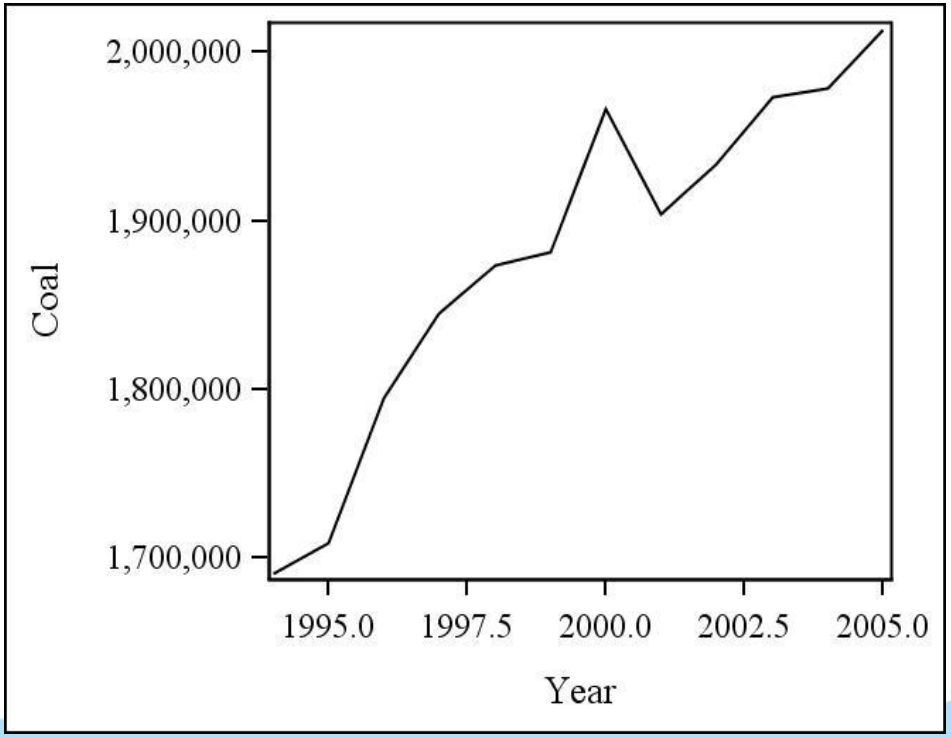


```
PROC SGPLOT DATA = Sashelp.Cars;
    REFLINE 3500 / AXIS = X LABEL = '3500';
    HBOX Weight / CATEGORY = Origin;
RUN;
```



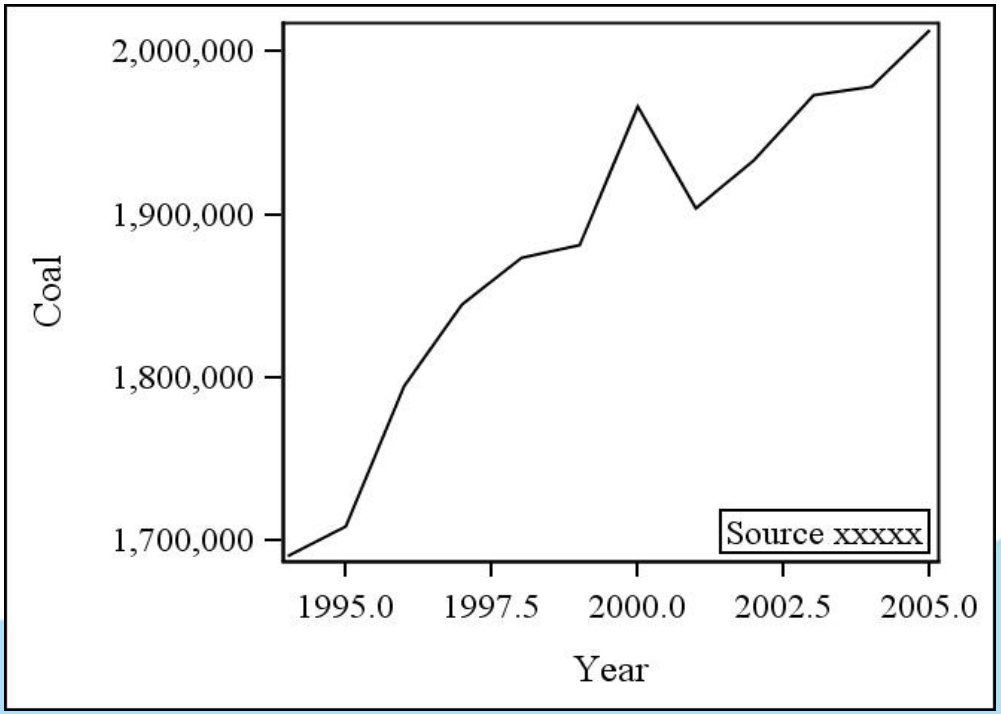
SERIES PLOT

```
PROC SGPLOT DATA = Sashelp.Electric (
    WHERE   = (Customer= "Residential"));
   SERIES    X = Year  Y = Coal;
RUN;
```

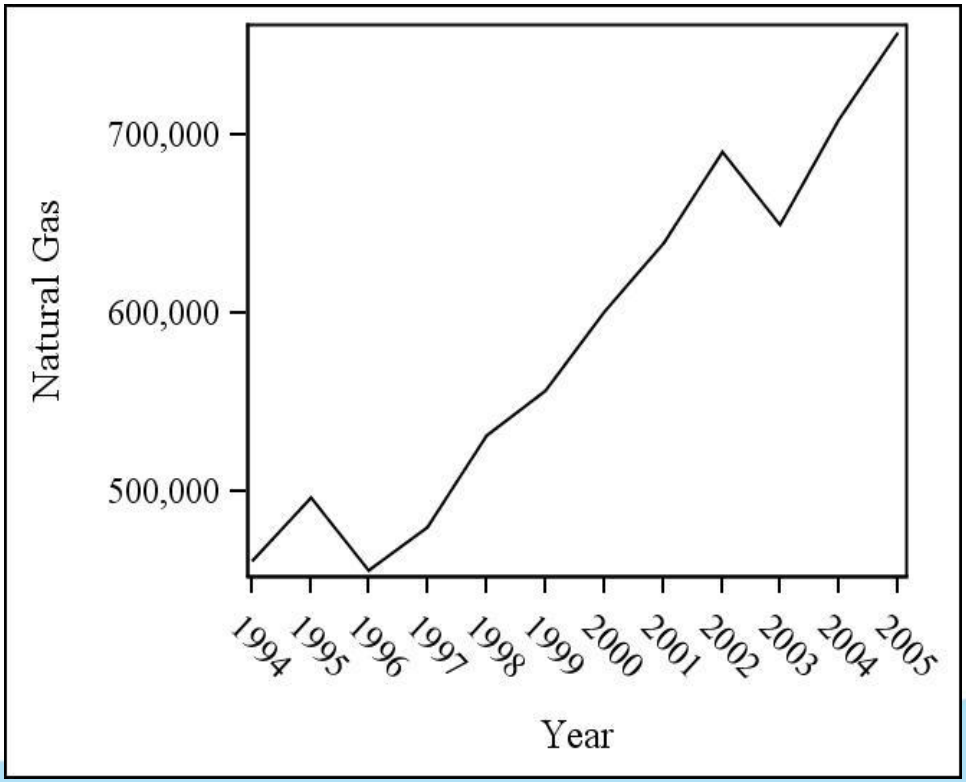


INSET STATEMENT

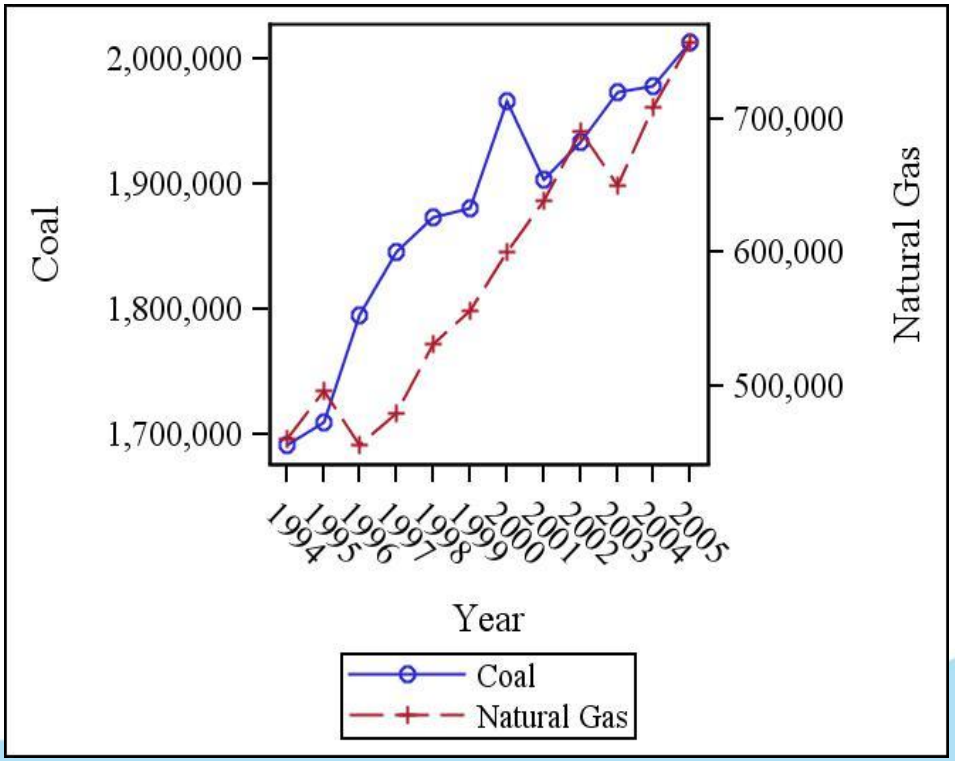
```
PROC SGPLOT  DATA  =  Sashelp.Electric (
          WHERE     =  (Customer= "Residential"));
    SERIES  X = Year     Y = Coal;
    INSET  'Source xxxxx'  /  POSITION  =  BOTTOMRIGHT
    BORDER;
RUN;
```



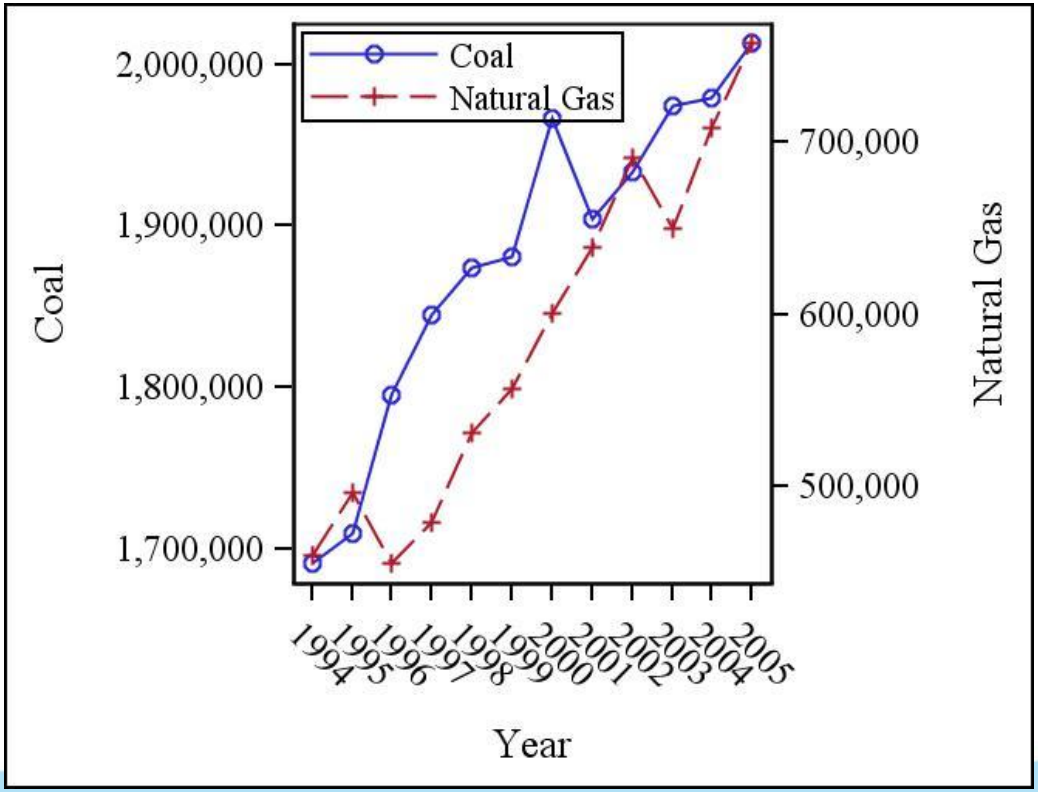
```
PROC SGPLOT DATA = Sashelp.Electric(
     WHERE = (Customer= "Residential"));
    XAXIS TYPE = DISCRETE;
    SERIES X = Year Y = Naturalgas;
RUN;
```



```
PROC  SGPLOT  DATA  =  Sashelp.Electric ( WHERE =
                                    (Customer= "Residential"));
XAXIS TYPE = DISCRETE;
SERIES X = Year  Y = Coal /  MARKERS;
SERIES X = Year  Y = Naturalgas  / MARKERS Y2AXIS;
RUN;
```

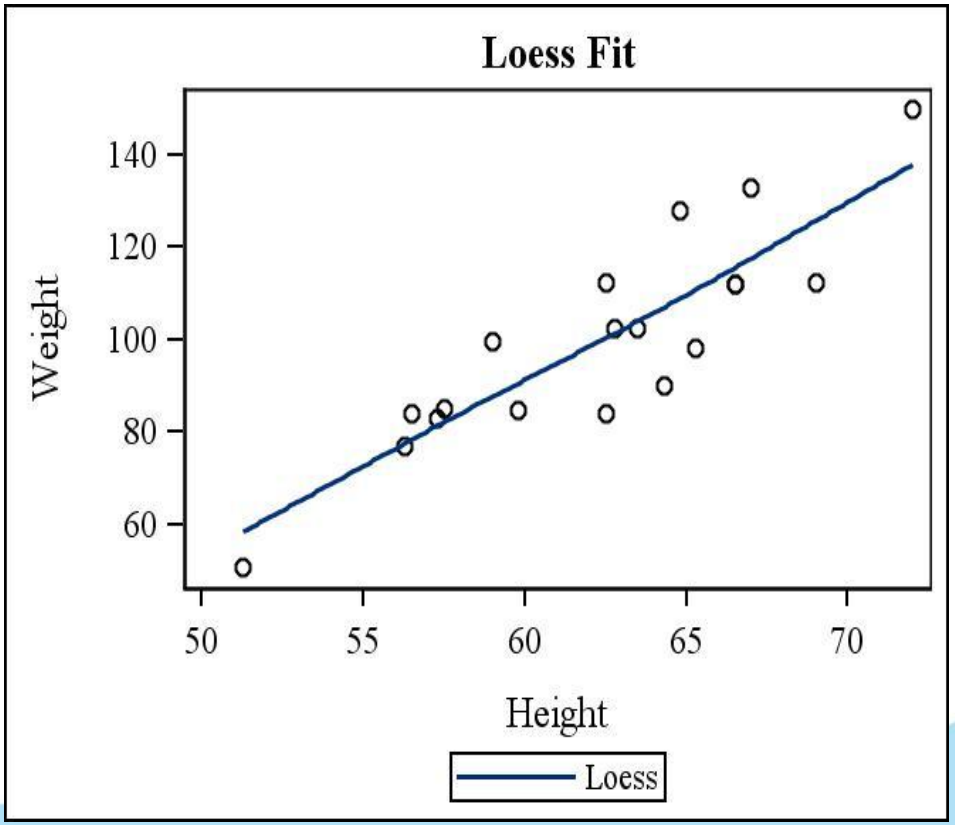


```
PROC SGPLOT DATA = Sashelp.Electric (WHERE=      (Customer =
    "Residential"));
    XAXIS TYPE = DISCRETE;
    SERIES X = Year  Y = Coal/ MARKERS;
    SERIES X = Year  Y =  Naturalgas / MARKERS Y2AXIS;
    KEYLEGEND  / LOCATION = INSIDE  POSITION = TOPLEFT;
RUN;
```



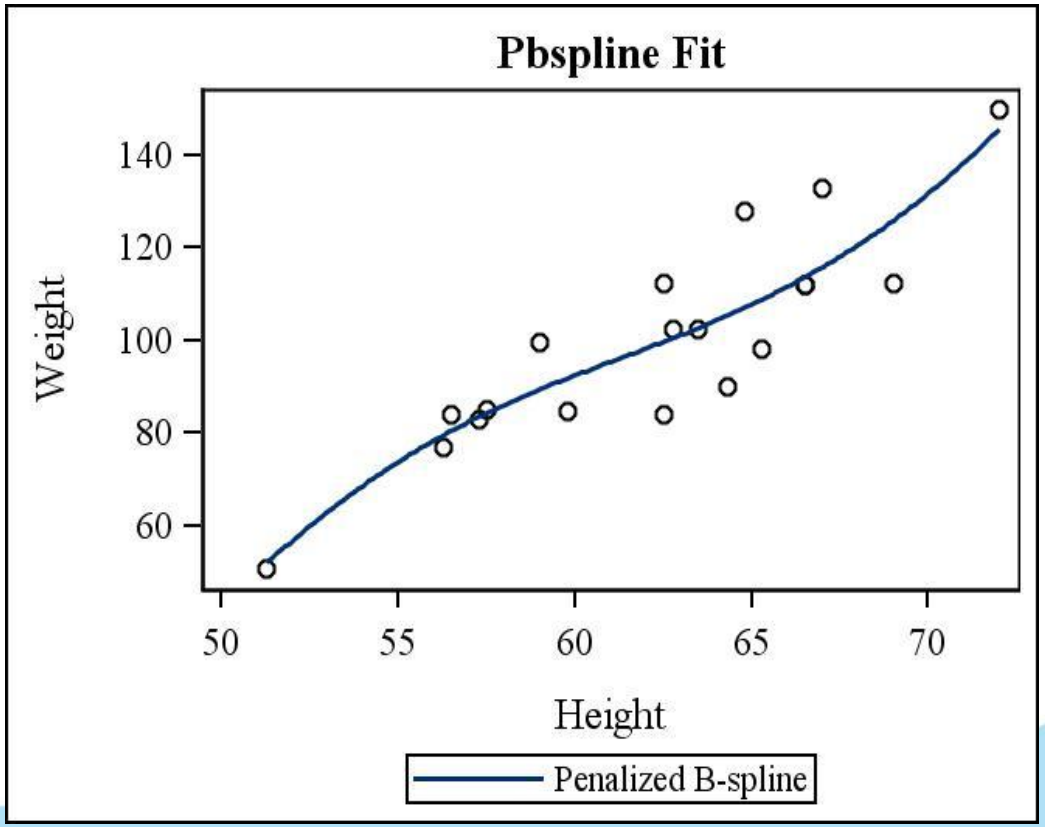
Loess Fit Plot

```
PROC SGPLOT  DATA  =  Sashelp.Class  ;
    TITLE  'Loess Fit' ;
    LOESS Y = Weight   X = Height;
RUN;
```



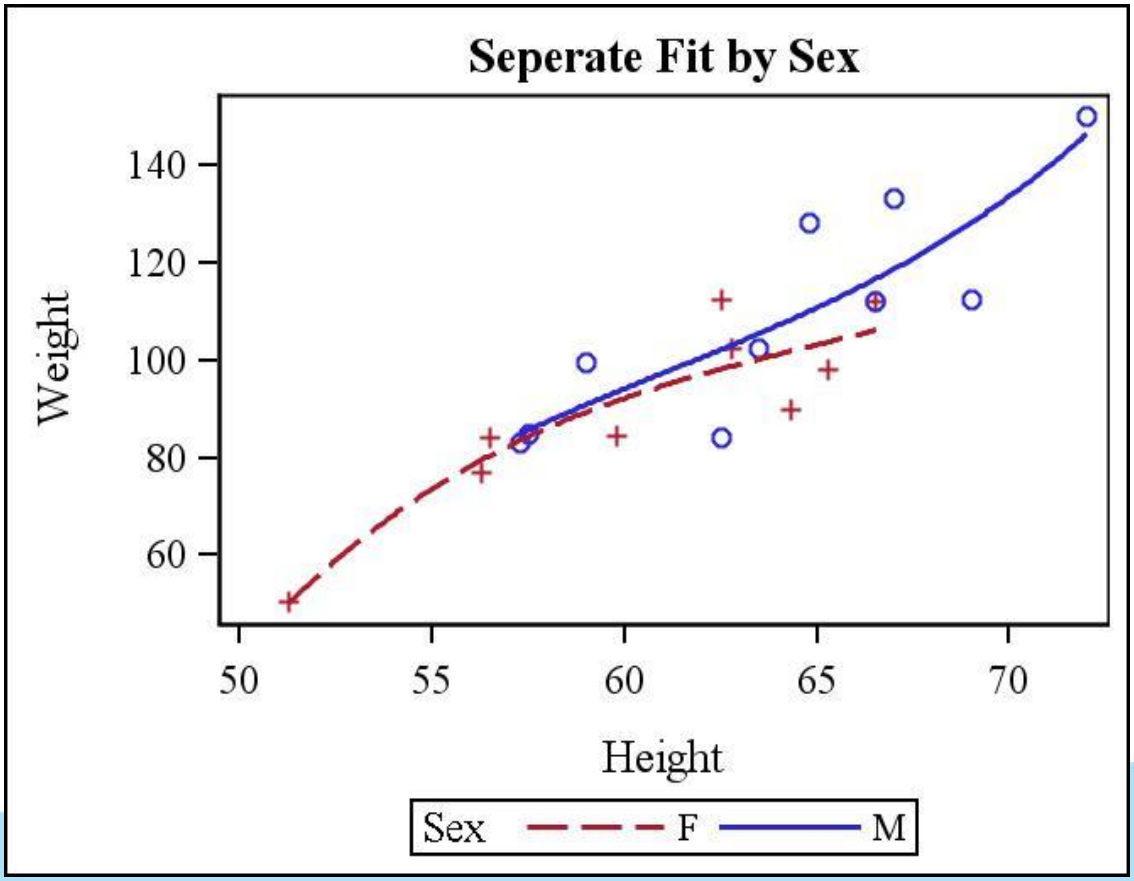
Penalized B-Spline

```
PROC SGPLOT DATA = Sashelp.Class ;
   TITLE 'Loess Fit' ;
   PBSPLINE Y = Weight  X = Height;
RUN;
```



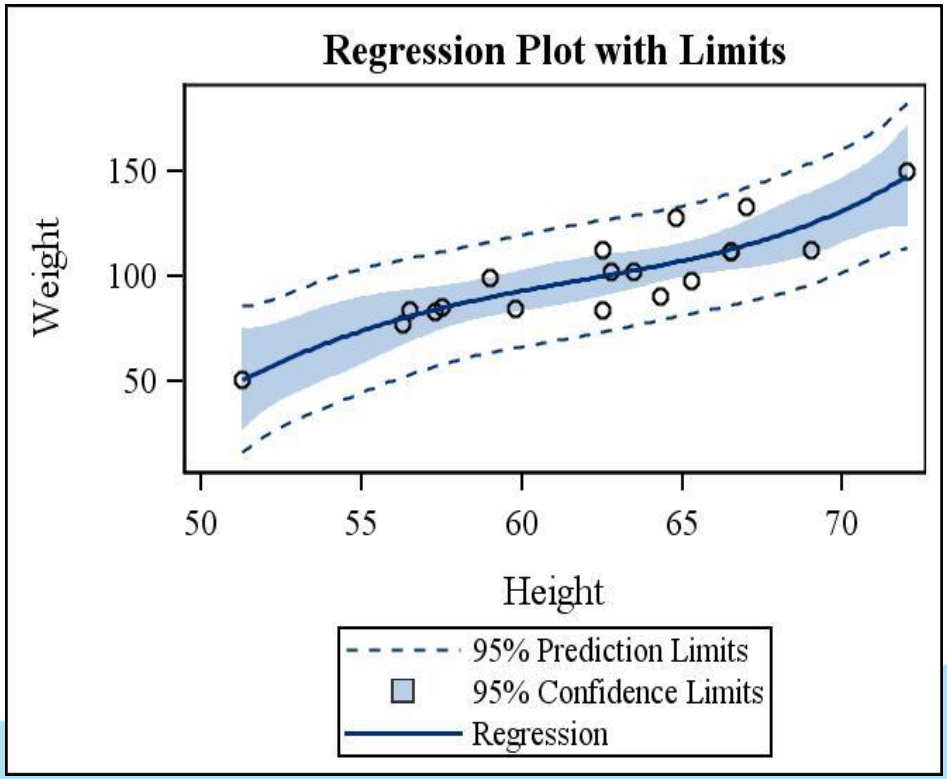
Grouped Regression Plot

```
PROC SGPLOT DATA  =  Sashelp.Class ;
    TITLE  'Separate Fit by Sex' ;
    REG  Y  = Weight  X  = Height  / GROUP  = Sex  DEGREE = 3;
RUN;
```



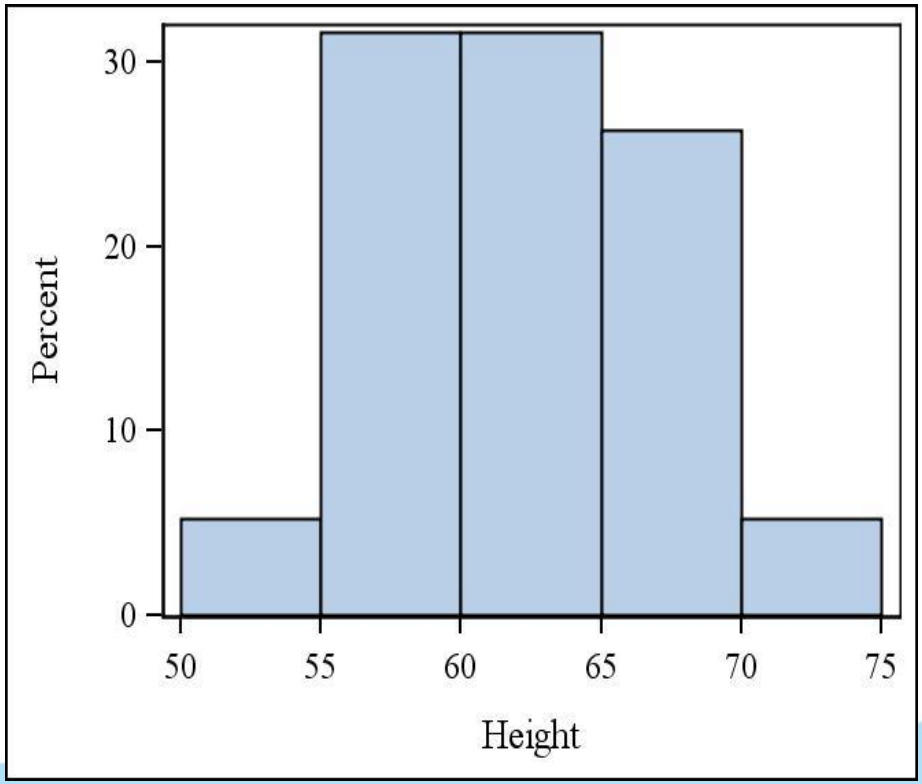
Regression Plot with Confidence and Prediction Limits

```
PROC SGPLOT DATA = Sashelp.Class ;
   TITLE 'Regression Plot with Limits' ;
   REG Y = Weight X = Height / DEGREE = 4 CLI CLM ;
RUN;
```

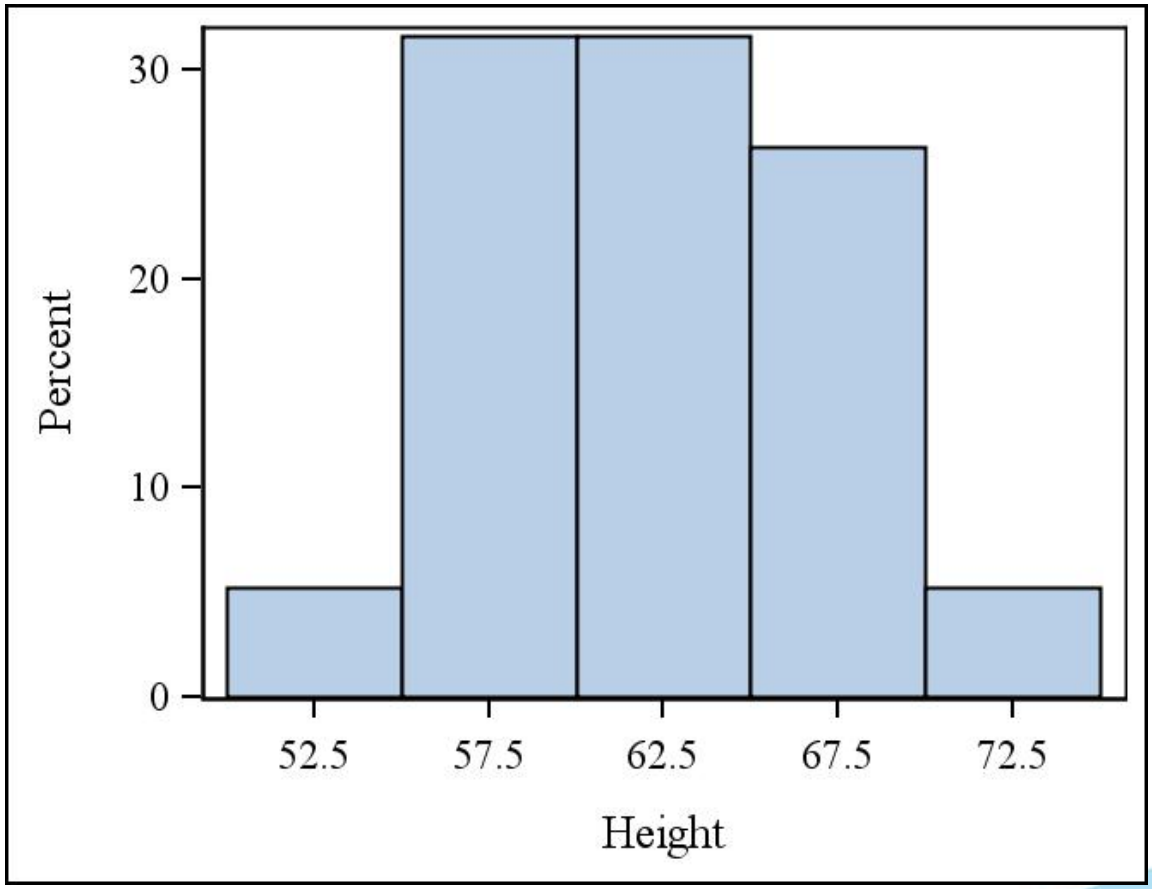


HISTOGRAM AND DENSITY CURVES

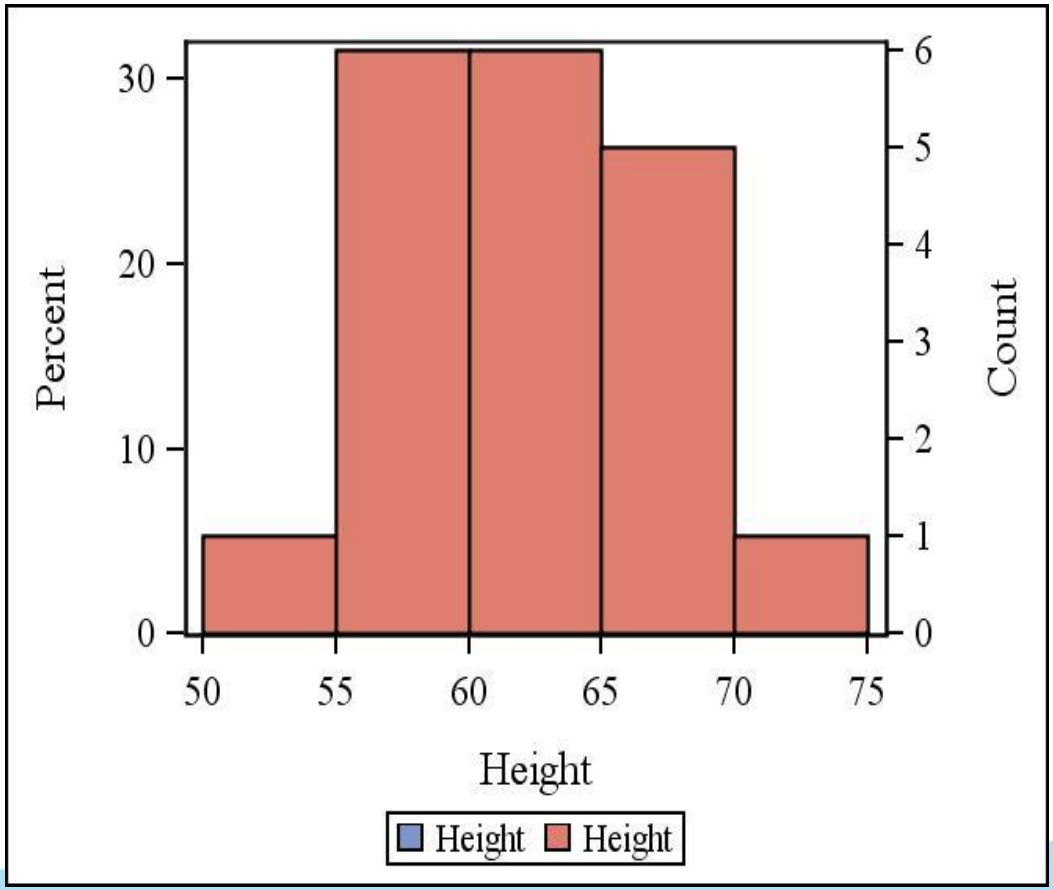
```
PROC SGPLOT DATA = Sashelp.Class;
   HISTOGRAM  Height;
RUN;
```



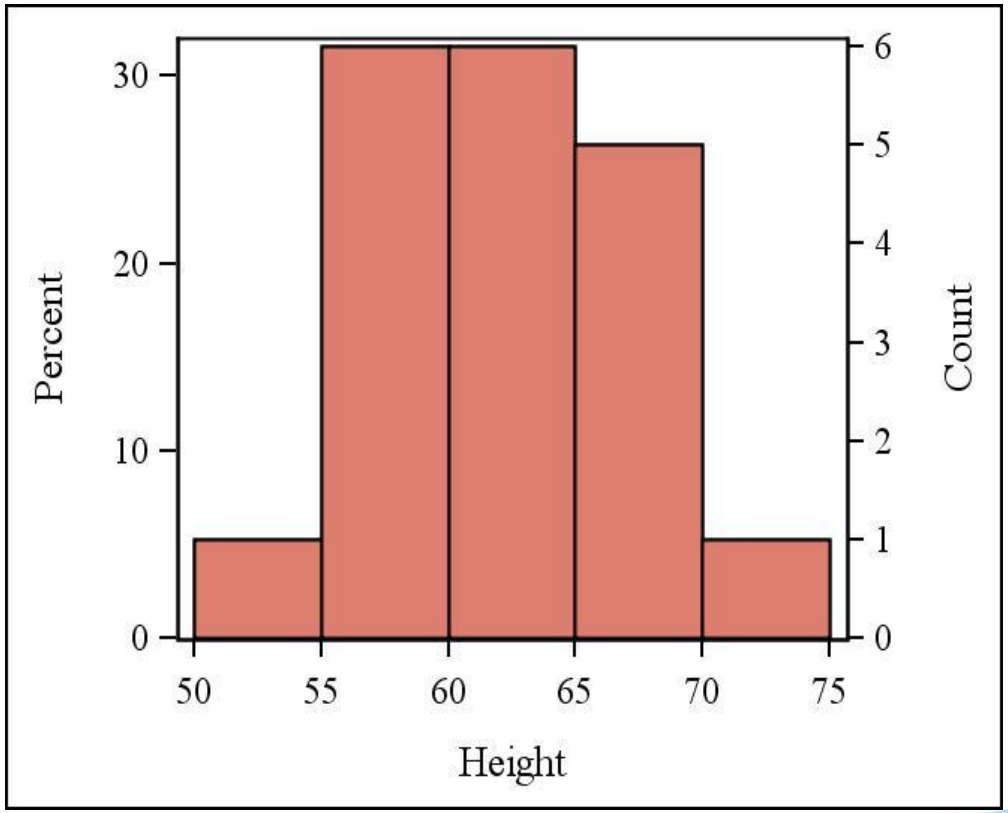
```
PROC SGPLOT DATA = Sashelp.Class;
   HISTOGRAM  Height  / SHOWBINS;
RUN;
```



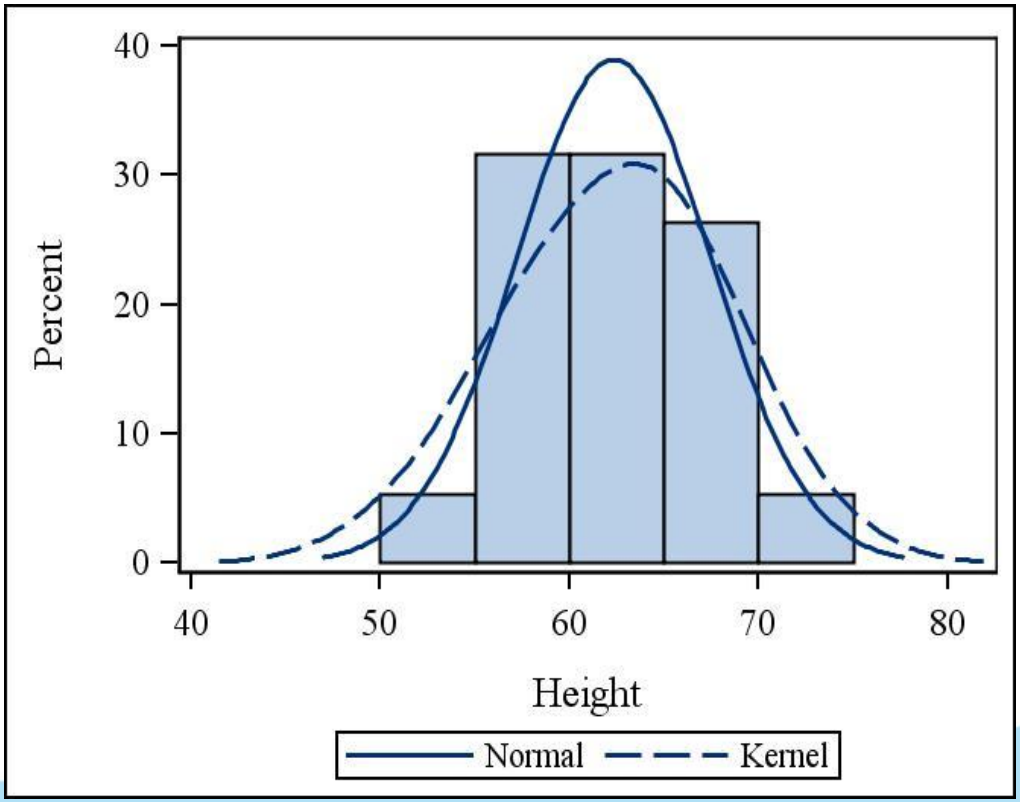
```
PROC SGPLOT DATA = Sashelp.Class;
   HISTOGRAM   Height;
   HISTOGRAM  Height / SCALE = COUNT Y2AXIS;
RUN;
```



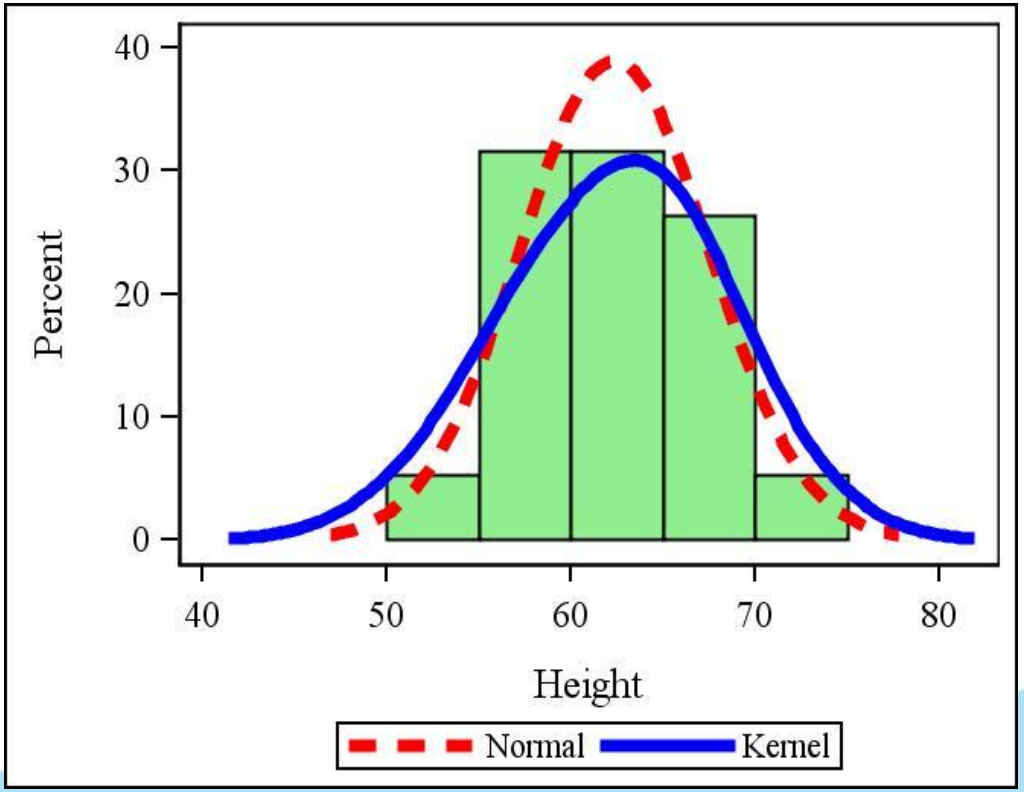
```
PROC SGPLOT DATA = Sashelp.Class  NOAUTOLEGEND;
    HISTOGRAM   Height;
    HISTOGRAM  Height  / SCALE= COUNT Y2AXIS;
RUN;
```



```
PROC  SGPLOT  DATA = Sashelp.Class;
    HISTOGRAM  Height;
    DENSITY  Height;
    DENSITY Height / TYPE = KERNEL;
RUN;
```

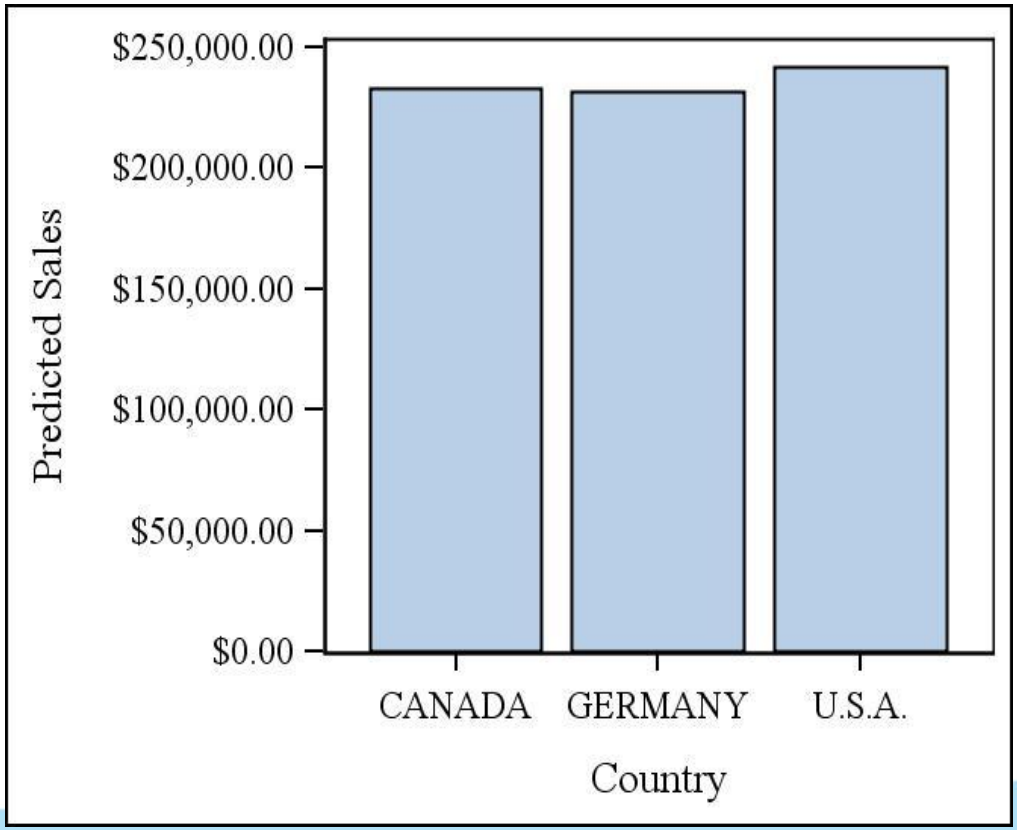


```
PROC SGPLOT DATA = Sashelp.Class;
    HISTOGRAM  Height  / FILLATTRS  =  (COLOR = LIGHTGREEN);
    DENSITY  Height  /  LINEATTRS = (COLOR=RED  PATTERN  =
          DASH THICKNESS =  5);
    DENSITY Height / TYPE = KERNEL LINEATTRS = (COLOR  =
          BLUE PATTERN  =  SOLID      THICKNESS  =  5);
RUN;
```

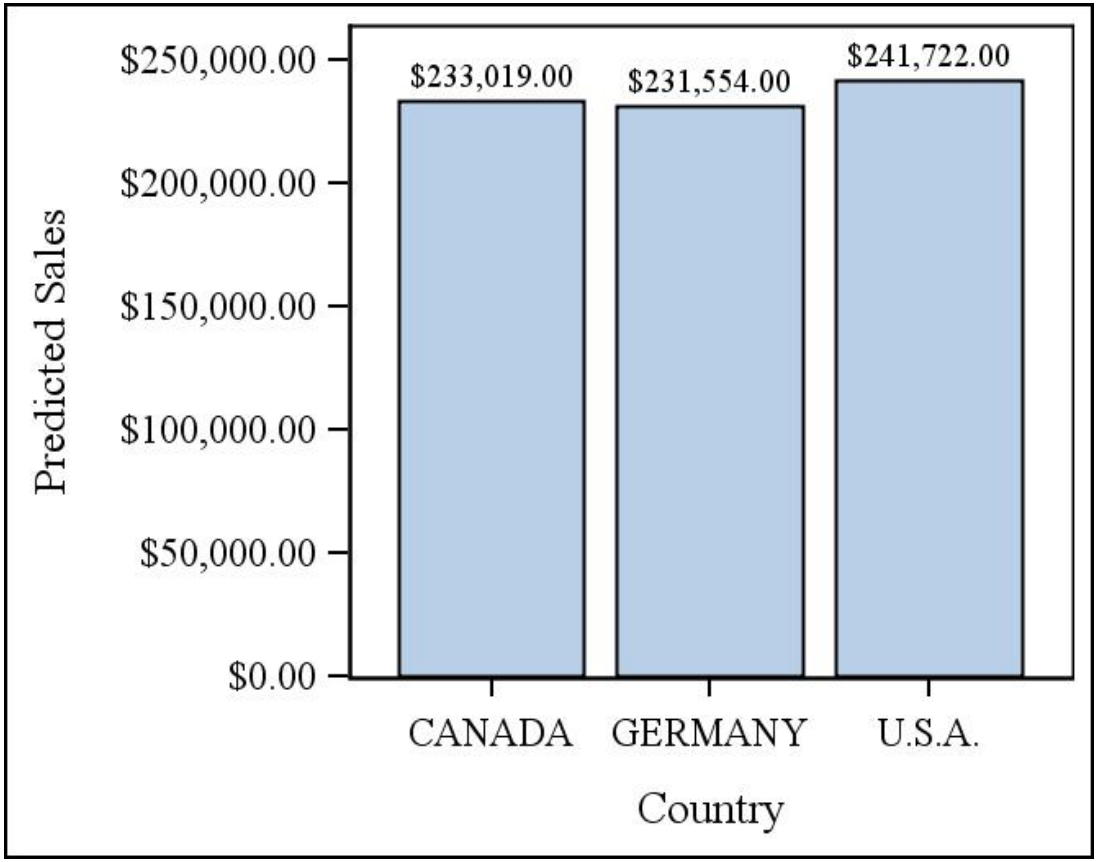


BAR CHART

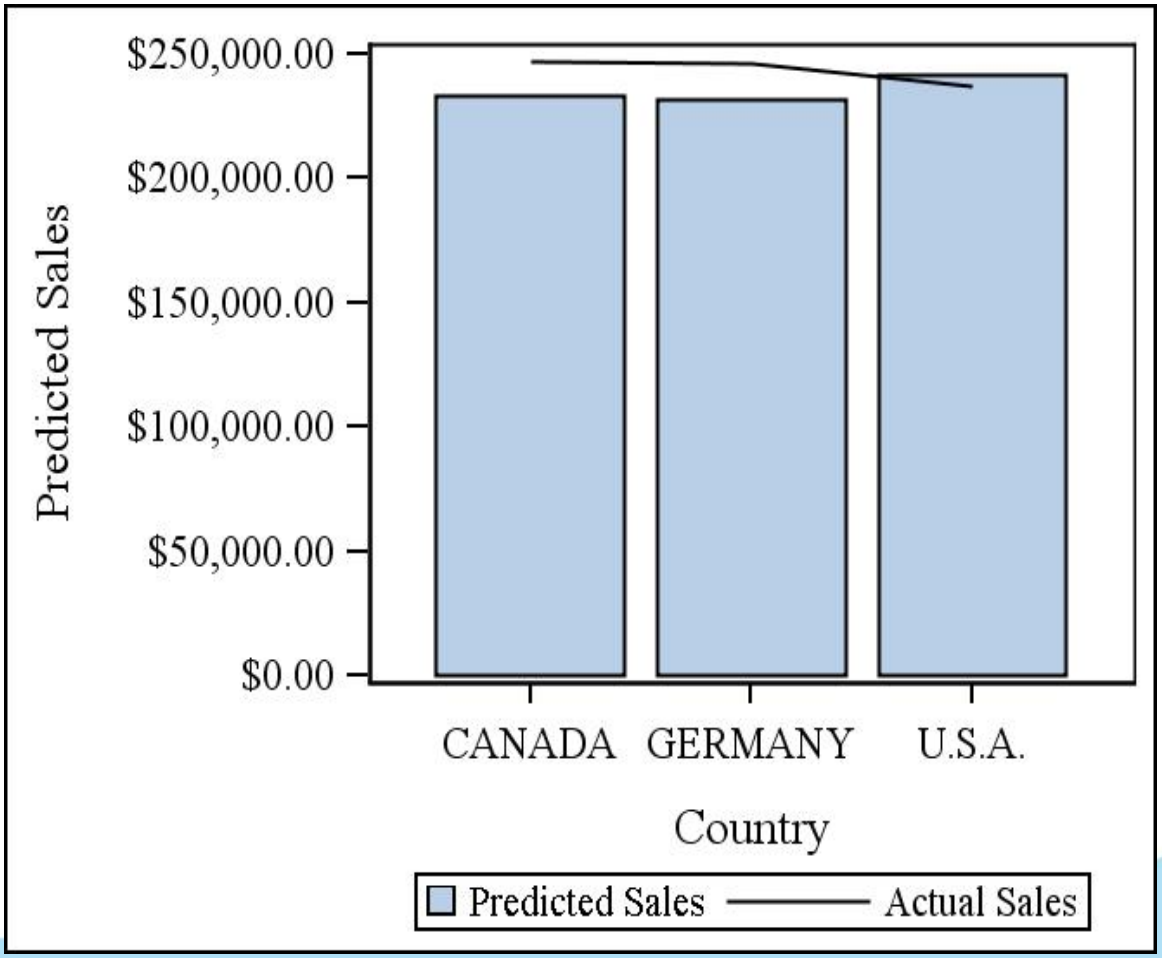
```
PROC SGPLOT DATA = Sashelp.Prdsale;
    VBAR   Country / RESPONSE= Predict;
RUN;
```



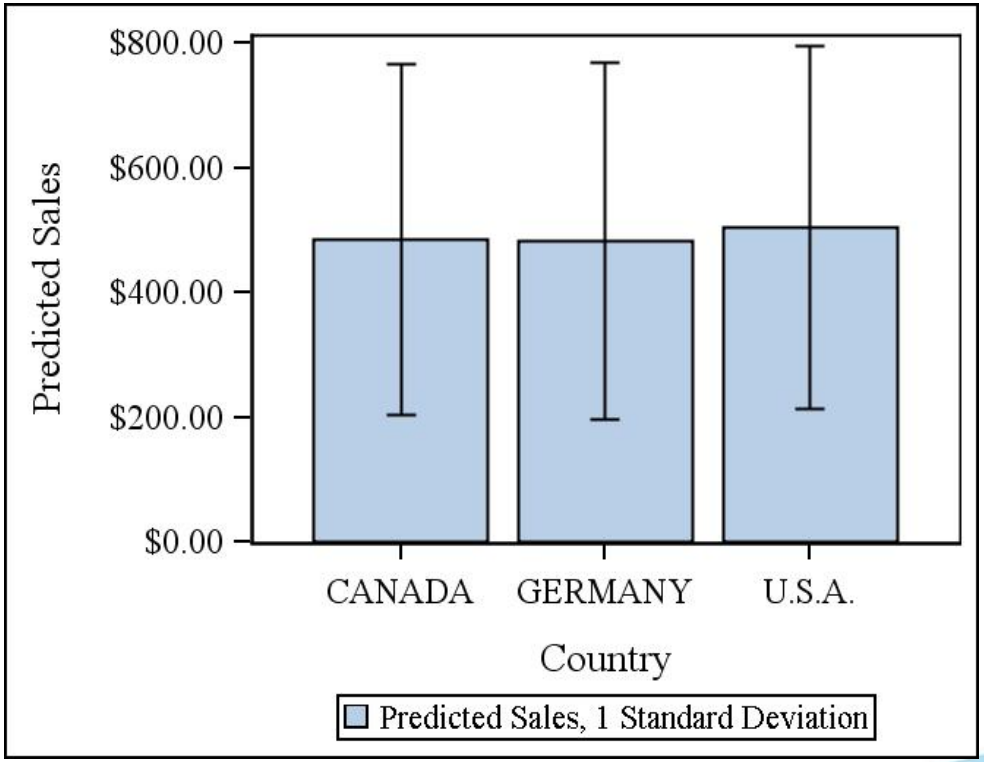
```
PROC SGPLOT DATA = Sashelp.Prdsale;
    VBAR   Country/ RESPONSE= Predict
           DATALABEL;
RUN;
```



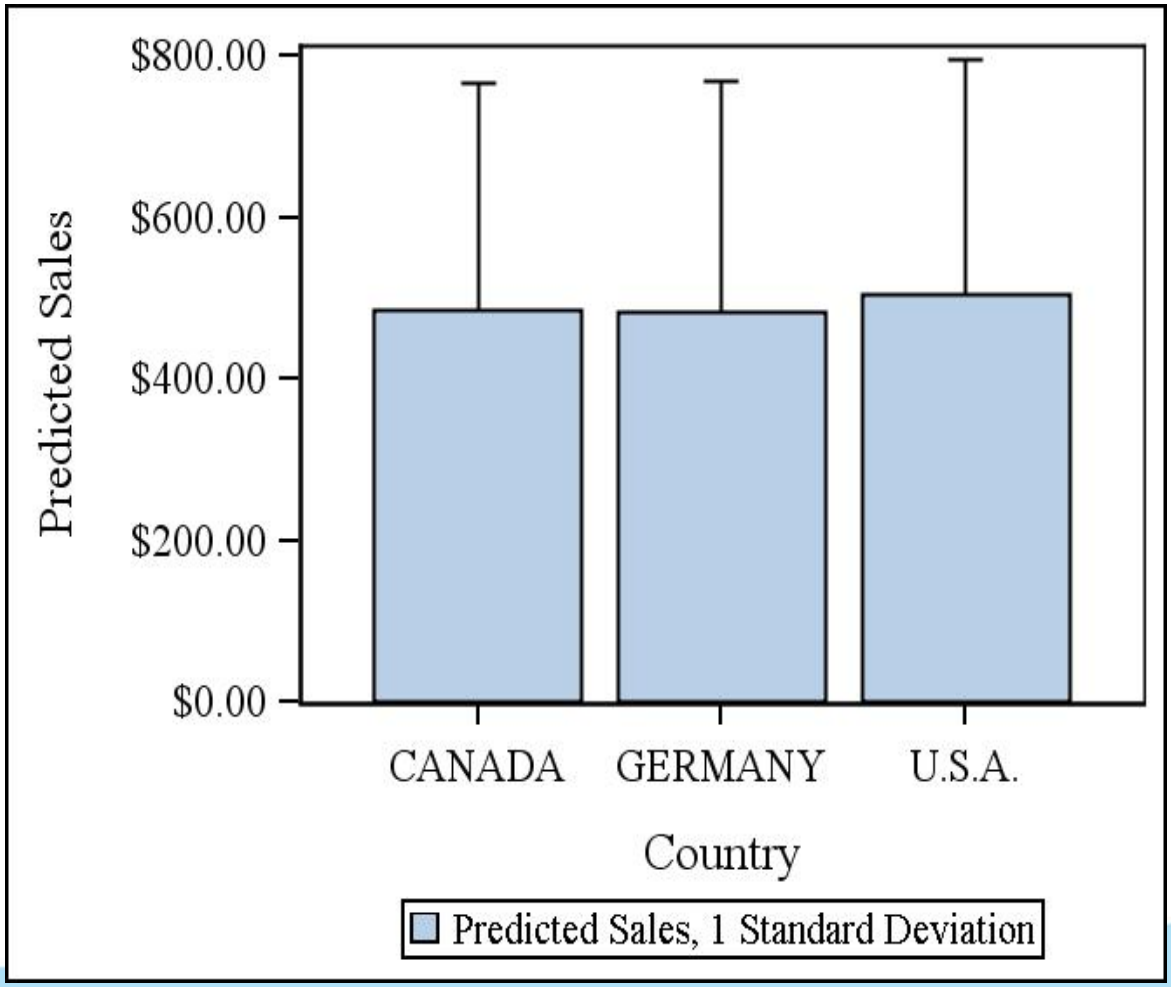
```
PROC SGPLOT DATA = Sashelp.Prdsale;
    VBAR   Country / RESPONSE  =  Predict;
    VLINE   Country / RESPONSE  =  Actual;
RUN;
```



```
PROC SGPLOT DATA = Sashelp.Prdsale;
   VBAR   Country / RESPONSE = Predict STAT = MEAN
   LIMITS = BOTH   LIMITSTAT =  STDDEV
        NUMSTD =1;
RUN;
```

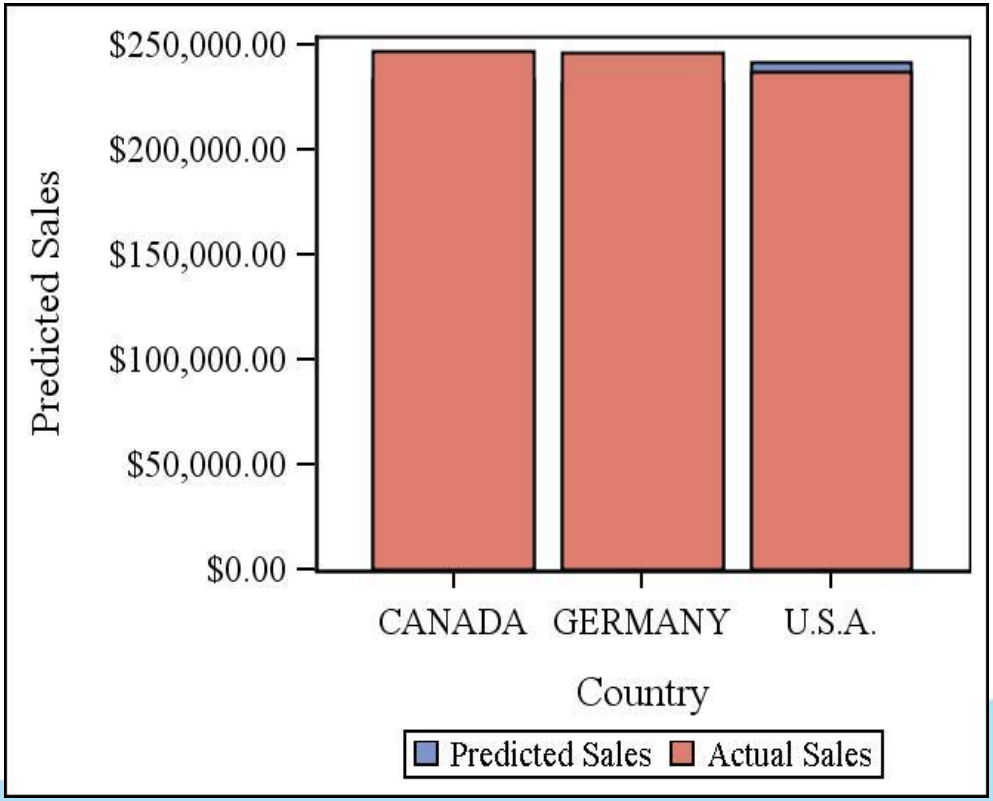


LIMITS = UPPERHALF

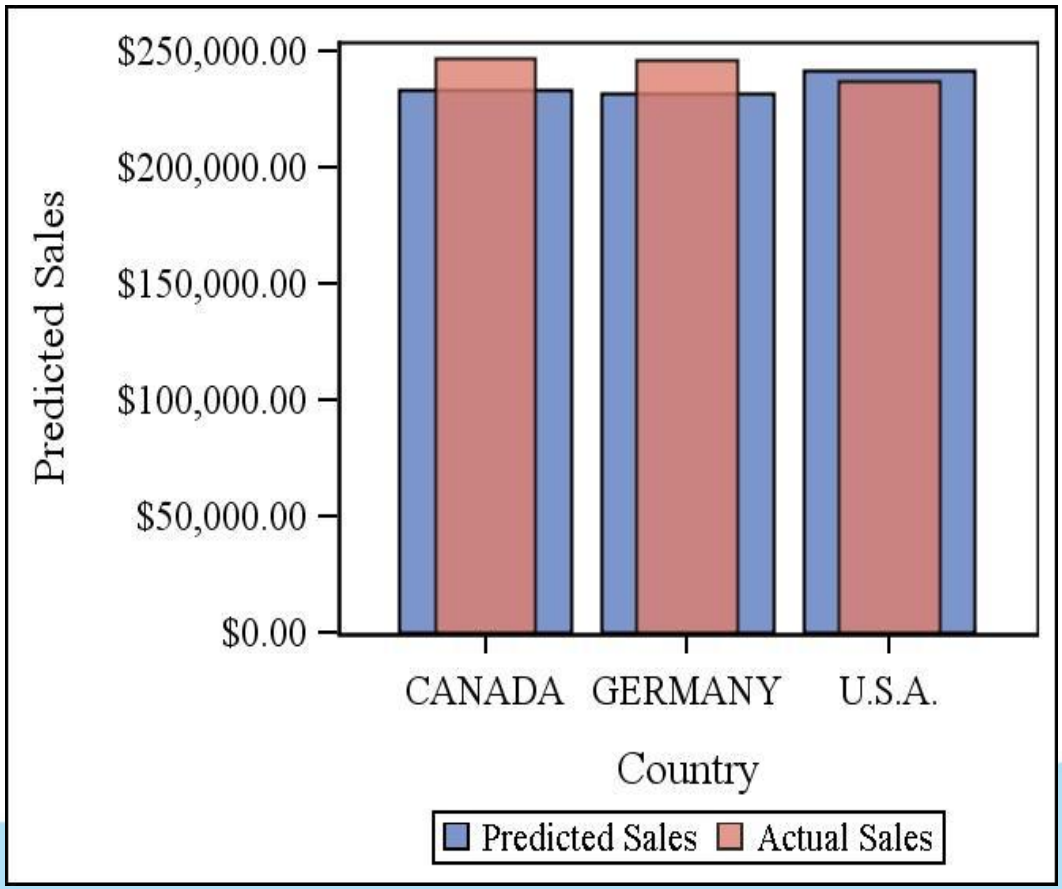


Overlaying

```
PROC SGPLOT DATA = Sashelp.Prdsale;
    VBAR Country  / RESPONSE = Predict;
    VBAR Country / RESPONSE = Actual;
RUN;
```

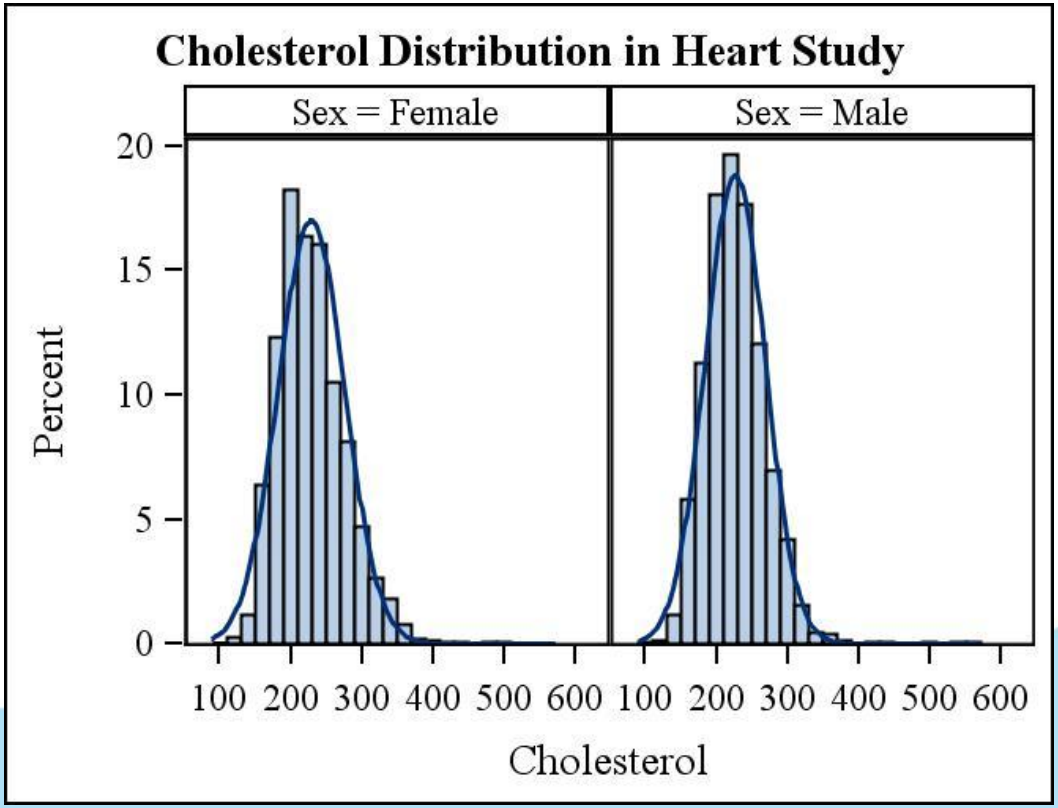


```
PROC SGPLOT DATA = Sashelp.Prdsale;
    VBAR Country / RESPONSE = Predict;
    VBAR Country / RESPONSE = Actual
        BARWIDTH =  0.5 TRANSPARENCY = 0.2;
RUN;
```

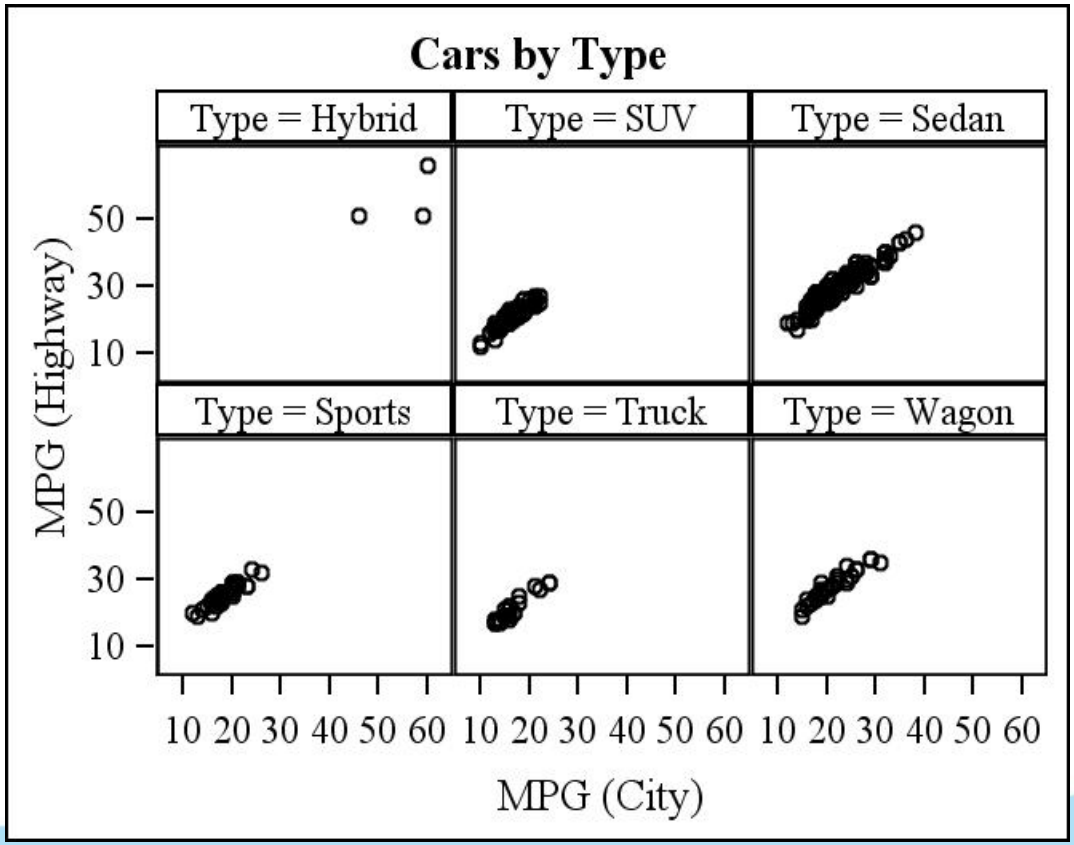


SGPANEL

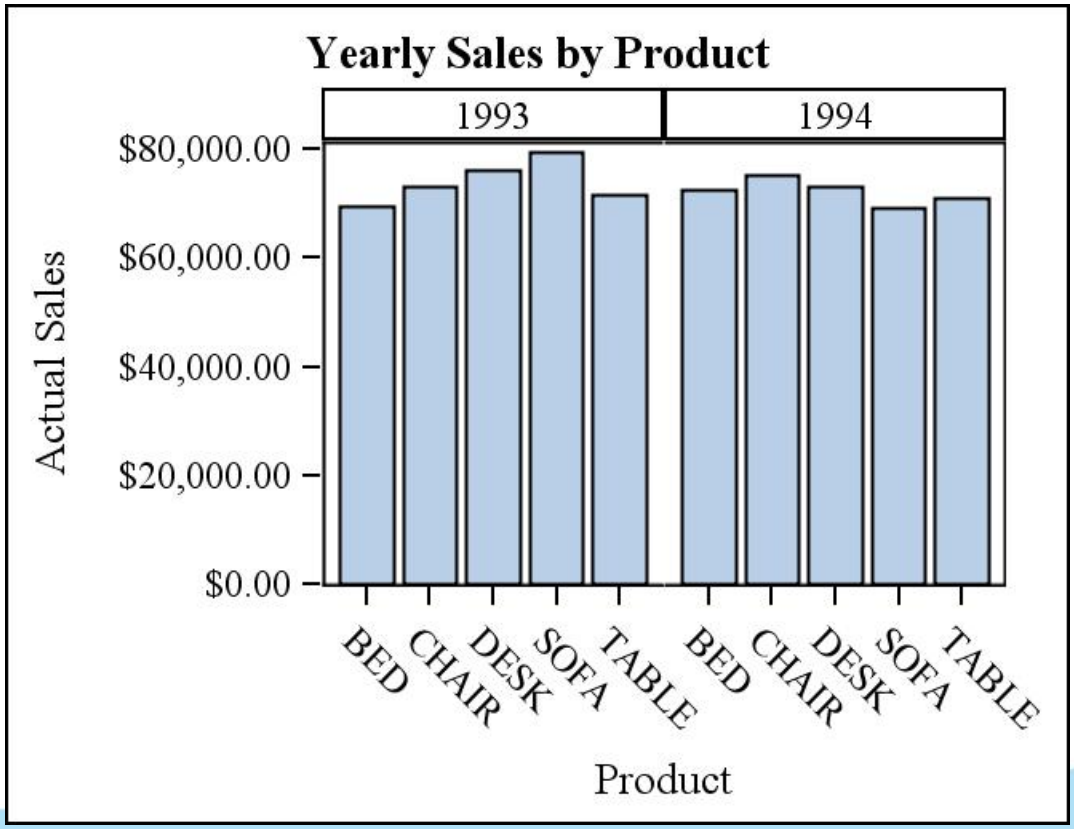
```
PROC SGPANEL DATA = Sashelp.Heart      NOAUTOLEGEND;
    TITLE "Cholesterol Distribution in Heart Study";
    PANELBY Sex;
    HISTOGRAM Cholesterol;
    DENSITY Cholesterol;
RUN;
```



```
PROC SGPANEL DATA = Sashelp.Cars;
    TITLE "Cars by Type";
    PANELBY Type / ROWS = 2 COLUMNS = 3;
    SCATTER X = Mpg_City Y = Mpg_Highway;
RUN;
```



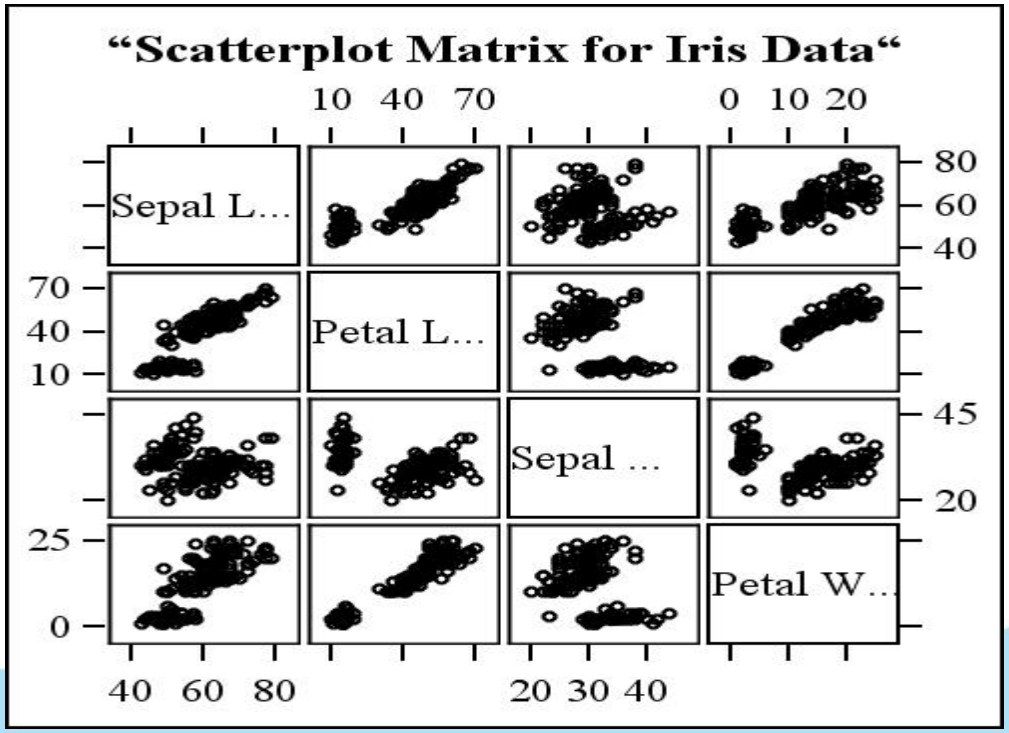
```
PROC SGPANEL DATA = Sashelp.Prdsale;
   TITLE "Yearly Sales by Product";
   PANELBY Year / NOVARNAME LAYOUT = COLUMNLATTICE
          NOBORDER;
   VBAR Product / RESPONSE = Actual;
RUN;
```



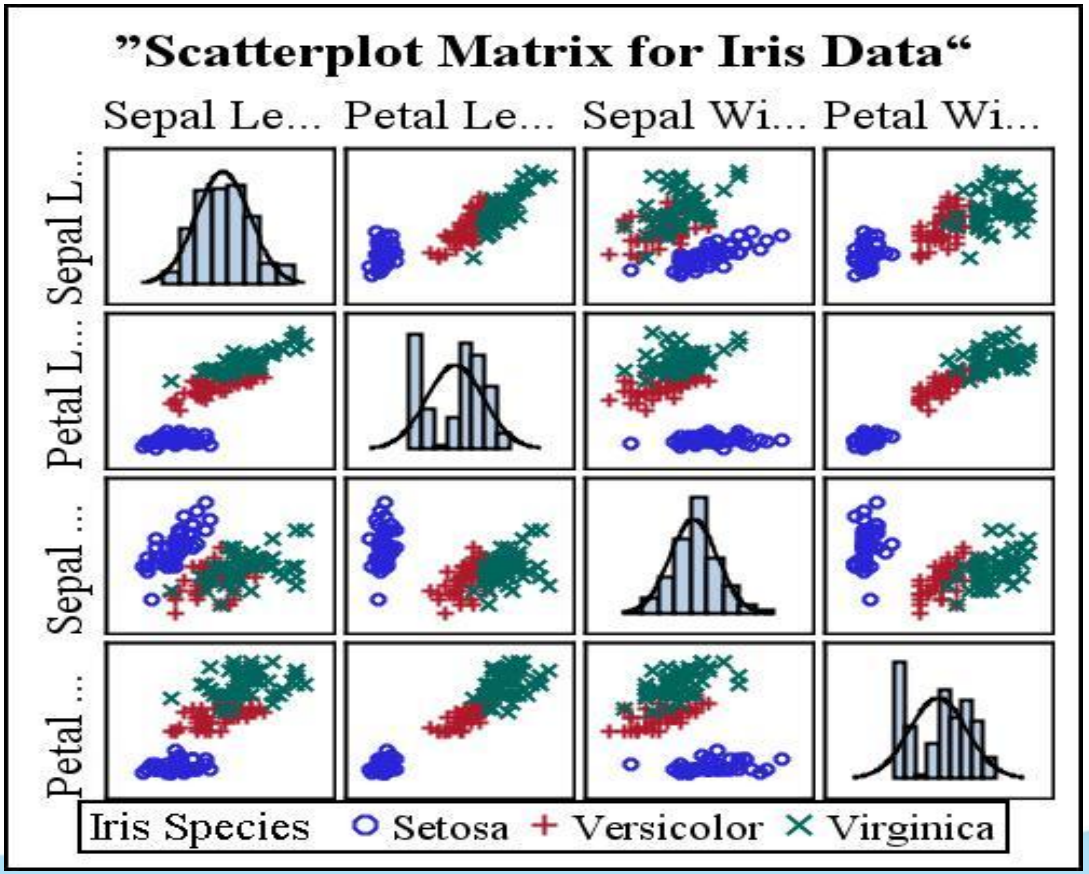
SGSCATTER

SCATTER PLOT MATRIX

```
PROC SGSCATTER DATA = Sashelp.Iris ;
    TITLE "Scatter plot Matrix for Iris Data" ;
    MATRIX Sepallength Petallength Sepalwidth Petalwidth ;
RUN ;
```

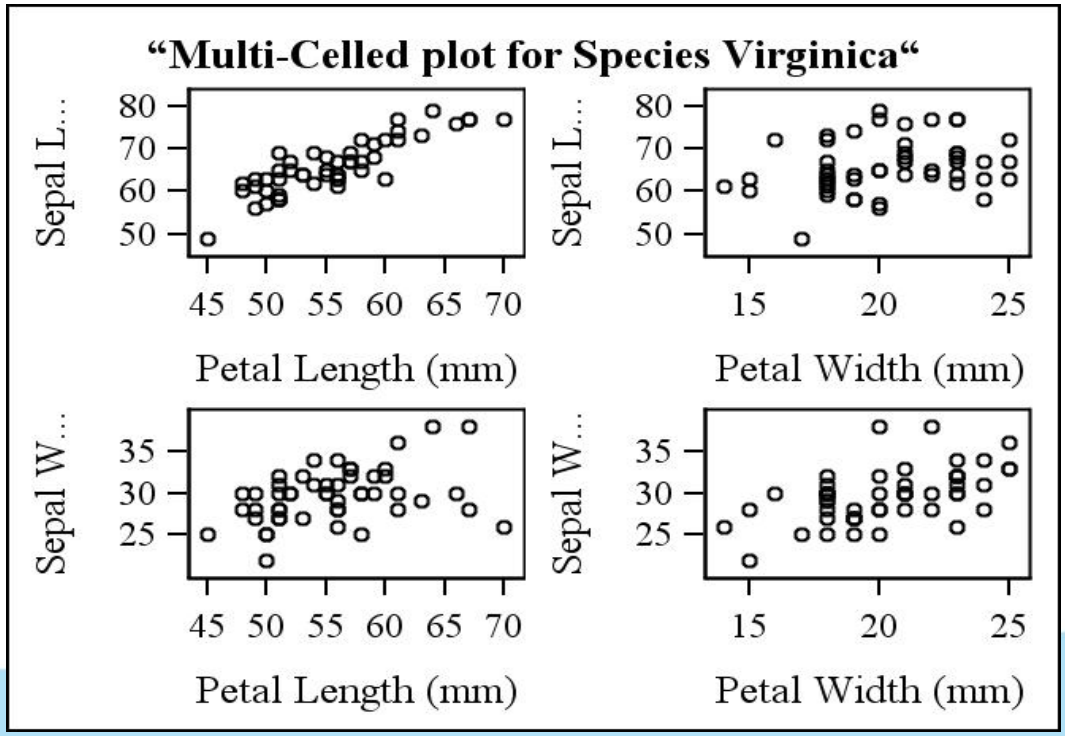


```
PROC SGSCATTER DATA = Sashelp.Iris ;
    TITLE ”Scatterplot Matrix for Iris Data“ ;
    MATRIX Sepallength Petallength Sepalwidth Petalwidth
        / GROUP= Species DIAGONAL= (HISTOGRAM
     NORMAL) ;
RUN ;
```



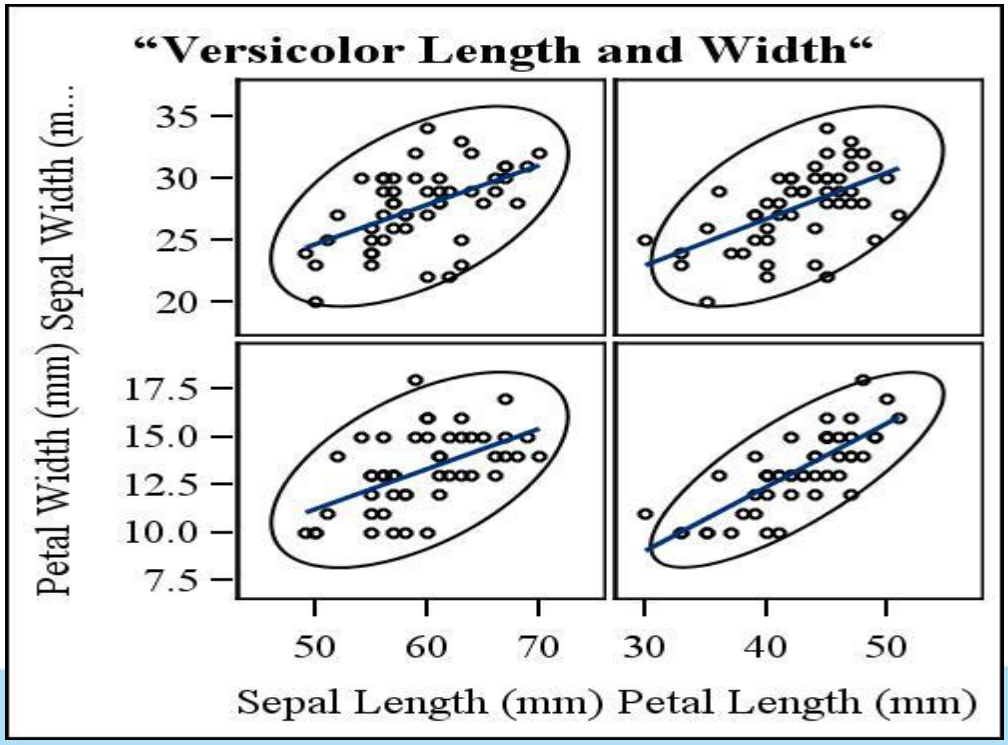
Comparative Panel Using PLOT Statement

```
PROC SGSCATTER DATA  =  Sashelp.Iris ( WHERE  = (Species =
    "Virginica")) ;
    TITLE  “Multi-Celled plot for Species Virginica“ ;
    PLOT  (Sepallength Sepalwidth) * (Petallength Petalwidth) ;
RUN ;
```



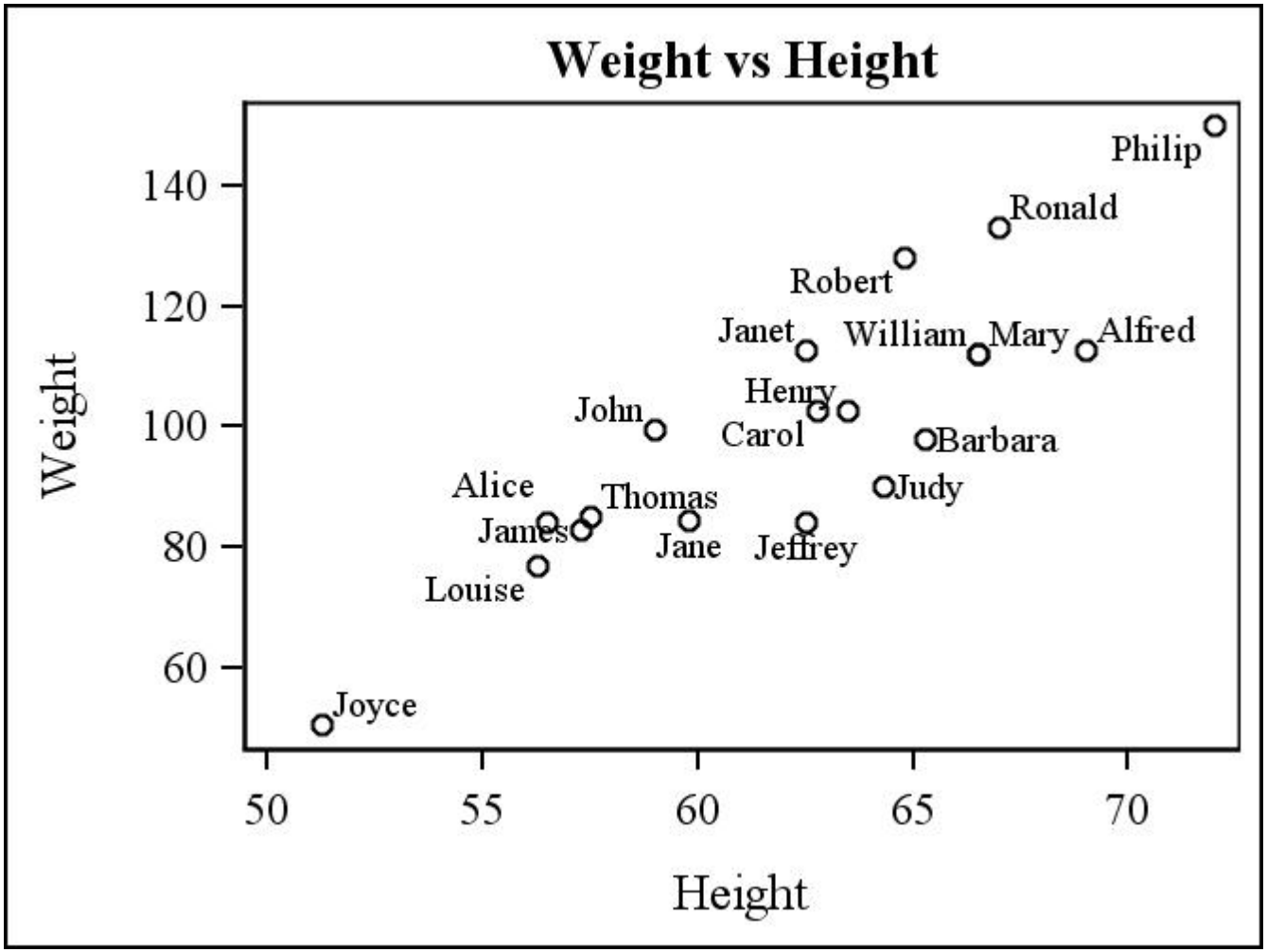
Comparative Panel Using Compare Statement

```
PROC SGSCATTER DATA  =  Sashelp.Iris (  WHERE = ( Species =
        "Versicolor")) ;
    TITLE   “Versicolor Length and Width“ ;
    COMPARE  Y = (Sepalwidth Petalwidth)
    X = (Sepallength Petallength)  /  REG ELLIPSE SPACING  =  4;
RUN ;
```

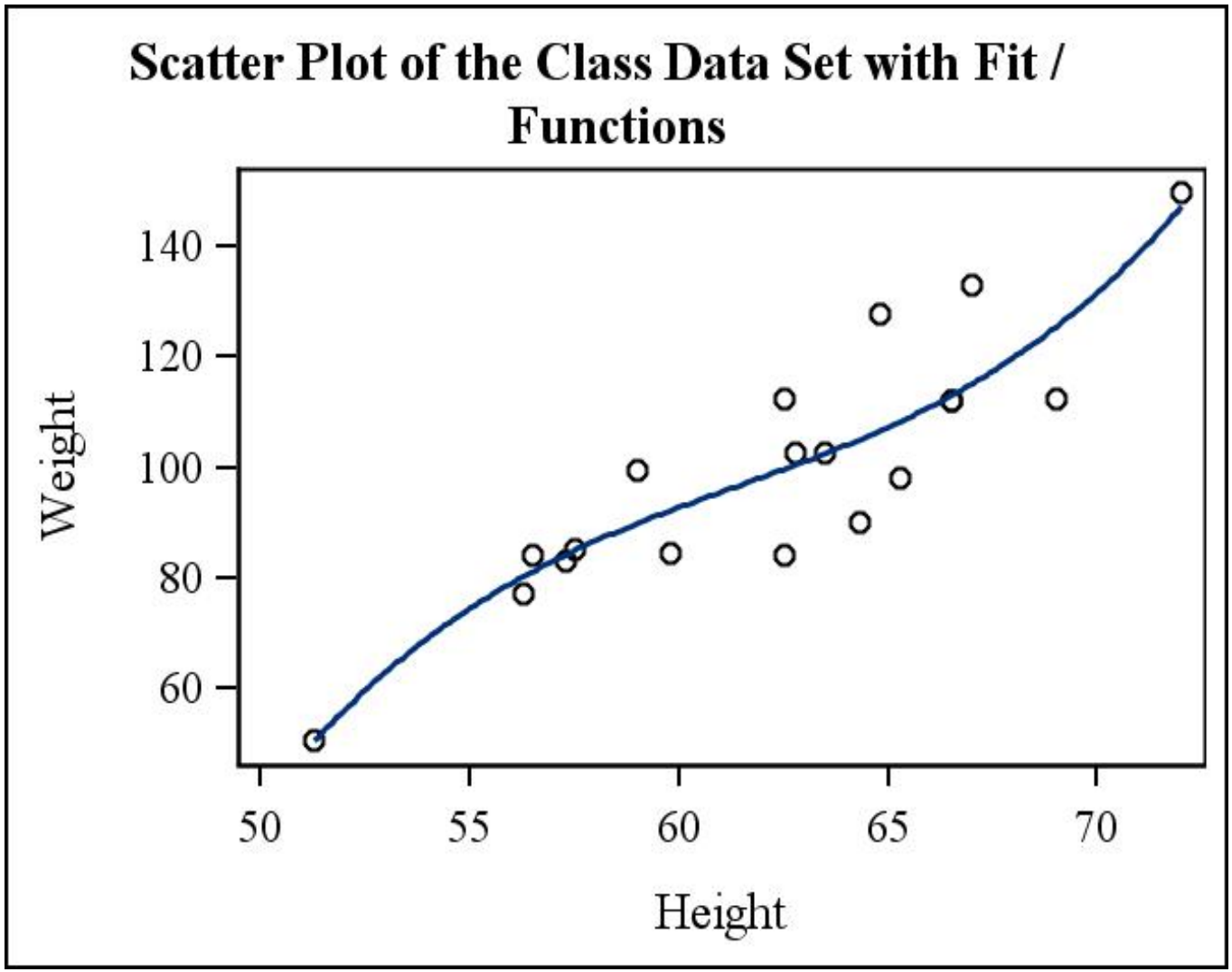


SGRENDER

```
PROC TEMPLATE ;
    DEFINE  STATGRAPH  Scatter ;
     BEGINGRAPH ;
          ENTRYTITLE   " Weight  vs Height  ";
          LAYOUT OVERLAY ;
               SCATTERPLOT   Y  =  Weight  X = Height  /
                                          DATALABEL =  Name;
          ENDLAYOUT ;
     ENDGRAPH;
    END ;
RUN ;

PROC SGRENDER DATA  =  Sashelp.Class  TEMPLATE  = Scatter;
RUN ;
```

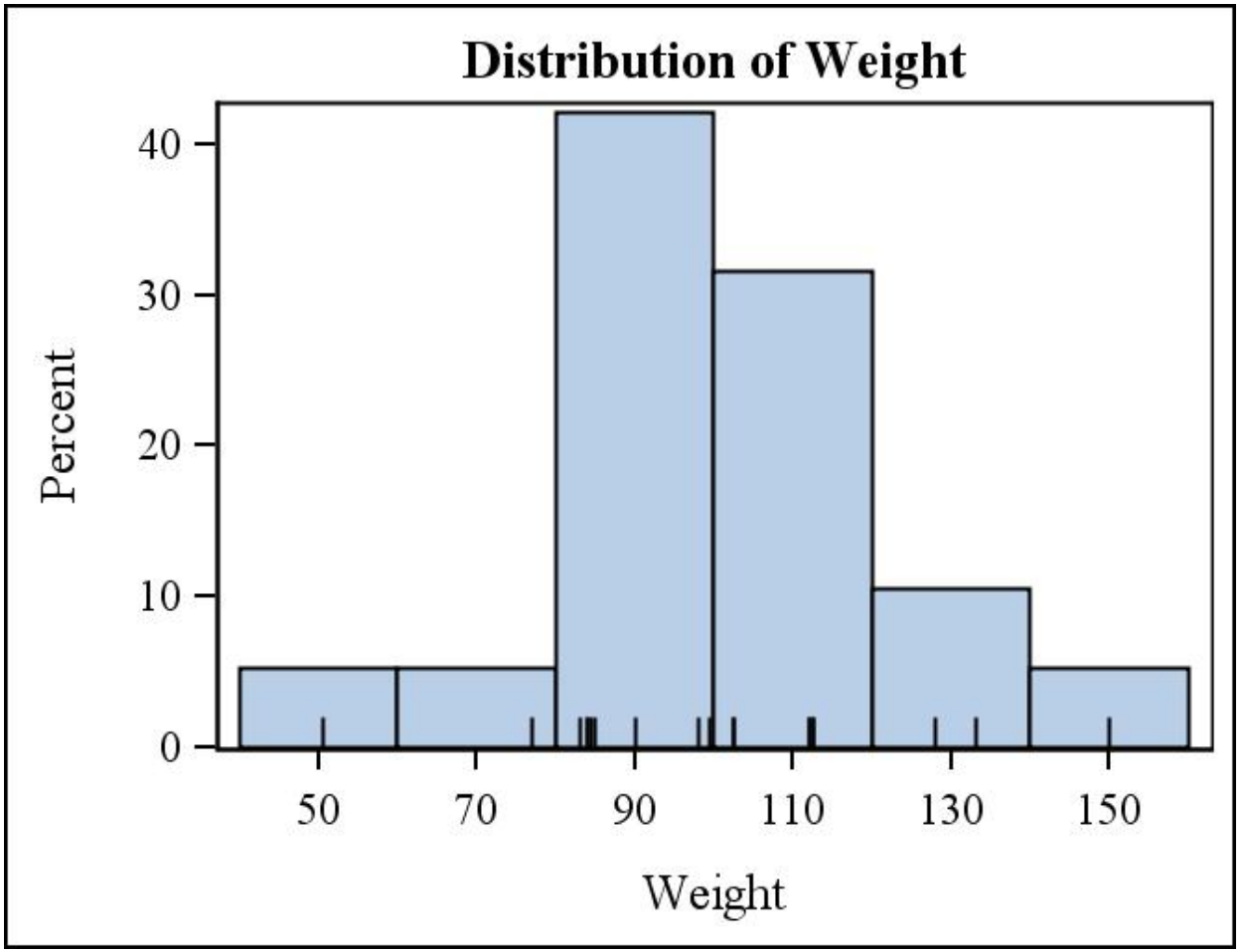

Weight vs Height
Weight
Height
140
120
100
80
60
50
55
60
65
70
Philip
Ronald
Robert
Janet
William
Mary
Alfred
Henry
John
Carol
Barbara
Alice
Thomas
Judy
James
Jane
Jeffrey
Louise
Joyce

```
PROC TEMPLATE;
    DEFINE STATGRAPH  Scatter ;
    BEGINGRAPH ;
ENTRYTITLE   “Scatter Plot of the Class Data Set with Fit Functions”;
          LAYOUT OVERLAY ;
               SCATTERPLOT   Y  = Weight  X = Height;
               REGRESSIONPLOT  Y  = Weight  X = Height  / DEGREE = 3;
          ENDLAYOUT ;
    ENDGRAPH ;
    END ;
RUN ;

PROC SGRENDER DATA = Sashelp.Class  TEMPLATE  =  Scatter;
RUN ;
```


Scatter Plot of the Class Data Set with Fit / Functions
Weight
140
120
100
80
60
50
55
60
65
70
Height

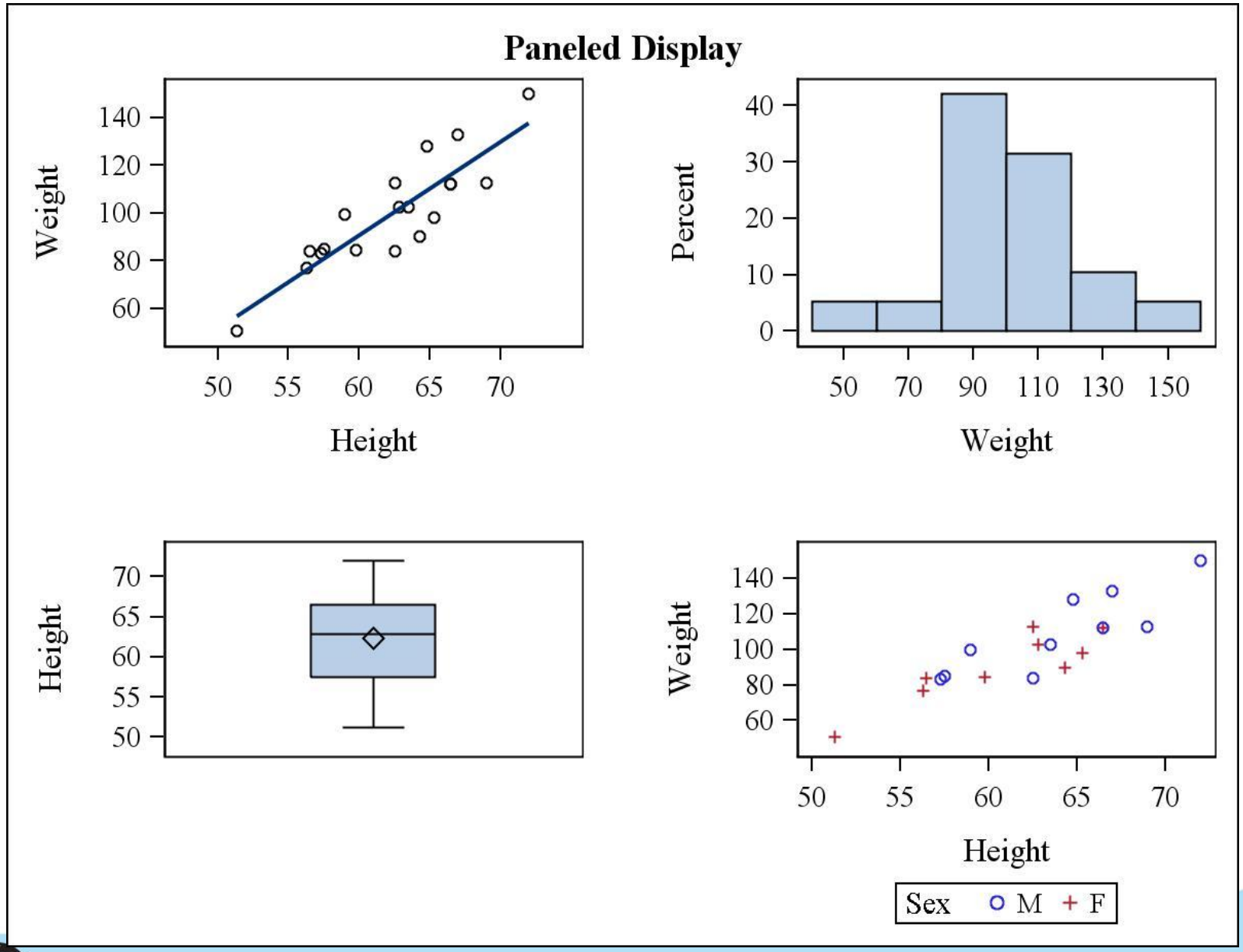
Fringe Plot

```
PROC TEMPLATE ;
    DEFINE  STATGRAPH  hist_fringe;
     BEGINGRAPH ;
     ENTRYTITLE   " Distribution of Weight ";
          LAYOUT OVERLAY ;
               HISTOGRAM Weight;
               FRINGEPLOT  WEIGHT;
          ENDLAYOUT ;
     ENDGRAPH;
     END ;
RUN ;

PROC  SGRENDER  DATA  =  Sashelp.Class  TEMPLATE  =
                                                  hist_fringe ;
RUN ;
```



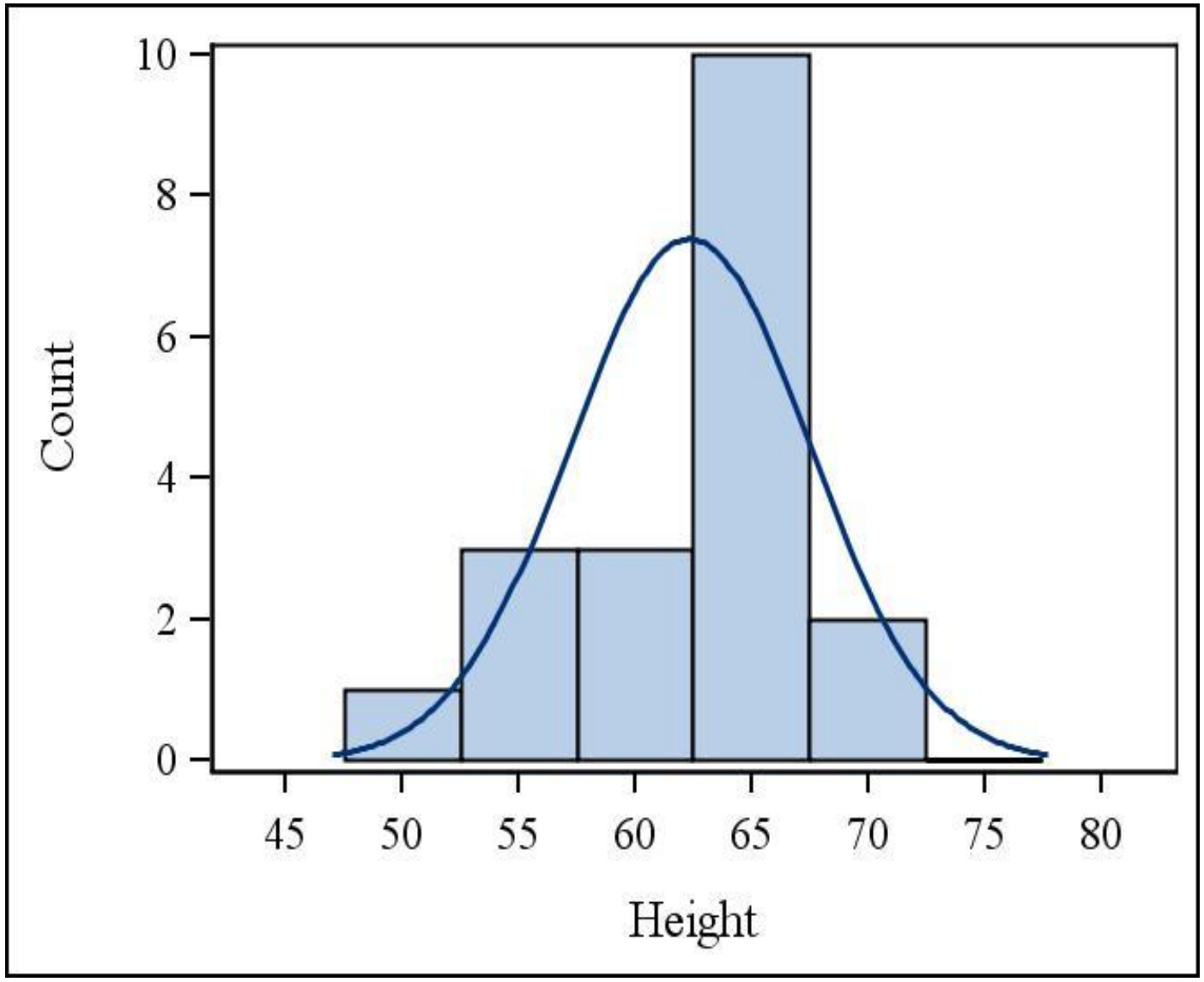
Fringe Plots Not available in PROC SGPLOT

```
PROC TEMPLATE;
      DEFINE STATGRAPH Panel;
        BEGINGRAPH;
              ENTRYTITLE "Paneled Display";

              LAYOUT LATTICE / ROWS= 2 COLUMNS= 2 ROWGUTTER = 10 COLUMNGUTTER =10;
                    LAYOUT OVERLAY;
                          SCATTERPLOT Y = Weight  X = Height;
                          REGRESSIONPLOT Y = Weight  X = Height;
                    ENDLAYOUT;

                    LAYOUT OVERLAY / XAXISOPTS = (LABEL= 'Weight');
                          HISTOGRAM Weight;
                    ENDLAYOUT;

                    LAYOUT OVERLAY / YAXISOPTS = (LABEL= 'Height');
                          BOXPLOT Y = Height;
                    ENDLAYOUT;

                    LAYOUT OVERLAY;
                          SCATTERPLOT Y = weight  X = height/ GROUP = sex  NAME = "Scat";
                          DISCRETELEGEND "Scat" / TITLE  = 'Sex';
                    ENDLAYOUT;

              ENDLAYOUT;
        ENDGRAPH;
      END;
RUN;

PROC SGRENDER DATA = Sashelp.Class TEMPLATE = Panel;
RUN;
```

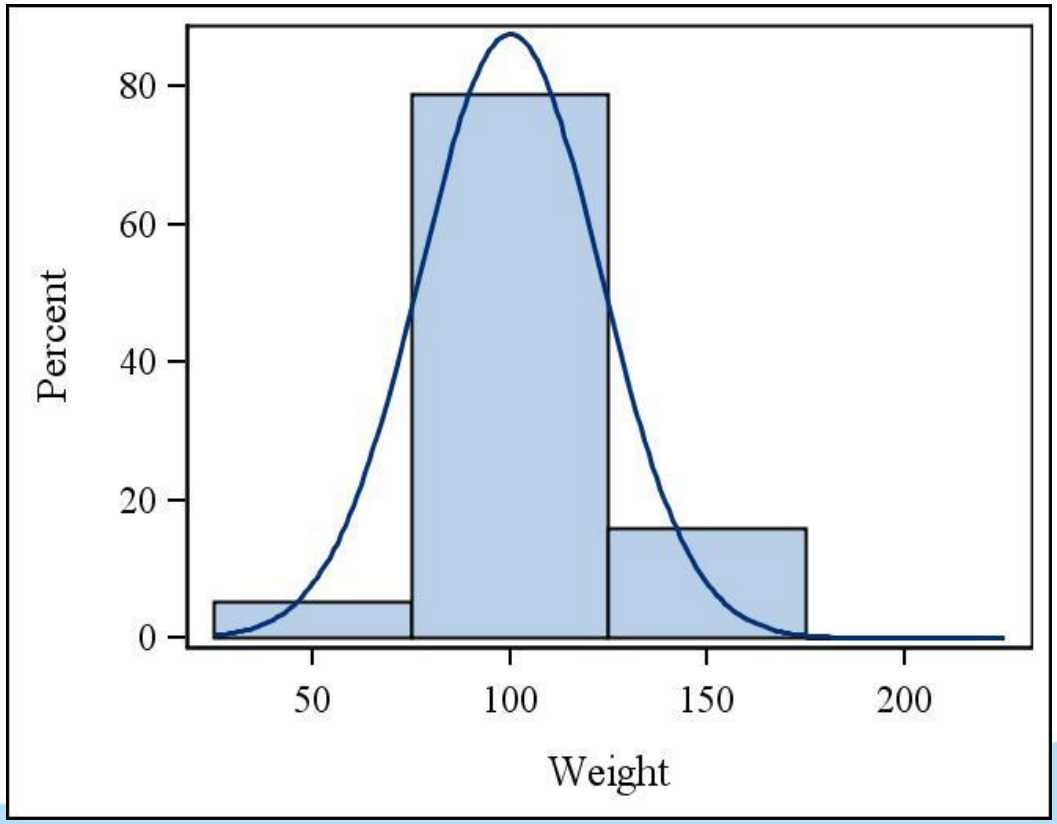

Paneled Display
Weight
140
120
100
80
60
50
55
60
65
70
Height
Percent
40
30
20
10
0
50
70
90
110
130
150
Weight
Height
70
65
60
55
50
Weight
140
120
100
80
60
50
55
60
65
70
Height
Sex
M
F

DYNAMICS AND MACRO VARIABLES

```
PROC TEMPLATE;
    DEFINE STATGRAPH  dynamic;
     BEGINGRAPH;
          MVAR  scale;
               NMVAR   bins;
               DYNAMIC  var ;
               LAYOUT OVERLAY ;
                    HISTOGRAM  var  / SCALE = scale          NBINS = bins;
                    DENISITYPLOT var;
               ENDLAYOUT;
     ENDGRAPH;
    END;
RUN;

%LET bins = 6;
%LET scale = count;
PROC SGRENDER DATA = sashelp.class TEMPLATE = dynamic;
    DYNAMIC var = 'Height';
RUN ;
```


Count
10
8
6
4
2
0
45
50
55
60
65
70
75
80
Height

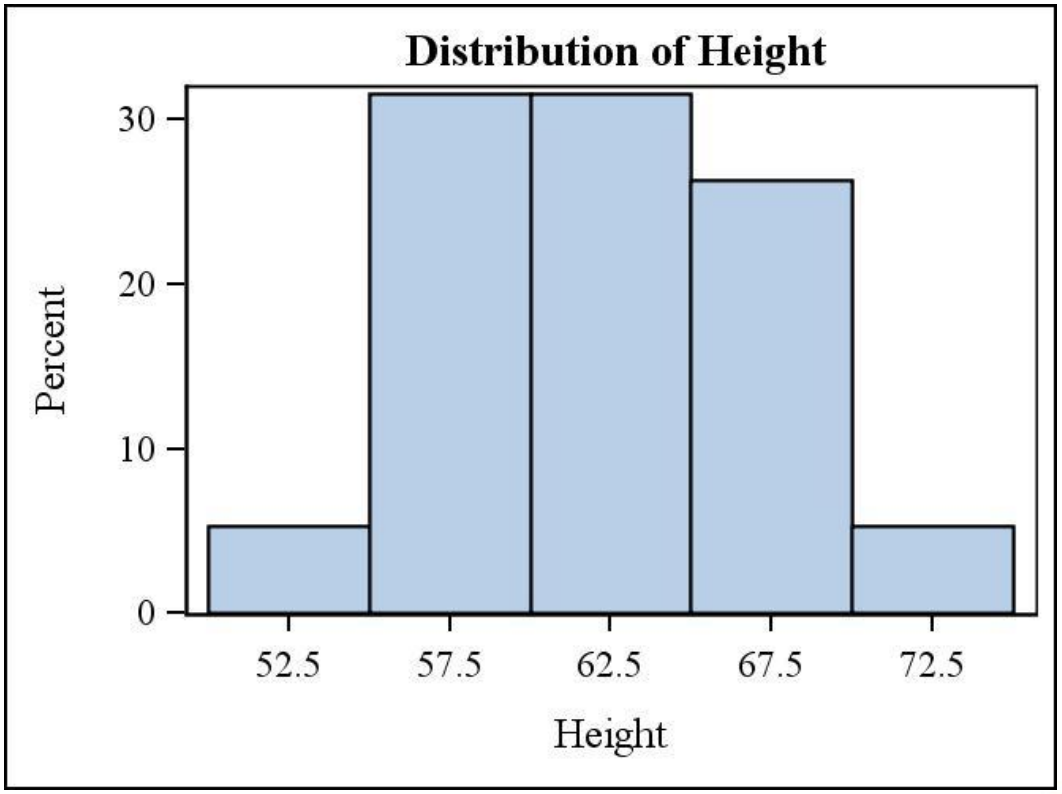
```
%LET bins = 4;
%LET  scale =  percent;
PROC SGRENDER DATA  =  sashelp.class TEMPLATE  =  dynamic;
    DYNAMIC  var  = 'Weight';
RUN ;
```



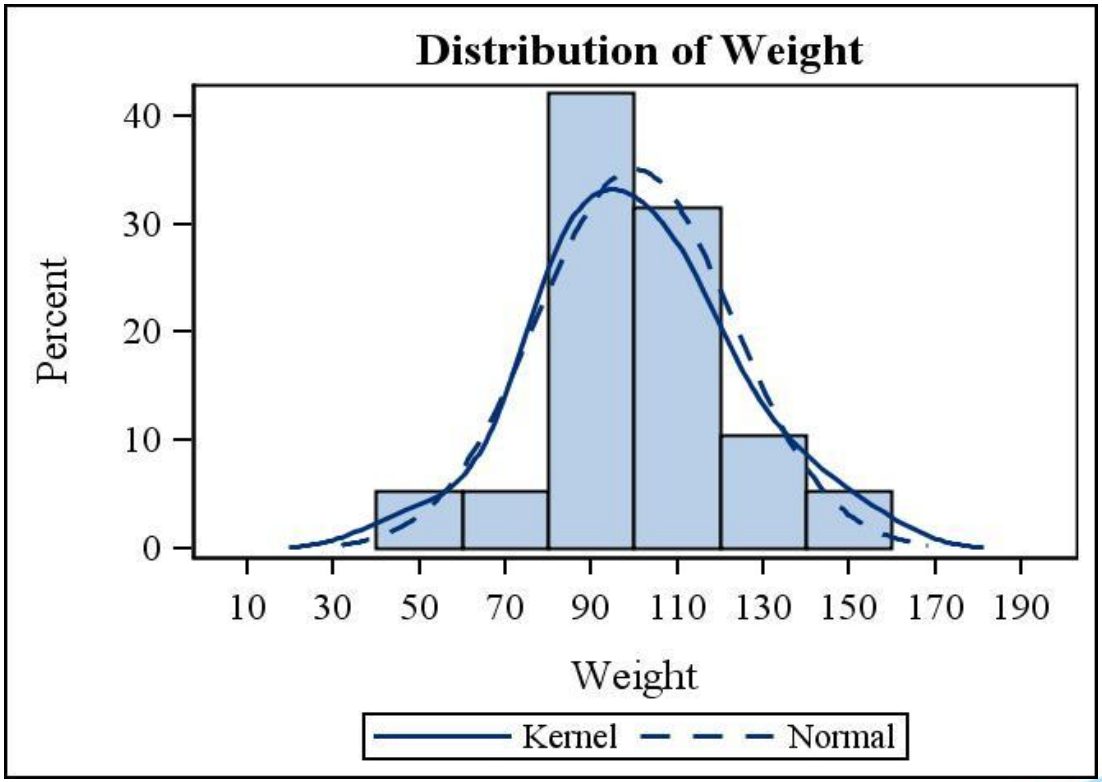
Graph on Condition

```
PROC TEMPLATE ;
      DEFINE  STATGRAPH  conditional;
      DYNAMIC  var  curve;
       BEGINGRAPH ;
            ENTRYTITLE   ' Distribution of '  var;
            LAYOUT OVERLAY ;
                  HISTOGRAM  VAR ;
                  IF ( UPCASE (curve)  =  'NONE' )
                       HISTOGRAM  var;
                  ENDIF;
                  IF ( UPCASE (curve) = 'ALL' )
                      DENSITYPLOT  VAR / KERNEL()  NAME= 'n' LEGENDLABEL = 'Normal' ;
                  ENDIF;
                  IF  (upcase(curve)  =  'ALL' )
                      DENSITYPLOT VAR  / NORMAL()  NAME = 'p'  LEGENDLABEL = 'Kernel'
                     LINEATTRS  =  ( PATTERN = DASH );
                  ENDIF;
                  DISCRETELEGEND  'n'  'k'  'p';
            ENDLAYOUT ;
       ENDGRAPH;
     END ;
RUN ;
```

```
PROC SGRENDER DATA  =  Sashelp.Class  TEMPLATE  =
                                                  conditional;
    DYNAMIC var = 'HEIGHT' curve ='None';
run;
```



```
PROC SGRENDER DATA  =  Sashelp.Class  TEMPLATE
                                    = conditional;
   DYNAMIC var = 'Weight' curve ='All';
run;
```



ODS DESTINATIONS FOR GRAPHS

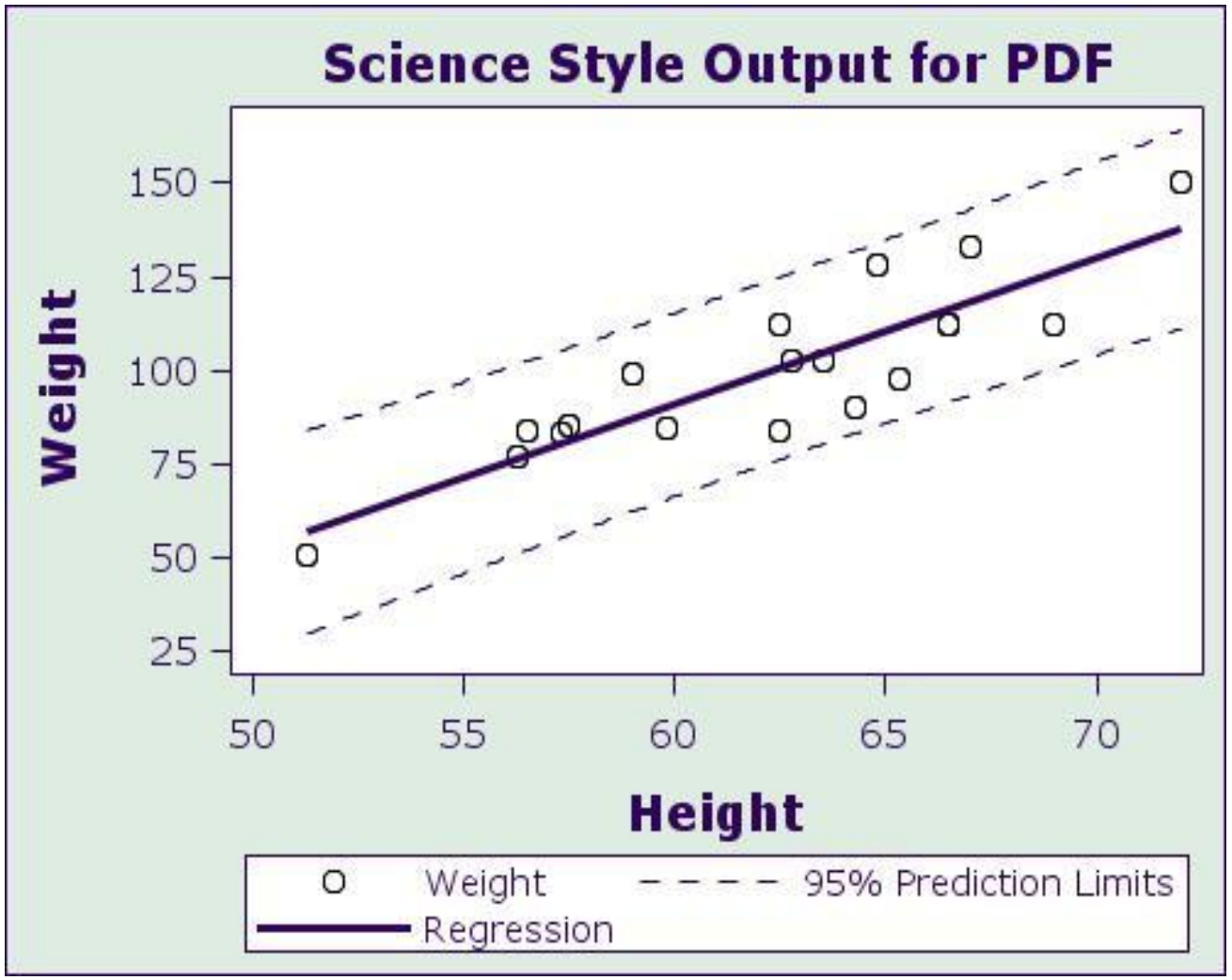
```
ODS PDF FILE = 'Scatter.pdf';

PROC SGPLOT DATA = Sashelp.Class;
SCATTER  X = Height  Y = Weight;
RUN;

ODS PDF CLOSE;
```

ODS STYLES

Some of them are

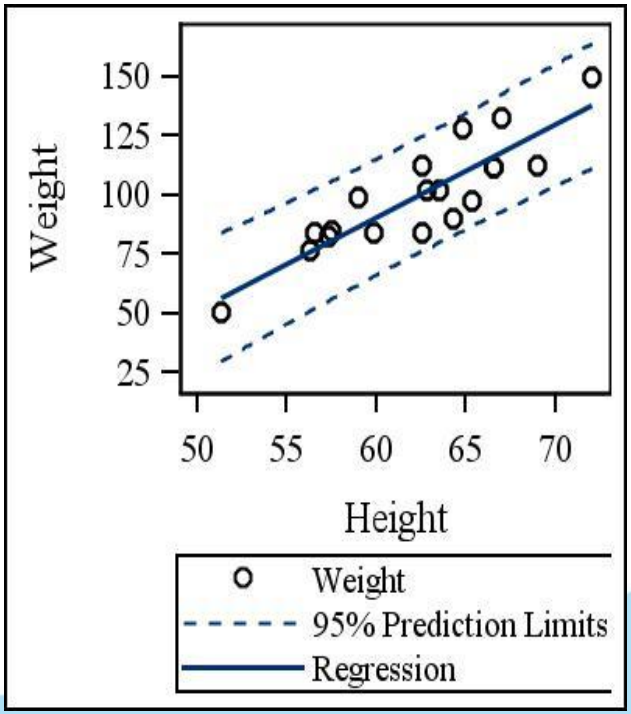
- Listing
- Default
- Statistical
- Analysis
- Journal
- Journal2
- Science

```
ODS PDF FILE = 'style.pdf' STYLE = SCIENCE;

PROC SGPLOT DATA = Sashelp.Class;
    TITLE 'Science Style Output for PDF';
    SCATTER X = Height  Y = Weight;
    REG X = Height  Y = weight / CLI;
RUN;

ODS PDF CLOSE;
```


Science Style Output for PDF
Weight
150
125
100
75
50
25
50
55
60
65
70
Height
Weight
95% Prediction Limits
Regression

SPECIFYING SIZE AND IMAGE FORMAT

```
ODS GRAPHICS / RESET IMAGEFMT = JPEG
HEIGHT= 2in  WIDTH= 2in;

ODS RTF  FILE = 'a.rtf';

PROC SGPLOT DATA = SASHELP.CLASS;
    SCATTER   X = HEIGHT   Y = WEIGHT;
    REG  X = HEIGHT   Y = WEIGHT/ CLI;
RUN;
ODS RTF CLOSE;
```



Some Annotation Functions for SG Procedures for SAS 9.3

Annotate Function	Description
Text	Draw a text on the graph
Image	Draw an image on the graph
Line	Draw a line on the graph
Arrow	Draw a line with an arrowhead on the graph
Rectangle	Draw a square or rectangle on the graph
Oval	Draw a circle or oval on the graph
Polygon	Draw a closed polygon on the graph

Conclusions

- SGPLOT, SGSCATTER, SGPANEL can be used quickly to graph most of the statistical graphics with high quality.
- Using GTL language and SGRENDER procedure can create customized graphics or layouts
- Many other graphs can be created using SG procedures.

United States Census Bureau

References

1. SAS Institute Inc. (2008). “SAS/GRAPH 9.2: Statistical Graphics Procedures Guide.”

2. Susan J. Slaughter, Lora D. Delwiche (2010), *“Using PROC SGPLOT for Quick High Quality Graphs,”* Proceedings of the 2010 -SAS Global Forum, April 2010

3. Heath, Dan. 2008. “Effective Graphics Made Simple using SAS/GRAPH ‘SG’ Procedures.”, *Proceedings of the 2008 SAS Global Forum,* April, 2008

4. Kuhfeld, W. F. (2010), Statistical Graphics in SAS— An Introduction to the Graph Template Language and the Statistical Graphics Procedures, SAS Institute Inc.

5. Sridharma, Selvaratnam (2010), “Introduction to Statistical Graphics Procedures” Nesug 2010 Proceedings

Contact Information

Speaker: Selvaratnam Sridharma

Company: Economic Planning and Coordination Division
U.S. Bureau of the Census
Washington, DC 20233-6100

Phone: 301-763-6774

Email: selvaratnam.sridharma@census.gov