

# SAS Workshop Series: Quality Graphics in SAS

Jason Brinkley

Department of Biostatistics

# SAS and Graphics

- SAS software has a sordid history with graphics. Being a code based software that was originally available for mainframe computers in the early 1970s, SAS was given a reputation for poor quality graphical output in the 1980s when point and click softwares became widely available.
- Modern users will find many different ways to obtain graphs in SAS, and SAS continues to add new options in newer versions.

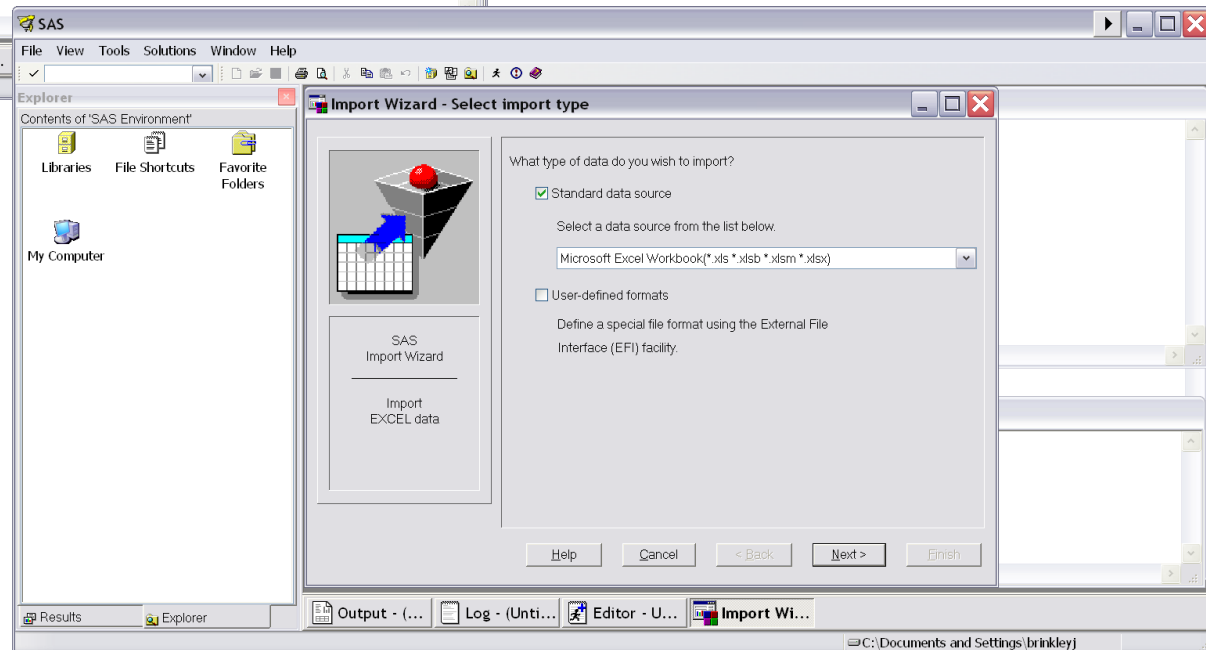
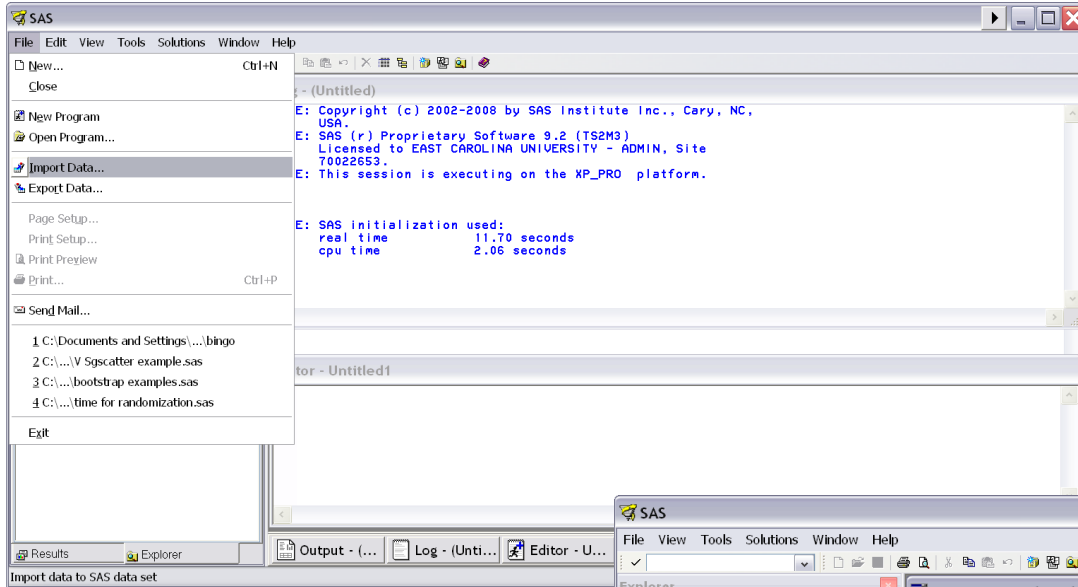
# Example

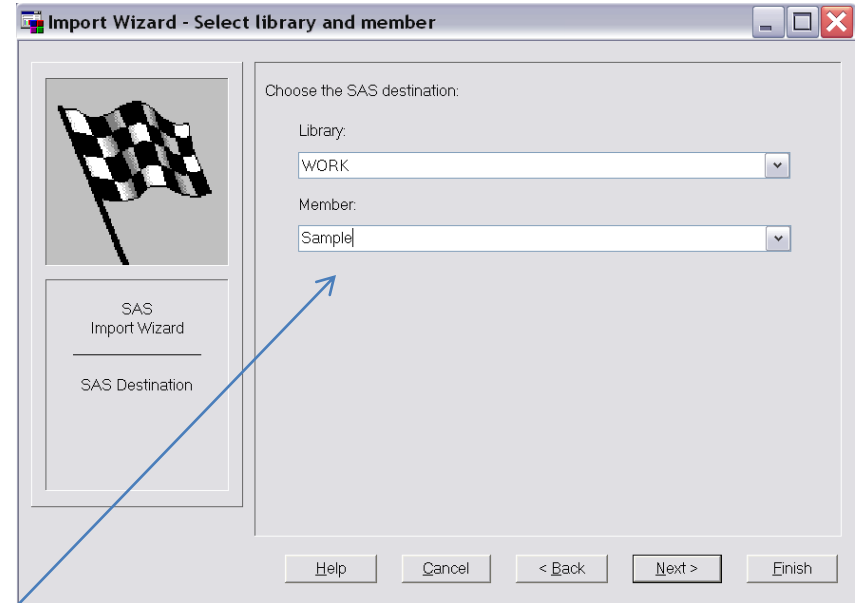
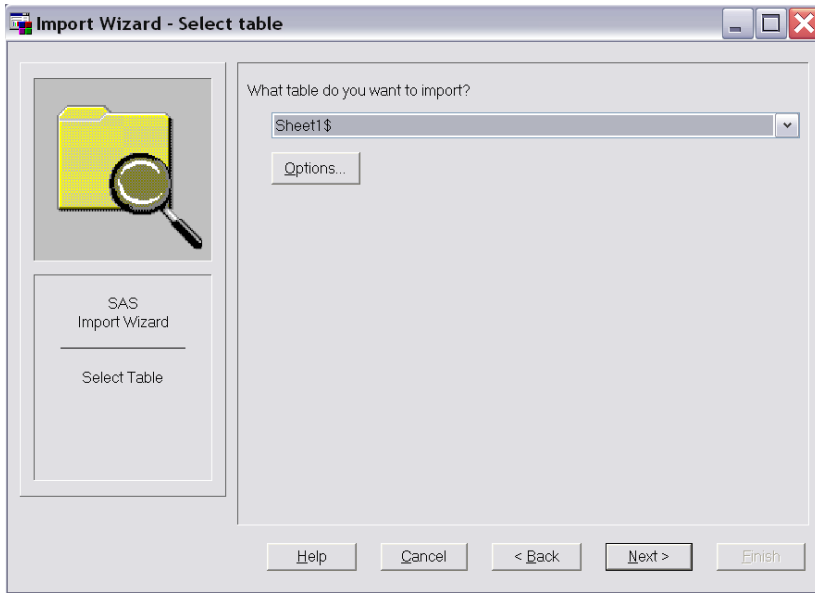
Students in an introductory statistics class, taught by Professor John Eccleston and Dr Richard Wilson at The University of Queensland, participated in a simple experiment. The students took their own pulse rate. They were then asked to flip a coin . If the coin came up heads, they were to run in place for one minute. Otherwise they sat for one minute. Then everyone took their pulse again. The pulse rates and other physiological and lifestyle data are given in the data.

Five class groups between 1993 and 1998 participated in the experiment. <http://www.statsci.org/data/oz/ms212.html>

ID	Height	Weight	Age	Gender	Smoker	Alcohol	Exercise	Ran	Pulse1	Pulse2	Year
1	173	57	18	Female	No	Yes	Moderate	Sat	86	88	93
2	179	58	19	Female	No	Yes	Moderate	Ran	82	150	93
3	167	62	18	Female	No	Yes	High	Ran	96	176	93
4	195	84	18	Male	No	Yes	High	Sat	71	73	93
5	173	64	18	Female	No	Yes	Low	Sat	90	88	93
6	184	74	22	Male	No	Yes	Low	Ran	78	141	93
7	162	57	20	Female	No	Yes	Moderate	Sat	68	72	93
8	169	55	18	Female	No	Yes	Moderate	Sat	71	77	93
9	164	56	19	Female	No	Yes	High	Sat	68	68	93
10	168	60	23	Male	No	Yes	Moderate	Ran	88	150	93
11	170	75	20	Male	No	Yes	High	Ran	76	88	93
12	178	58	19	Male	No	No	Low	Sat	74	76	93
13	170	68	22	Male	Yes	Yes	Moderate	Sat	70	71	93
14	187	59	18	Male	No	Yes	High	Sat	78	82	93
15	180	72	18	Male	No	Yes	Moderate	Sat	69	67	93
16	185	110	22	Male	No	Yes	Low	Sat	77	73	93
17	170	56	19	Male	No	No	Low	Sat	64	63	93

# Loading Data into SAS (Quickly)





This creates a temporary file called "Sample" and is located in your Work directory. It will NOT be saved and you will have to reimport data into SAS again at startup. But SAS does give you the option to save import code

# Review: Descriptives

- We will focus on three variables for now: Height (cm), Weight (kg), and Gender.
- We can start by reviewing how to print and summarize data.
- We expect to see difference in heights and weights by gender.

\*This simple code tells SAS that the dataset Sample is the one I want to work with this session;

```
Data Sample;
set Sample;
run;
```

\*Print the Data;

```
Proc Print;
Run;
```

The MEANS Procedure

Variable	Label	N	Mean	Std Dev	Minimum
Height	Height	110	171.5818182	16.0768691	68.0000000
Weight	Weight	110	66.3318182	15.1568674	27.0000000

\*Summarize the Data;

```
Proc Means;
var Height Weight;
Run;
```

The MEANS Procedure

Variable	Label	N	Mean	Std Dev	Minimum
Height	Height	51	164.4901961	12.6733935	93.0000000
Weight	Weight	51	56.5980392	9.3755104	27.0000000

\*Sort the data;

```
Proc Sort;
by Gender;
run;
```

\*Summaries by Gender;

```
Proc Means;
by Gender;
Var Height Weight;
run;
```

Variable	Label	Maximum
Height	Height	180.0000000
Weight	Weight	78.0000000

# Histograms

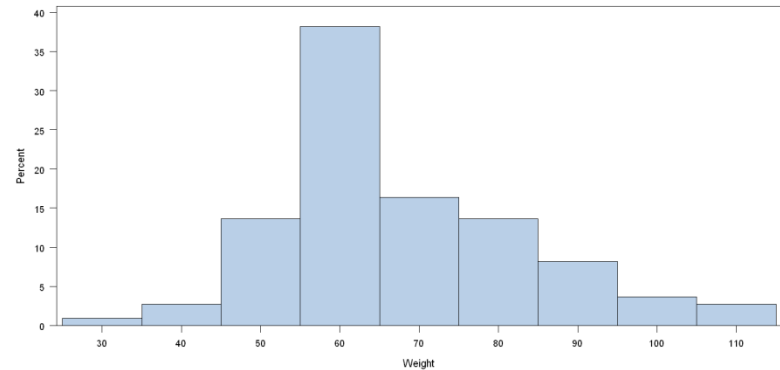
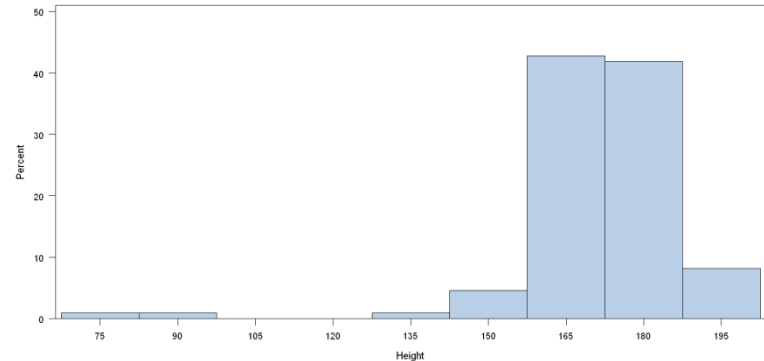
\*Simple histogram;

**Proc univariate;**

var height weight;

histogram height weight;

**run;**



Too much other output! Proc Univariate gives you LOTS of information!

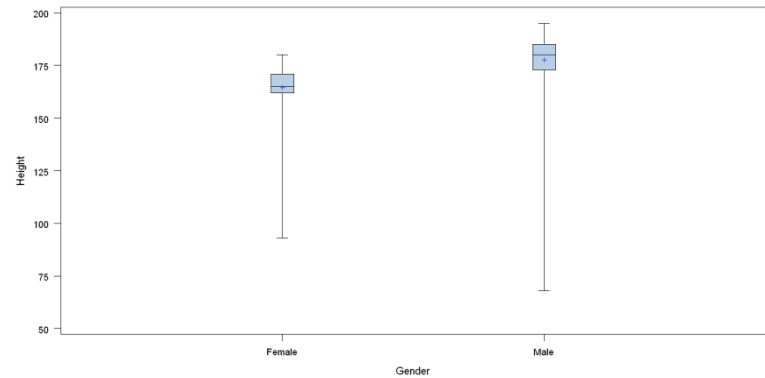
# Boxplot – Height by Gender

\*Simple boxplot;

**Proc boxplot;**

**plot height\*gender;**

**run;**



# Better Boxplot

\*Better Boxplot;

**proc boxplot ;**

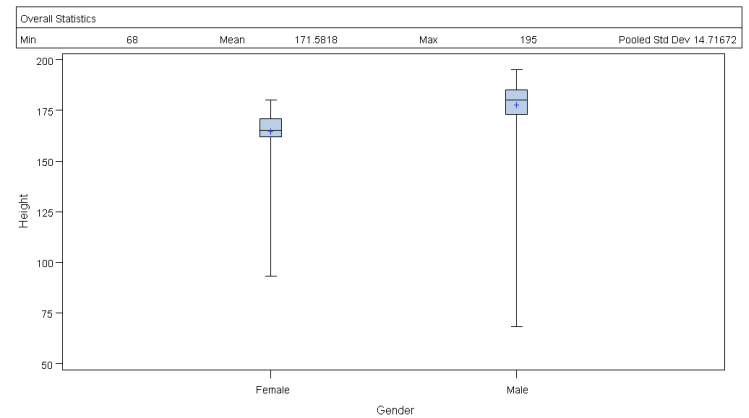
plot height\*gender ;

inset min mean max stddev /

header = 'Overall Statistics'

pos = tm;

**run;**



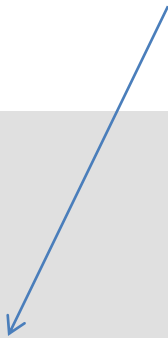
# Advances in graphics and display of output - ODS

- Newer versions of SAS have incorporated a new system of graphics referred to as ODS or “Output Delivery System”. This is a way to make better looking text output and visualizations. But the output is not stored in the usual places and takes more computing time.
- Generally speaking, SAS users look at raw output first and then decide what needs to be prettied up. Unless what they have a good understanding of what they want to look at and then they go straight to the “pretty” reports.

# ODS and Text - Example

```
ods html body = "output.html";  
proc means;  
var height weight;  
run;  
ods html close;
```

Tables can be copied and pasted into excel or read into other SAS datafiles!



## *The SAS System*

### *The MEANS Procedure*

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
Height	Height	110	171.5818182	16.0768691	68.0000000	195.0000000
Weight	Weight	110	66.3318182	15.1568674	27.0000000	110.0000000

# Best Boxplot – Using ODS

```
ods graphics on;
```

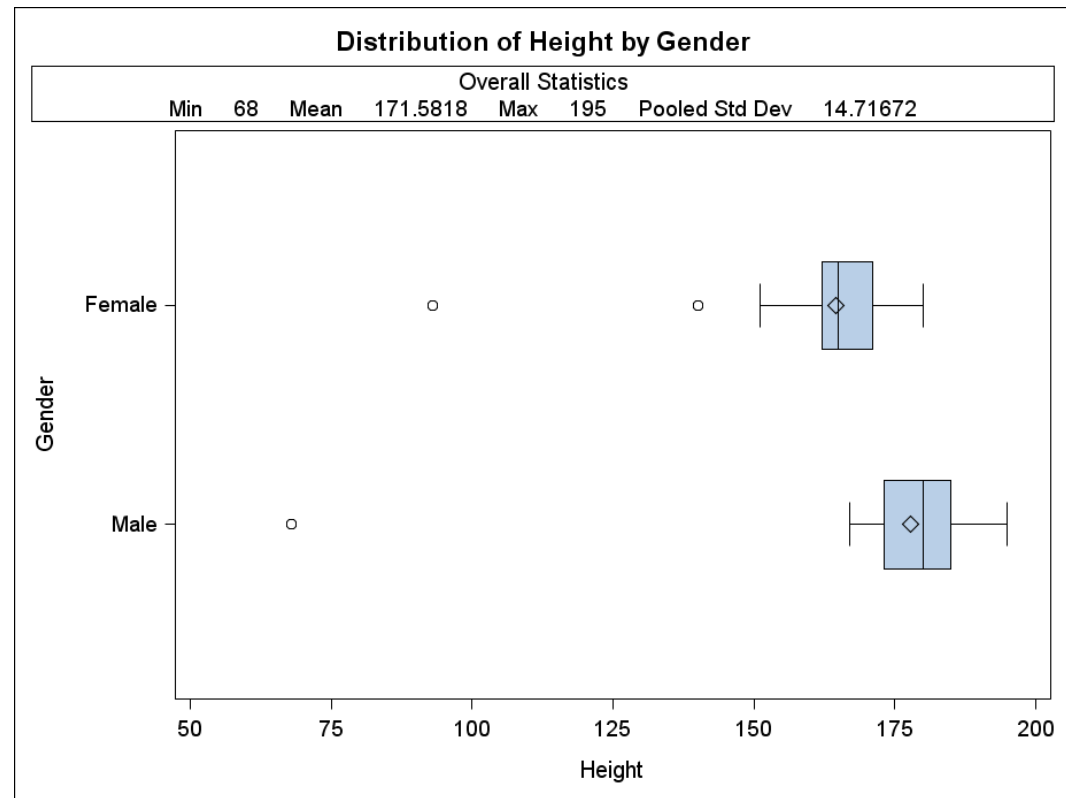
```
*Best Boxplot;
```

```
proc boxplot ;
```

```
  plot height*gender /  
  boxstyle=scematic horizontal;  
  inset min mean max stddev /  
    header = 'Overall Statistics'  
    pos   = tm;
```

```
run;
```

```
ods graphics off;
```



# Charts - Simple

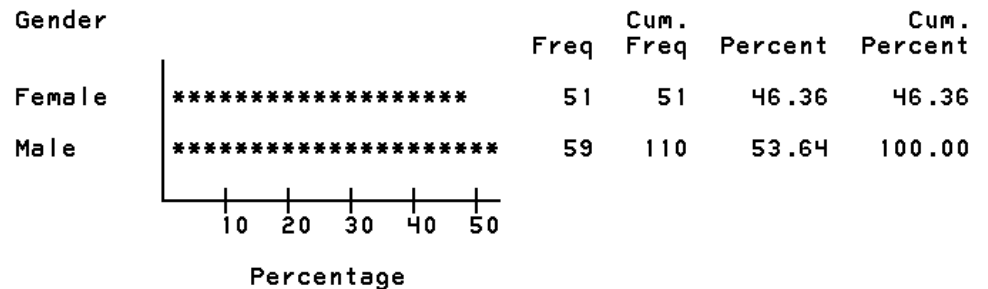
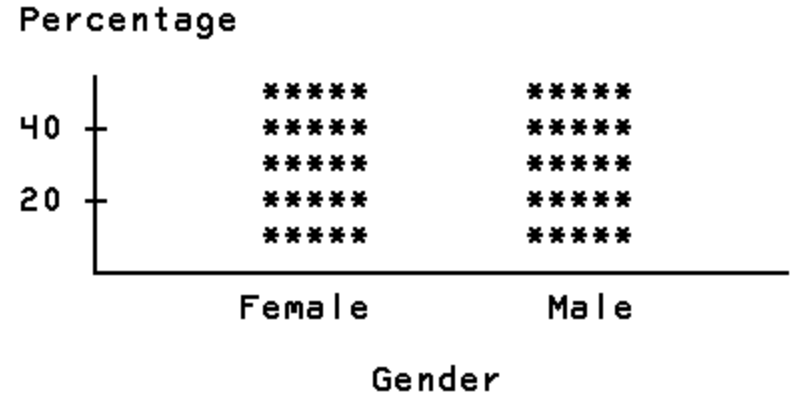
\*Categorical variables;

**Proc chart;**

vbar gender / type=percent;

hbar gender;

**run;**



# Charts - Better

\*Categorical variables;

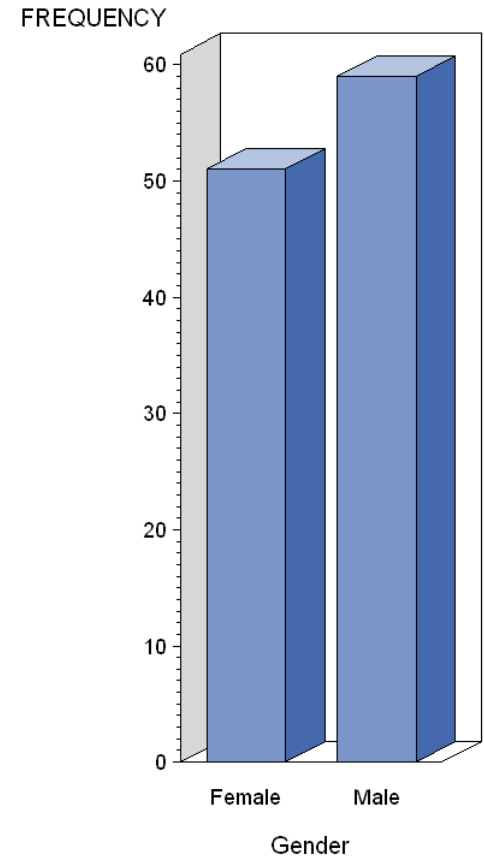
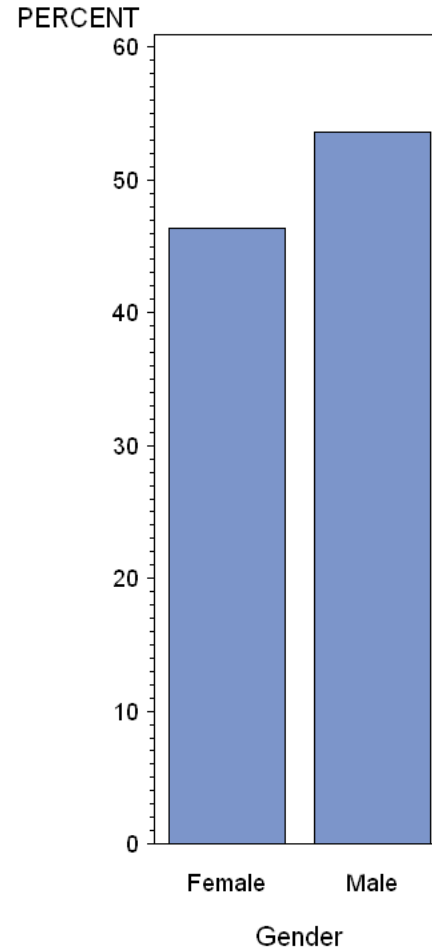
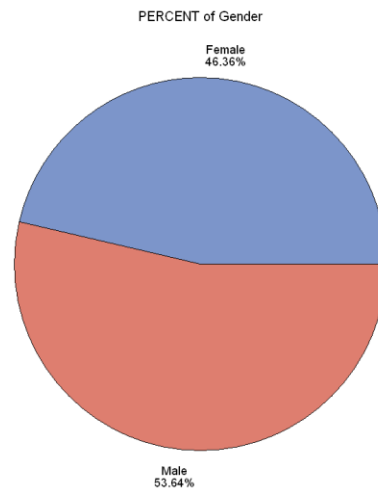
**Proc gchart;**

```
vbar gender / type=percent;
```

```
vbar3d gender;
```

```
pie gender / type = percent;
```

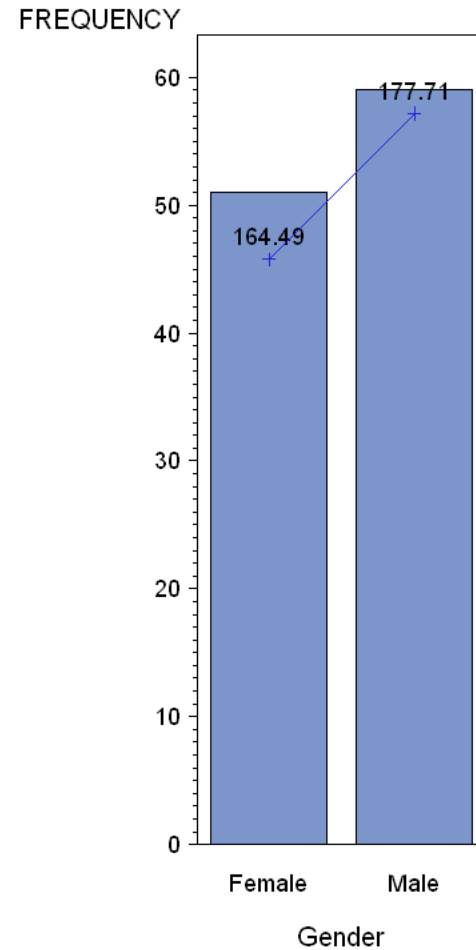
**run;**



Best? We will look later.

# Bar-Line Chart

```
proc gbarline;  
bar gender;  
plot/ sumvar=height mean noaxis;  
run;
```



# Scatterplots

- Scatterplots are by far some of the most useful graphs for looking at the relationship between variables.
- We can see the two way relationship between quantitative variables, and we can add a third categorical variable for different aspects of the same graph.

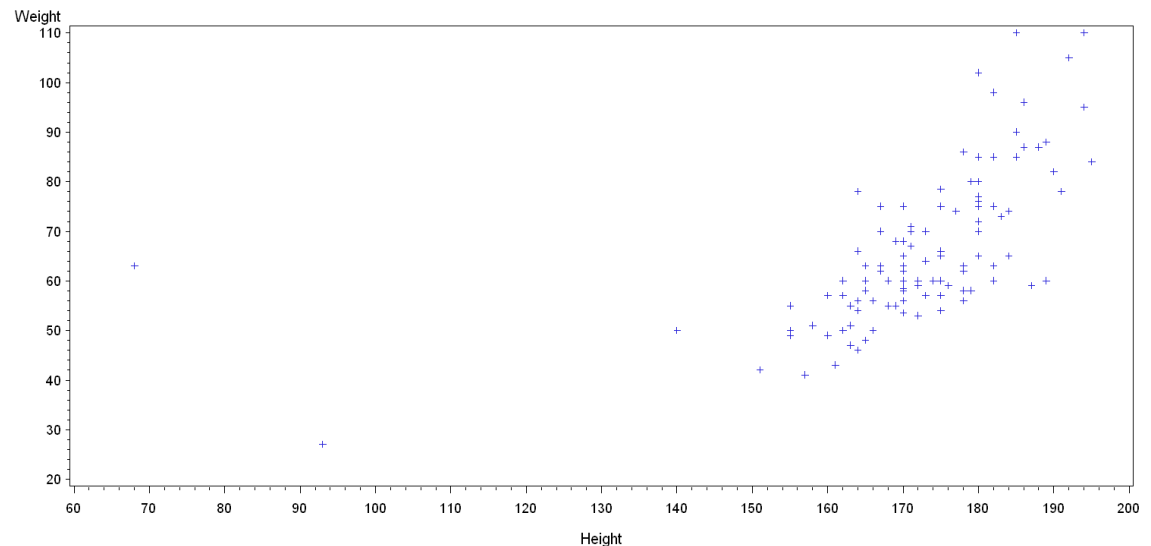
# Simple Plots

\*simple plot y by x;

**proc gplot;**

plot weight\*height;

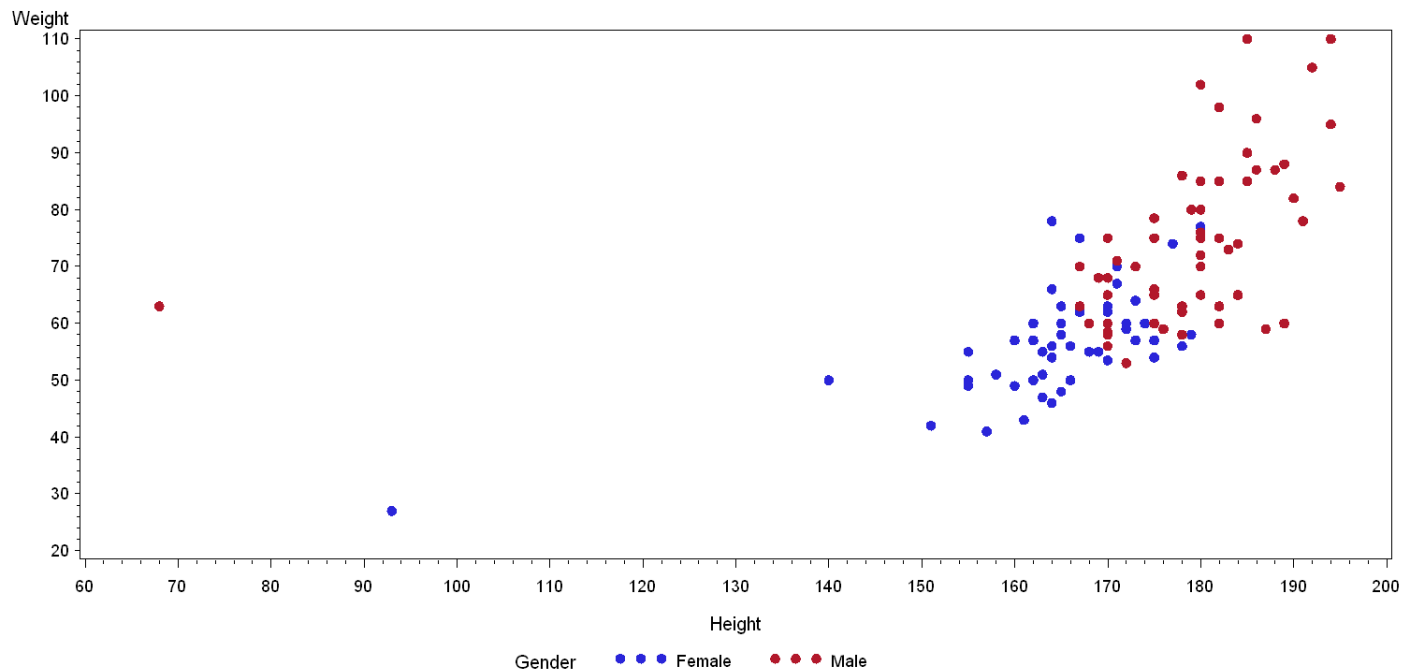
**run;**



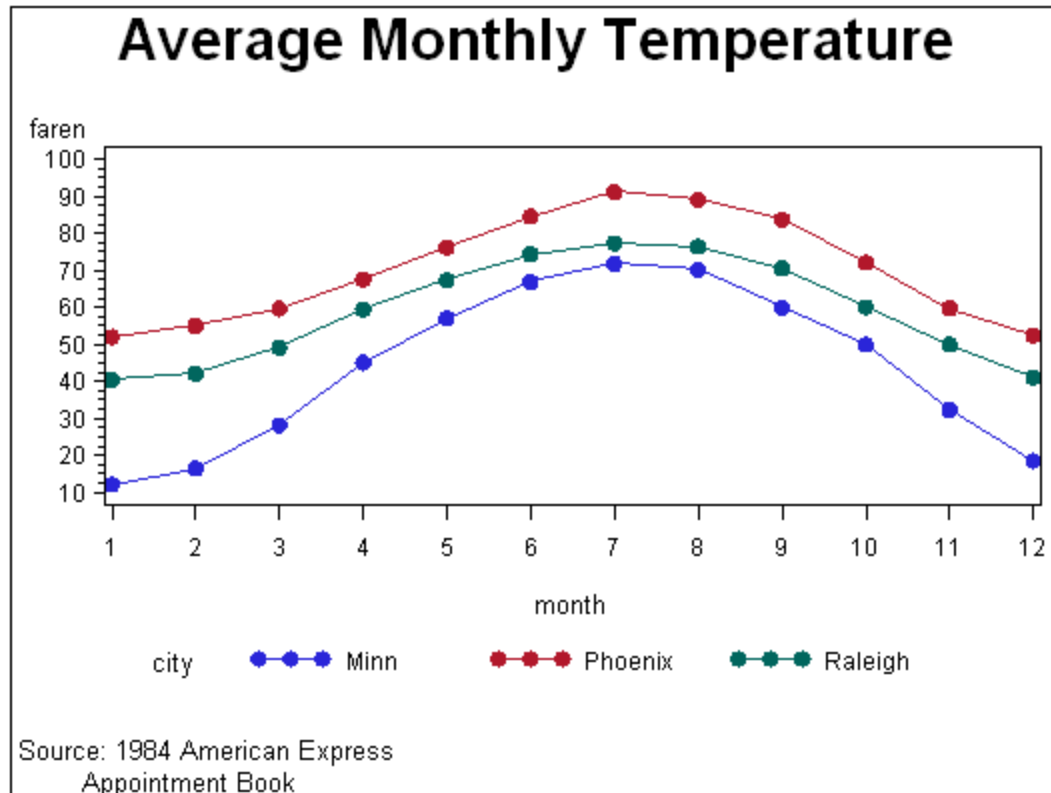
# Better Plots (with options)

- \*scatterplot with options;
- title1 "Plot Height Versus Weight";
- symbol1 value=dot;
- **proc gplot;**
- plot weight\*height=gender;
- **run;**

Plot Height Versus Weight



# Other Options



# New in SAS 9.2 – Statistical Graphics

- Starting in SAS 9.2 a set of powerful new graphical options became available to users. They combine ODS delivery with easy to code graphics and a powerful set of options.
- The most useful (and novel) is that in some cases statistical analysis has been added to the visualizations.

# Proc SGPlot

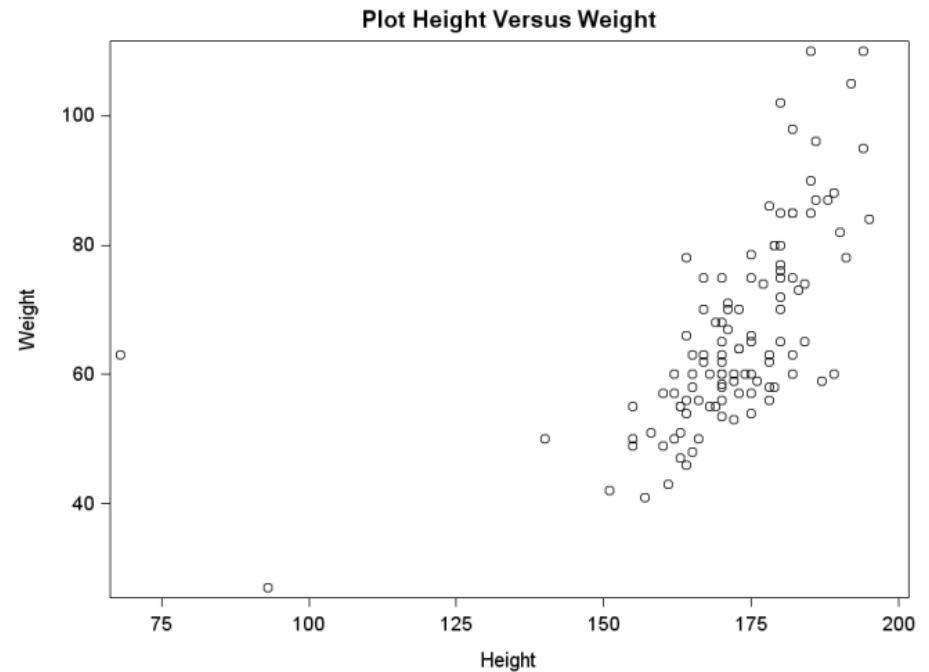
- Proc SGPlot does it all: histograms, bar charts, box plots, and scatterplots.
- Let's look at the height versus weight example.

# Scatterplot - Best

**Proc SGplot;**

**scatter x=height y=weight;**

**run;**



```
proc sgplot;
```

```
  reg x=height y=weight / CLM CLI;
```

```
run;
```



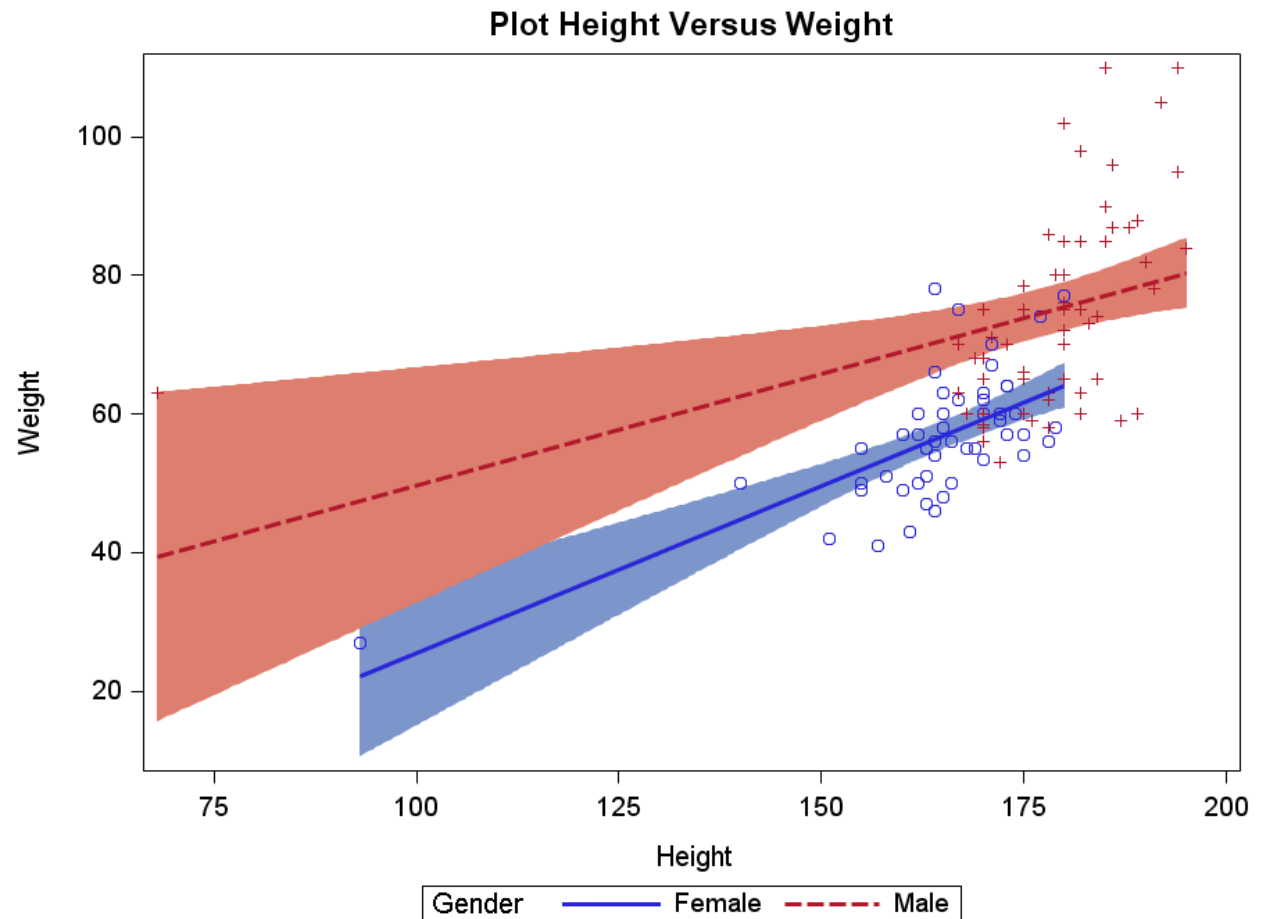
```
proc sgplot;
```

```
  reg x=height y=weight/ group=gender
```

```
  CLM;
```

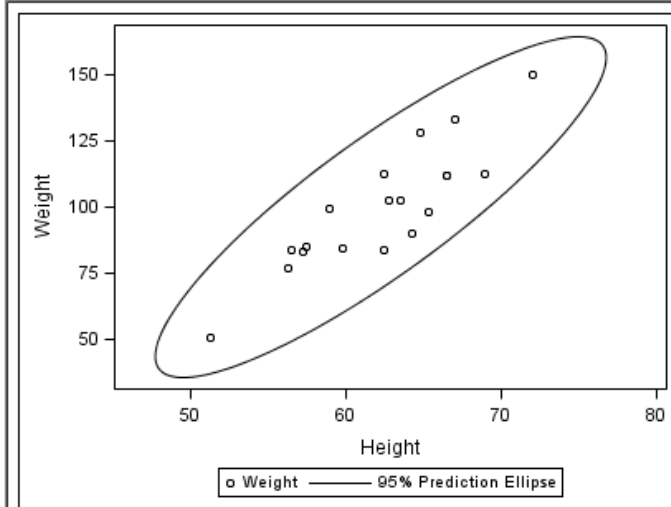
```
run;
```

Different Regression  
Analysis for Different  
Genders.



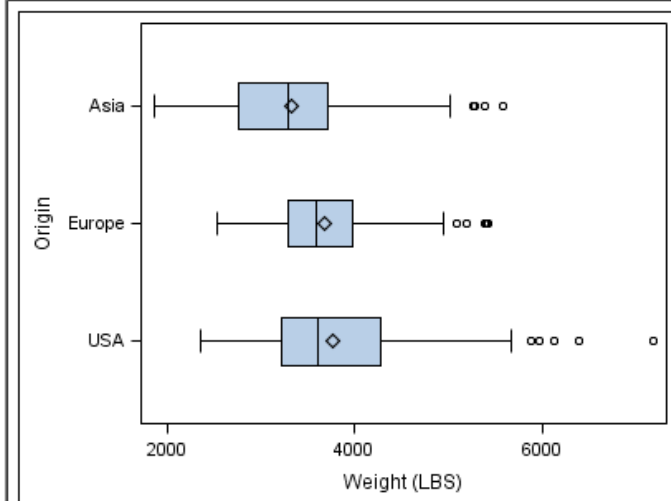
# Examples from SAS Online Help

## Examples of Graphs that Can Be Generated by the SGPLOT Procedure



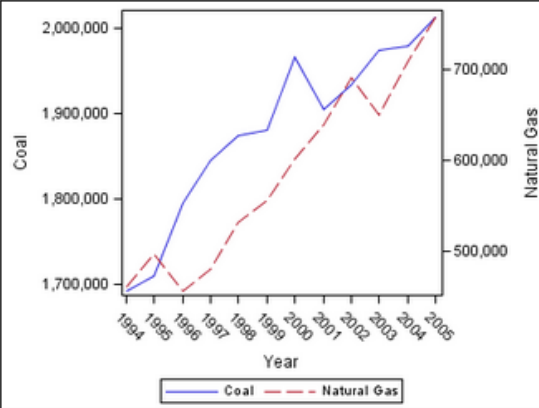
The following code creates an ellipse plot:

```
proc sgplot data=sashelp.class;  
  scatter x=height y=weight;  
  ellipse x=height y=weight;  
run;
```



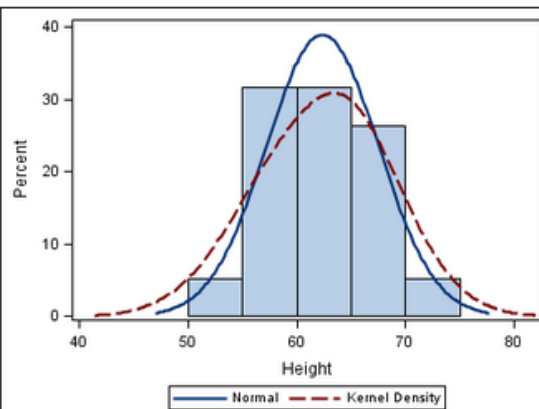
The following code creates a horizontal box plot:

```
proc sgplot data=sashelp.cars;  
  hbox weight / category=origin;  
run;
```



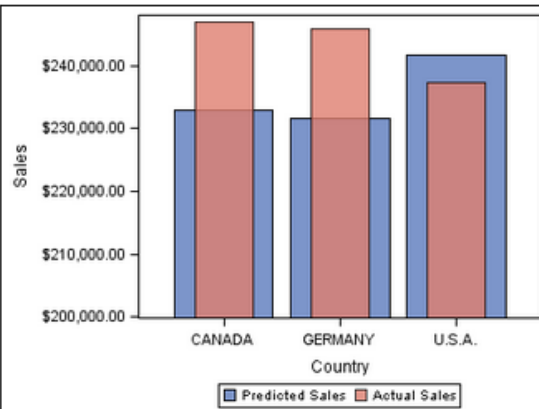
The following code creates a graph with two series plots:

```
proc sgplot data=sashelp.electric(
  where=(customer="Residential"));
  xaxis type=discrete;
  series x=year y=coal;
  series x=year y=naturalgas / y2axis;
run;
```



The following code creates a graph with a histogram, a normal density curve, and a kernel density curve:

```
proc sgplot data=sashelp.class;
  histogram height;
  density height;
  density height / type=kernel;
run;
```



The following code creates a graph with two bar charts:

```
proc sgplot data=sashelp.prdsale;
  yaxis label="Sales" min=200000;
  vbar country / response=predict;
  vbar country / response=actual
    barwidth=0.5
    transparency=0.2;
run;
```

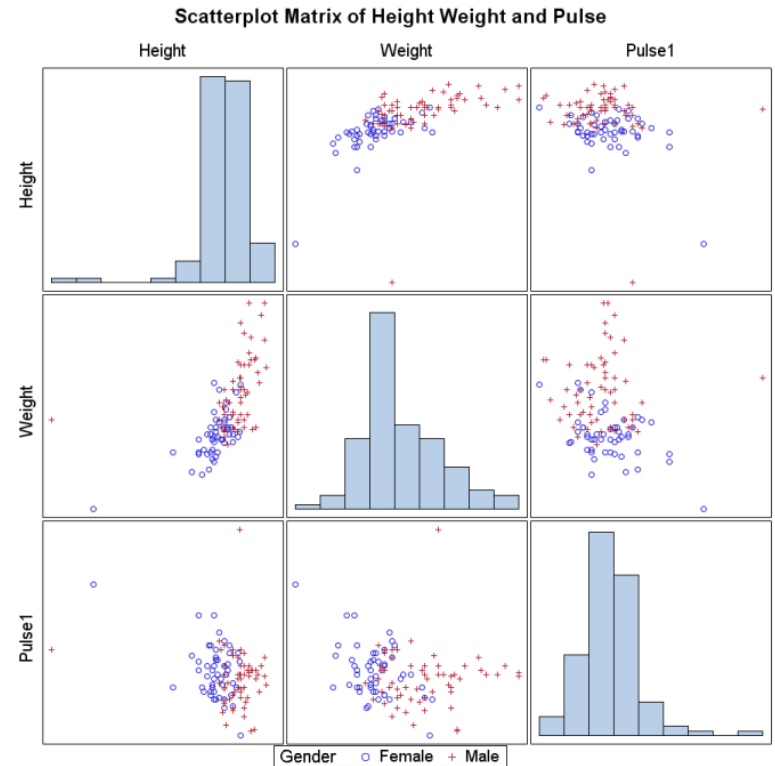
# Proc SGScatter – Scatterplot Matrix

```
title1 "Scatterplot Matrix of Height Weight and Pulse";
```

```
proc sgscatter;
```

```
matrix height weight pulse1/group=gender  
diagonal=(histogram);
```

```
run;
```



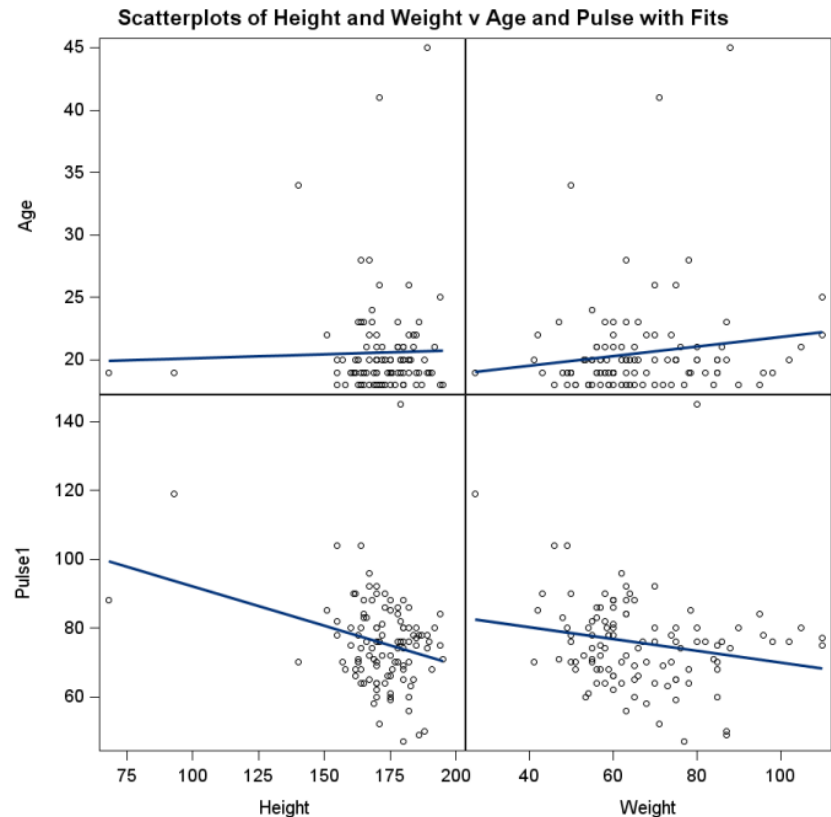
# SGScatter - Compare

```
title1 "Scatterplots of Height and Weight v Age and Pulse with Fits";
```

```
proc sgscatter;
```

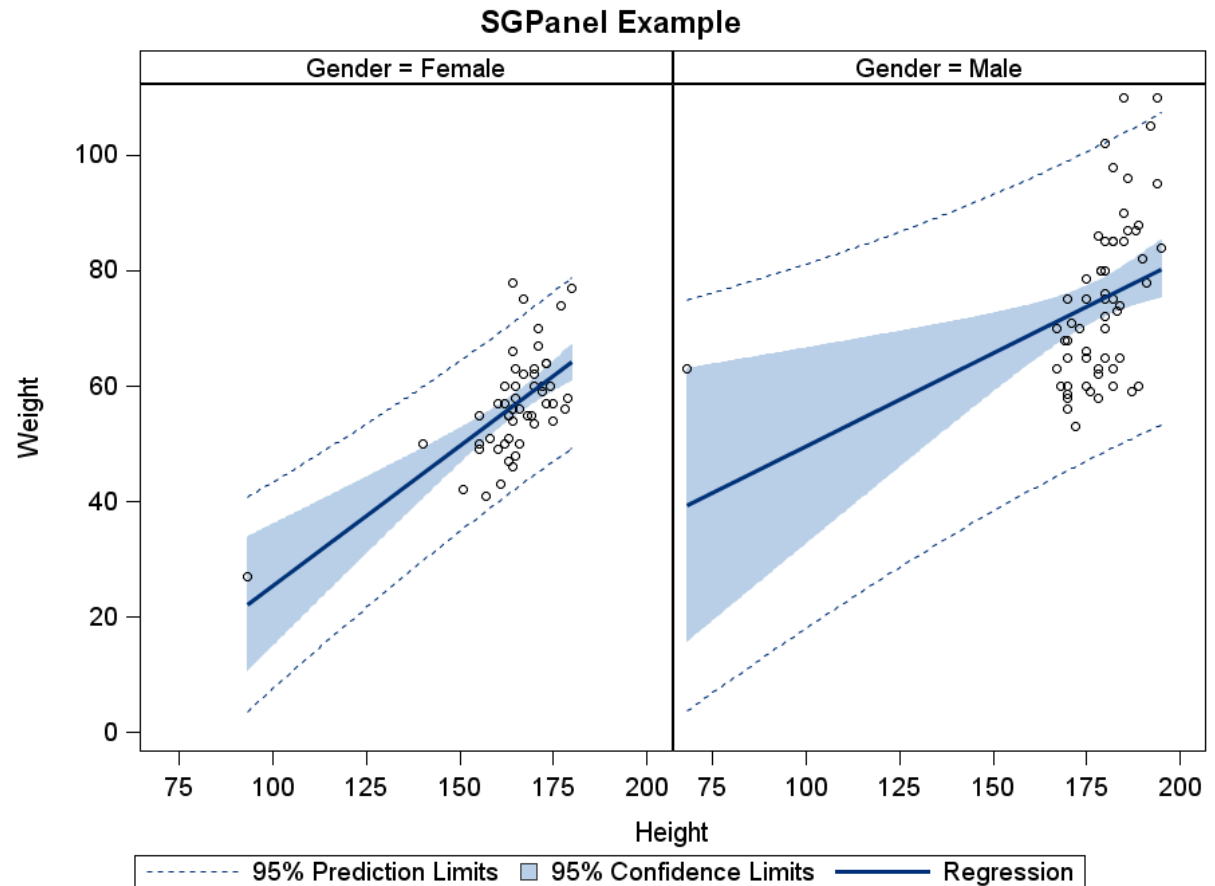
```
Compare x=(height weight) y=(age pulse1)/reg;
```

```
run;
```



# SGPanel – Other ways to group Graphics

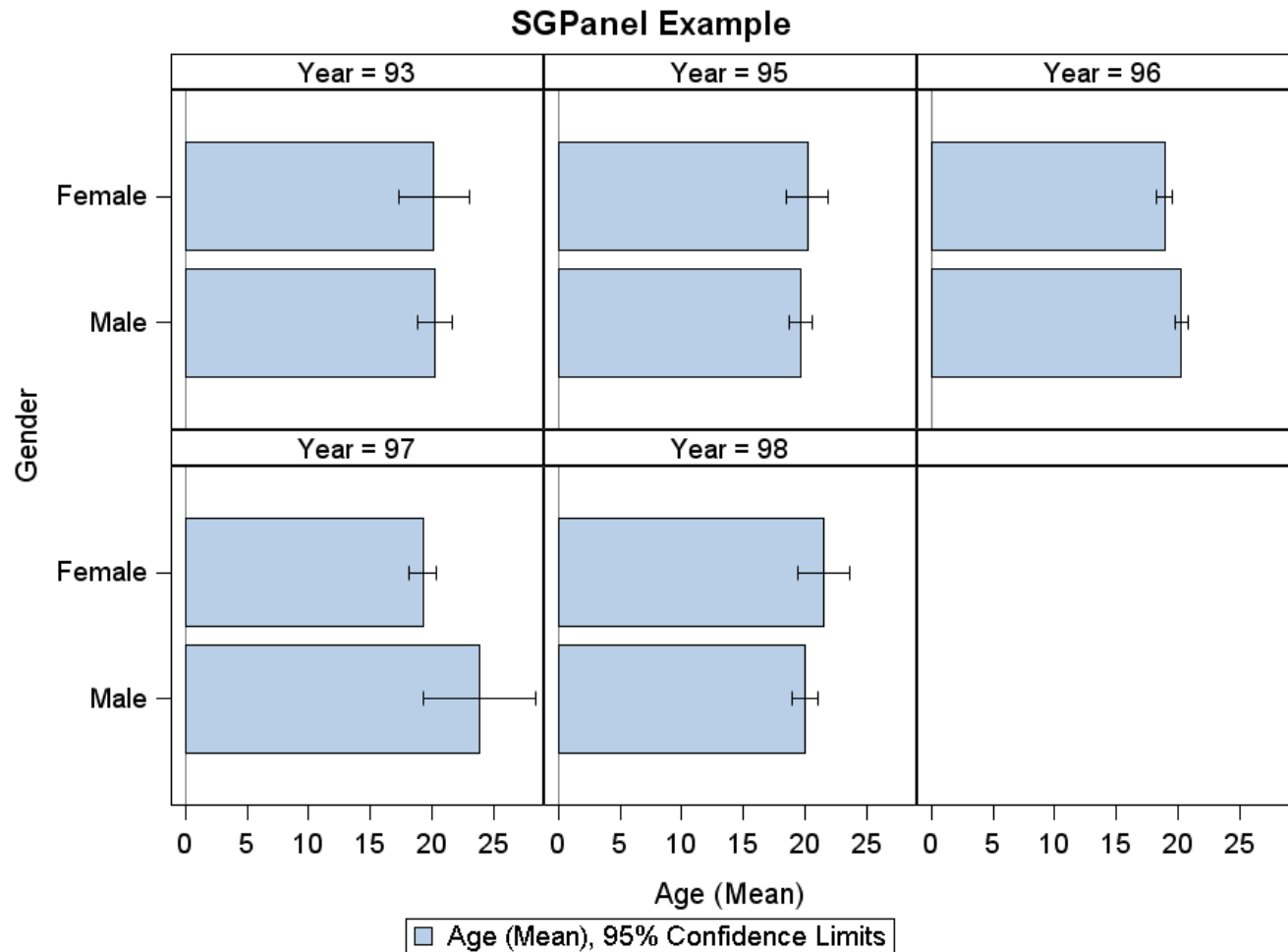
```
title1 "SGPanel Example";  
proc SGpanel;  
panelby gender;  
reg x=height y=weight/cli clm;  
run;
```



```

proc sgpanel;
panelby year;
hbar gender / response=age
stat=mean limits=both;
run;

```



# Additional Resources

- Introduction to Graphics Using SAS/GRAPH® Software – Mike Kalt  
<http://analytics.ncsu.edu/sesug/2010/RIV13.slides.Kalt.pdf>
- SAS Online Documentation  
<http://support.sas.com/cdlsearch?ct=80000>