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.

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

*Print the Data;
Proc Print;
Run;

*Summarize the Data;
Proc Means;
var Height Weight;
Run;

*Sort the data;
Proc Sort;
by Gender;
run;

*Summaries by Gender;
Proc Means;
by Gender;
Var Height Weight;
run;
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;

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;

* The SAS System *
* The MEANS Procedure *

<table>
<thead>
<tr>
<th>Variable</th>
<th>Label</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height</td>
<td>Height</td>
<td>110</td>
<td>171.5818182</td>
<td>16.0768691</td>
<td>68.0000000</td>
<td>195.0000000</td>
</tr>
<tr>
<td>Weight</td>
<td>Height</td>
<td>110</td>
<td>66.3318182</td>
<td>15.1568674</td>
<td>27.0000000</td>
<td>110.0000000</td>
</tr>
</tbody>
</table>
ods graphics on;
*Best Boxplot;
proc boxplot ;
    plot height*gender / boxstyle=schematic 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

```sas
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

![Average Monthly Temperature Graph]

Source: 1984 American Express Appointment Book
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:
```plaintext
proc sgplot data=sashelp.class;
    scatter x=height y=weight;
    ellipse x=height y=weight;
    run;
```

The following code creates a horizontal box plot:
```plaintext
proc sgplot data=sashelp.cars;
    hbox weight / category=origin;
    run;
```
The following code creates a graph with two series plots:

```r
proc gplot data=nsbloy._electric( where=(customer="Residential"));
  title1 "\series y\=\textit{Coal}",
  title2 "\series y\=\textit{Natural Gas}";
  plot y\*year / vaxis=year;
run;
```

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

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

The following code creates a graph with two bar charts:

```r
proc sgplot data=sashelp.prdsale;
  yaxis label="Sales" min=0 max=240000;
  vbar country / response=predict;
  vbar country / response=actual
    barwidth=0.5
    transparency=0.2;
run;
```
title1 "Scatterplot Matrix of Height Weight and Pulse";
proc sgscatter;
matrix height weight pulse1/group=gender
diagonal=(histogram);
run;
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

• SAS Online Documentation
  http://support.sas.com/cdlsearch?ct=80000