Creating Standard R Plots

SP4R04d02.sas

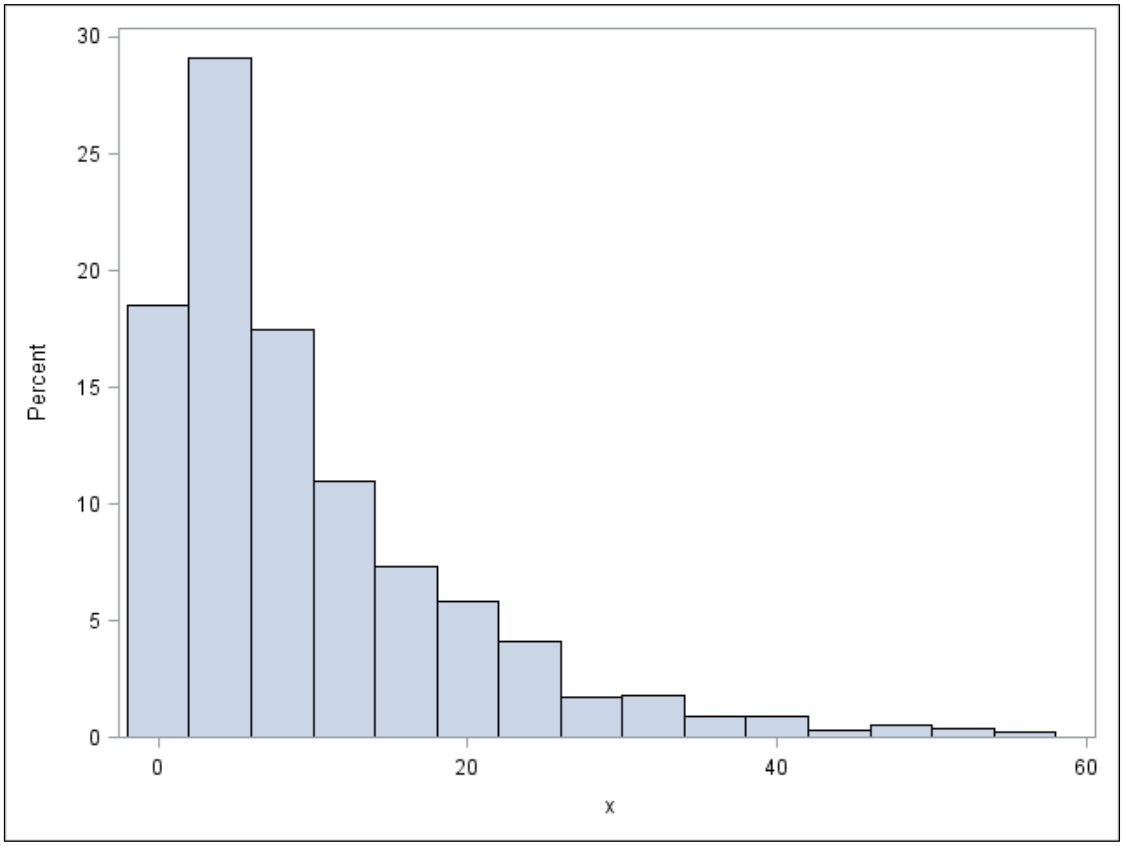
1. Create a data table of 1000 random deviates from an exponential distribution with a mean of 10. Create a histogram of the new variable.

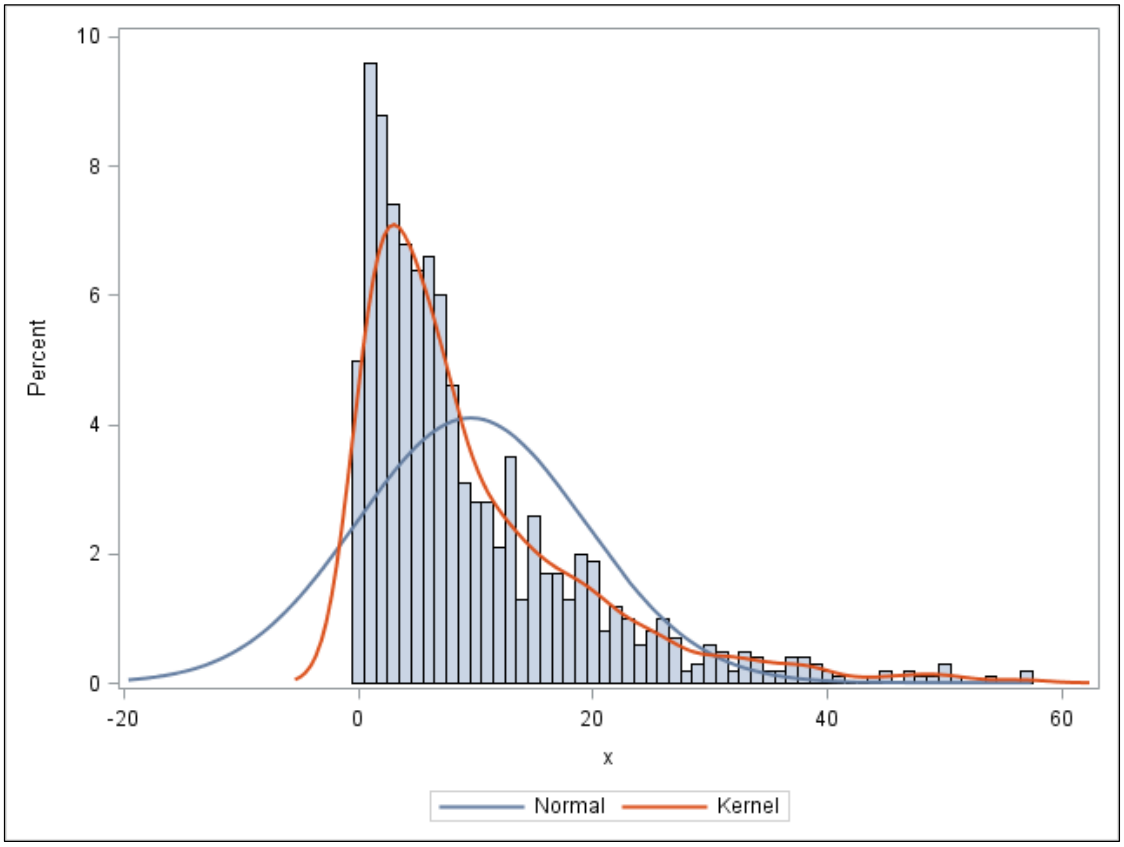
```
data sp4r.hist_data;
   call streaminit(123);
   do i=1 to 1000;
      x = rand('exponential')*10;
      output;
   end;
run;

proc sgplot data=sp4r.hist_data;
   histogram x;
run;

proc sgplot data=sp4r.hist_data;
   histogram x / binwidth=1;
   density x / type=normal;
   density x / type=kernel;
run;
```

The density estimate is overlaid on the histogram when both statements appear in the SGPLOT procedure.

Selected PROC SGPLOT statements and options:

HISTOGRAM	creates a histogram of the proceeding variable. Use the BINWIDTH= option to alter the bin size.
DENSITY	creates a density estimate of the proceeding variable. Use the TYPE= option to estimate the normal distribution or the kernel distribution.



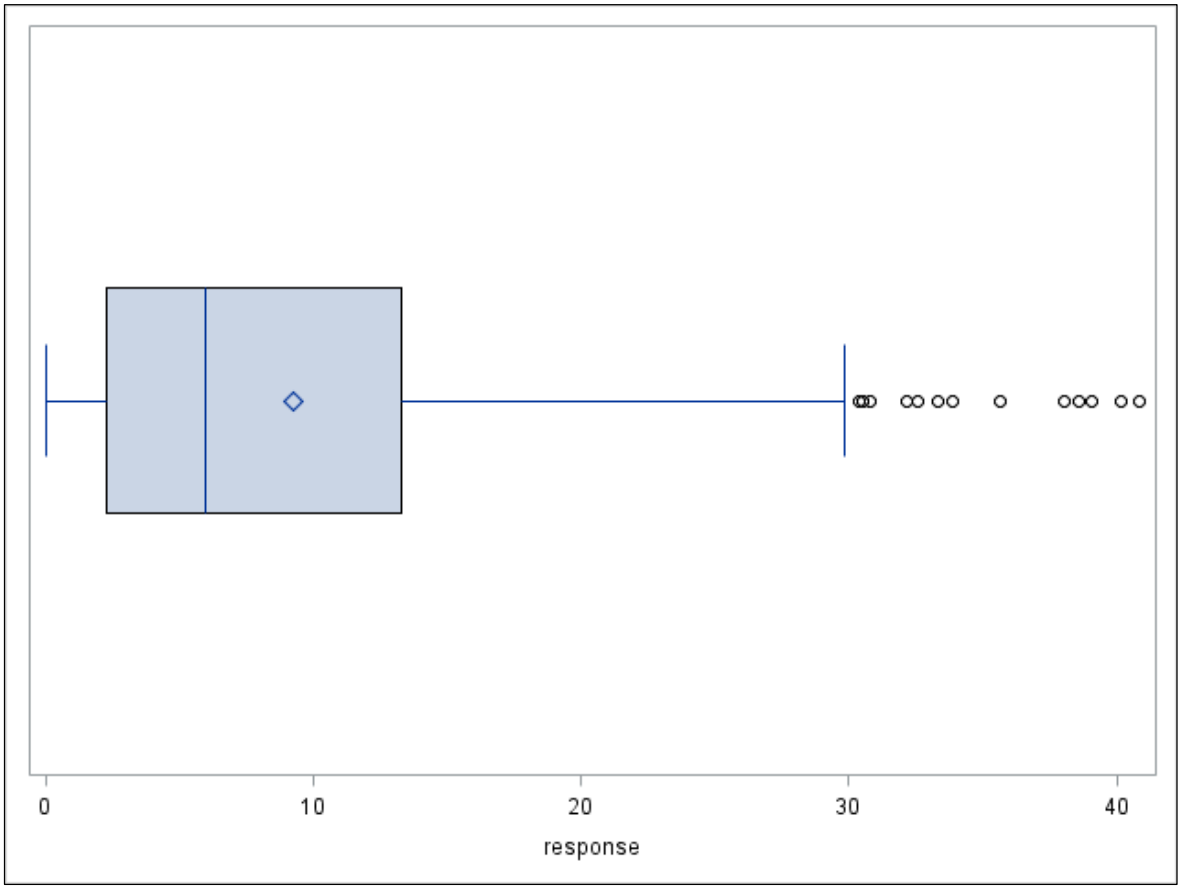


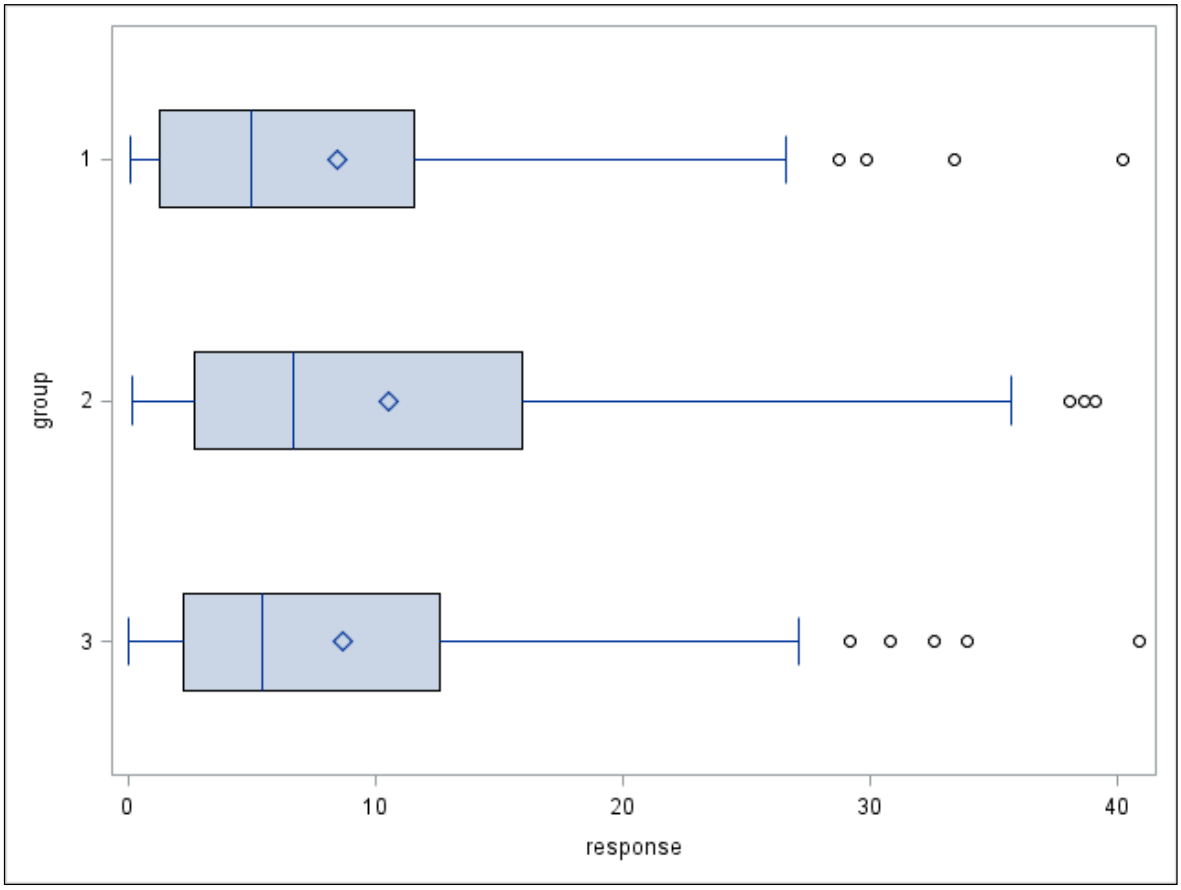
2. Create a new data table with three groups of 100 observations each. Let each observation be a random deviate from an exponential distribution with a mean of 10. First, create a horizontal box plot that ignores the **group** variable. Second, create a horizontal box plot with the CATEGORY= option.

```
data sp4r.boxplot_data (drop=rep);
   call streaminit(123);
   do group=1 to 3;
      do rep=1 to 100;
         response = rand('exponential')*10;
         output;
      end;
   end;
run;

proc sgplot data=sp4r.boxplot_data;
   hbox response;
run;

proc sgplot data=sp4r.boxplot_data;
   hbox response / category=group;
run;
```

Selected PROC SGPLOT statement and option:

HBOX creates a horizontal box plot of the response variable. Use the CATEGORY= option to create a box plot for each category of a classification variable.



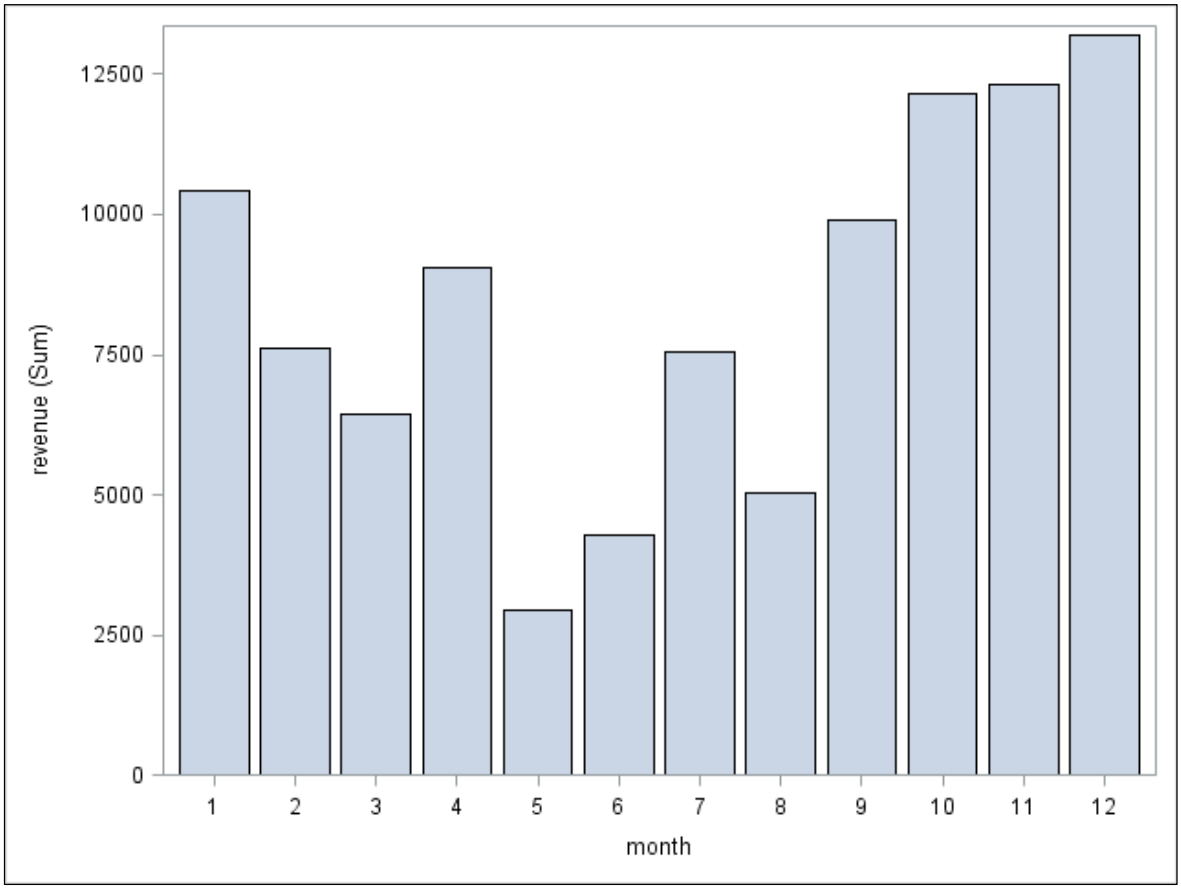


3. Create a data table with random deviates generated from a normal distribution with a mean of 10,000 and a standard deviation of 5000. This corresponds to the revenue for a 12-month period. Create a bar chart of the newly generated data.

```
data sp4r.sales;
   call streaminit(123);
   do month=1 to 12;
      revenue = rand('Normal',10000,5000);
      output;
   end;
run;

proc sgplot data=sp4r.sales;
   vbar month / response=revenue;
run;
```

Selected PROC SGPLOT statement and option:

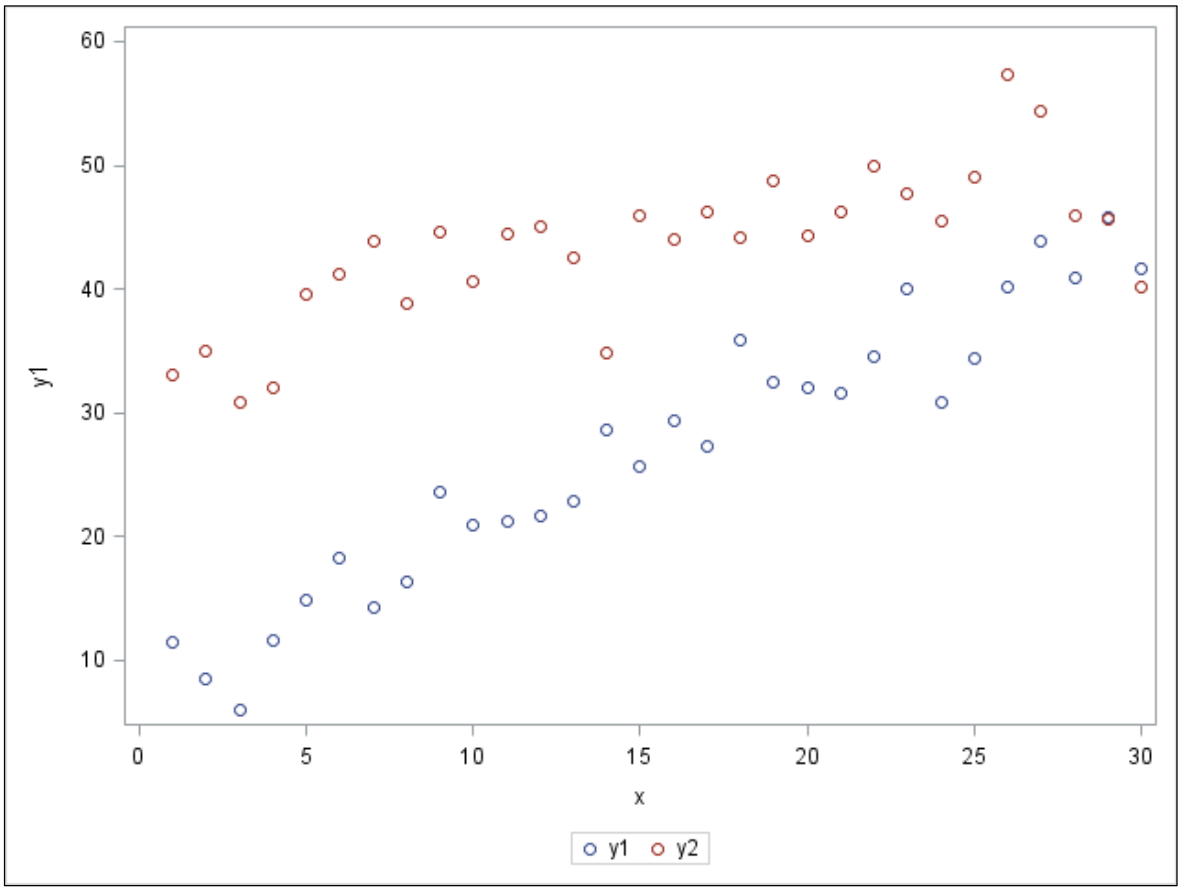
VBAR creates a vertical bar chart. The proceeding variable is the category. Use the RESPONSE= option to specify the response for each category to be plotted.



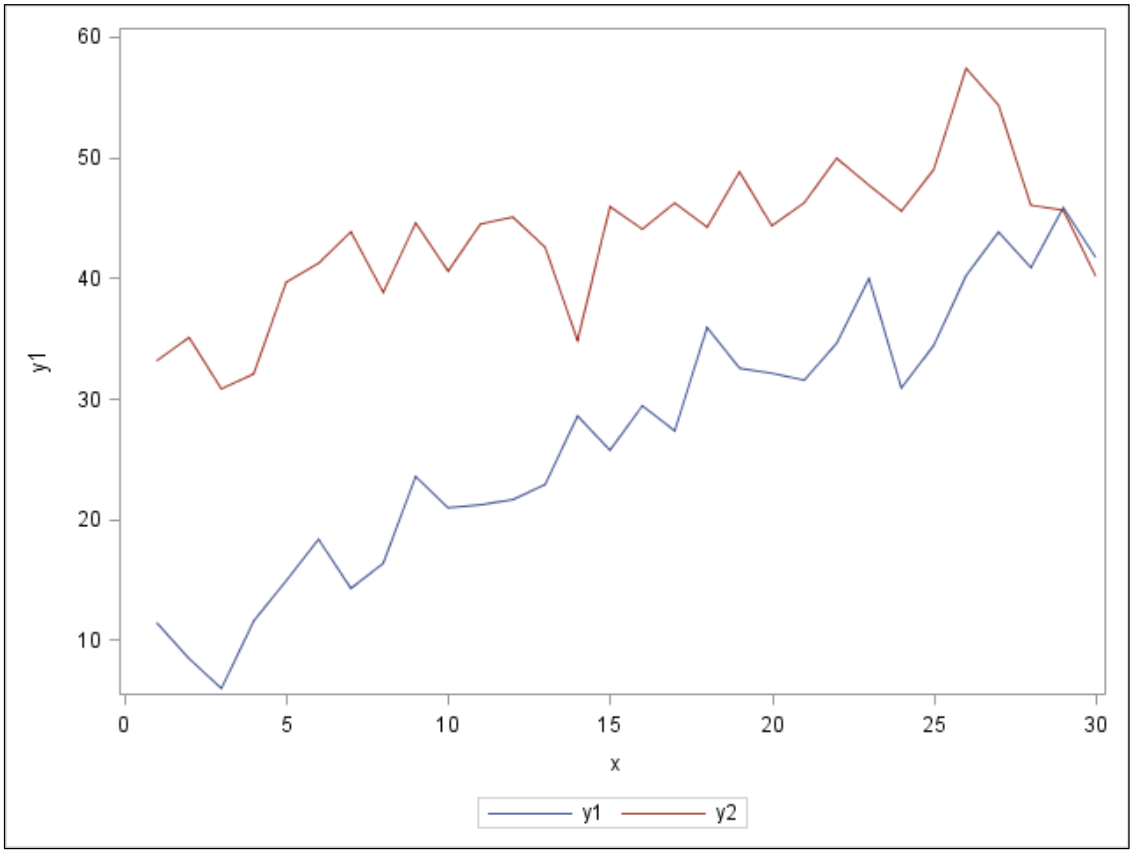
4. Create a data table with $Y_1 = 10 + X + \varepsilon$ and $Y_2 = 35 + \frac{X}{2} + \varepsilon$ where ***X*** ranges from 1 to 30 and $\varepsilon \sim N(0, \sigma = 5)$. Keep only the variables ***X***, ***Y1***, and ***Y2***. Use PROC SGPLOT to create a scatter plot and a series plot separately for both variables ***Y1*** and ***Y2***. Then merge the two plots by providing multiple SCATTER and SERIES statements to a single SGPLOT procedure.

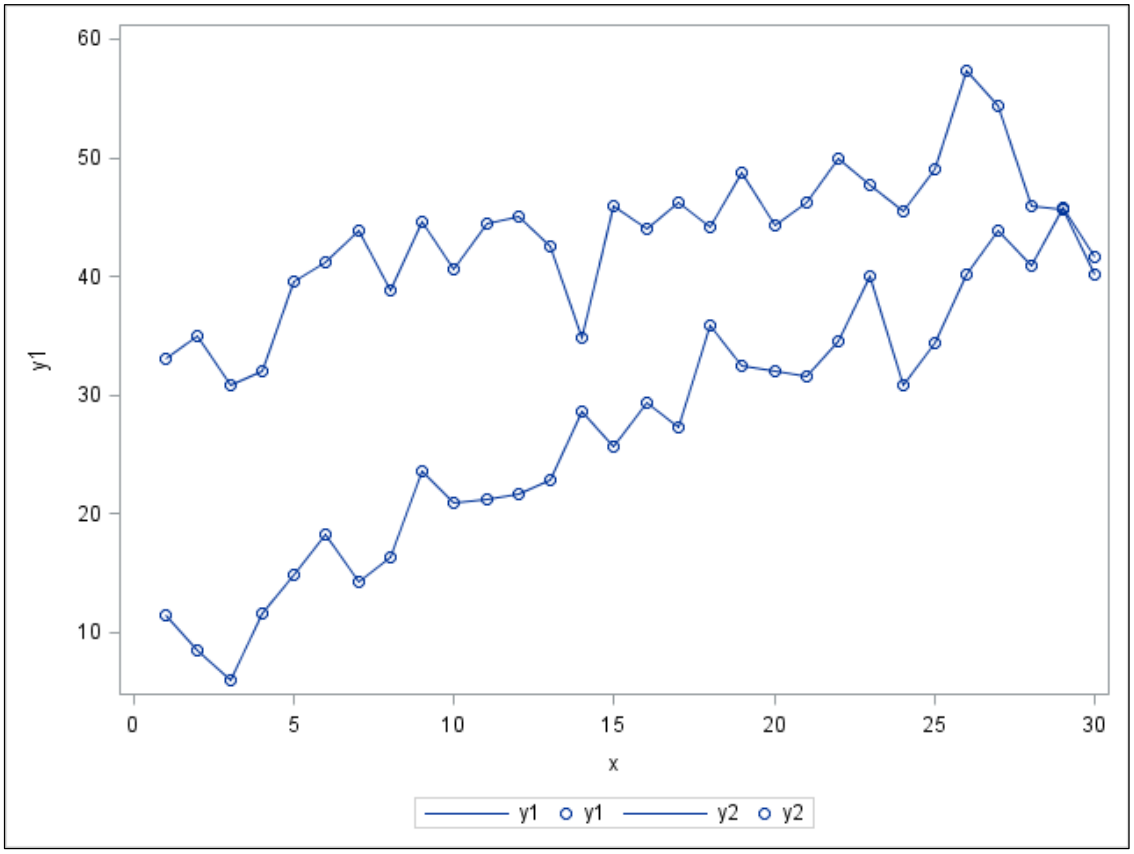
```
data sp4r.series_data (keep=x y1 y2);
   call streaminit(123);
   do x=1 to 30;
      beta01 = 10;
      beta11 = 1;
      y1 = beta01 + beta11*x + rand('Normal',0,5);
      beta02 = 35;
      beta12 = .5;
      y2 = beta02 + beta12*x + rand('Normal',0,5);
      output;
   end;
run;

proc sgplot data=sp4r.series_data;
   scatter x=x y=y1;
   scatter x=x y=y2;
run;
```

```
proc sgplot data=sp4r.series_data;
   series x=x y=y1;
   series x=x y=y2;
run;

proc sgplot data=sp4r.series_data;
   series x=x y=y1;
   scatter x=x y=y1;
   series x=x y=y2;
   scatter x=x y=y2;
run;
```



Notice that the Y-axis label uses the first Y-axis variable input name. This can be altered using the YAXIS LABEL= option.





Notice that the legend specifies a name for both SCATTER and SERIES statements. This can be altered by using the KEYLEGEND statement and the NAME option.

5. Using the data table from step 4, add a line of best fit to the scatter plot for both Y_1 and Y_2.

```
proc sgplot data=sp4r.series_data;
   reg x=x y=y1 / clm cli;
   reg x=x y=y2 / clm cli;
run;
```

Selected PROC SGPLOT statement and options:

REG creates a fitted regression line to a scatter plot. Use the CLM and CLI options to include prediction and confidence limits. Limits corresponding to alpha=0.05 is the default.