

Chapter 3 SAS Graphs

This chapter introduces a SAS procedure called **PROC SGPLOT**. The SGPLOT procedure can create a wide variety of plot types and can overlay plots together to produce many different types of graphs, such as histogram, boxplot, scatterplot, line plot, and bar chart.

We will use the datasets in SASHELP library. No need for us to import data into SAS. Check out SASHELP library to see the existing datasets. This documents describe each dataset:
<https://support.sas.com/documentation/tools/sashelpug.pdf>.

We use PROC PRINT to take a look at the first five observations (obs=5) of *sashelp.class*.

Code Editor:

```
*Dataset class in SASHELP library*;
proc print data=sashelp.class (obs=5);
run;
```

Results Window:

Obs	Name	Sex	Age	Height	Weight
1	Alfred	M	14	69.0	112.5
2	Alice	F	13	56.5	84.0
3	Barbara	F	13	65.3	98.0
4	Carol	F	14	62.8	102.5
5	Henry	M	14	63.5	102.5

Then look at frequency of division and product types using PROC FREQ.

Code:

```
*Descriptive statistics*;
proc means data=sashelp.class;
  var age height weight;
run;

proc freq data=sashelp.class;
  tables sex;
run;
```

Results:

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
Age	19	13.3157895	1.4926722	11.0000000	16.0000000
Height	19	62.3368421	5.1270752	51.3000000	72.0000000
Weight	19	100.0263158	22.7739335	50.5000000	150.0000000

The FREQ Procedure

Sex	Frequency	Percent	Cumulative Frequency	Cumulative Percent
F	9	47.37	9	47.37
M	10	52.63	19	100.00

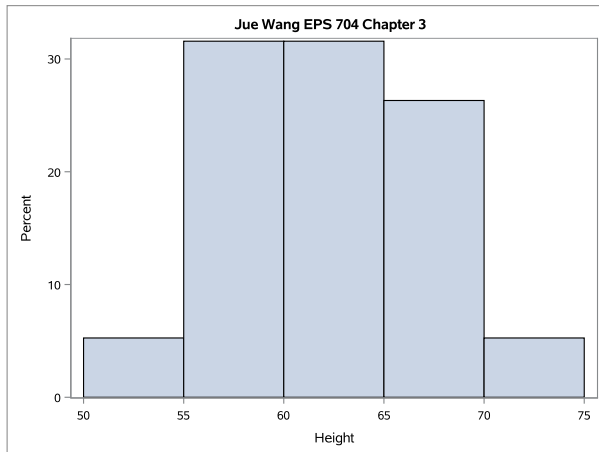
3.1 Histogram – Frequency distribution of continuous variable (Height)

- PROC SGPLOT line: identify the dataset.
- HISTOGRAM statement: simply specify the variable name.

Code Editor:

```
*Simple Histogram for Height*;
proc sgplot data=sashelp.class;
  histogram height;
run;
```

Results Window:

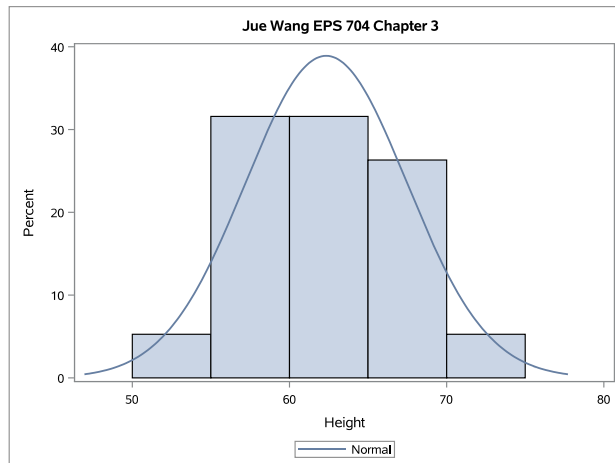


We can also add a normal density curve using DENSITY statement.

Code:

```
*Histogram for Height with normal density curve*;
proc sgplot data=sashelp.class;
  histogram height;
  density height;
run;
```

Results:



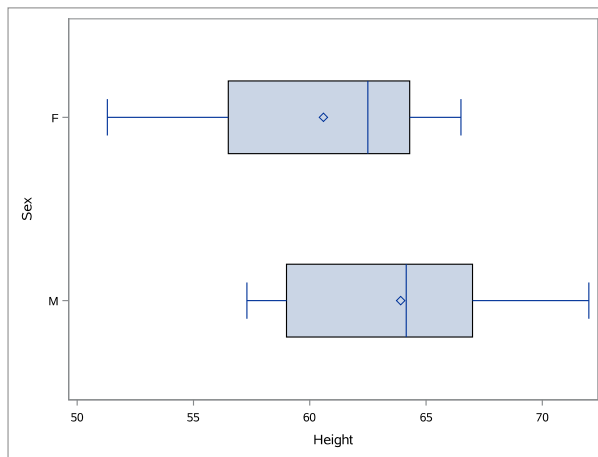
3.2 Boxplot – Frequency distribution of continuous variable (Height by Gender)

Code:

```
*Boxplot -- Height by gender*;
proc sgplot data=sashelp.class;
  hbox height / category=sex;
run;
```

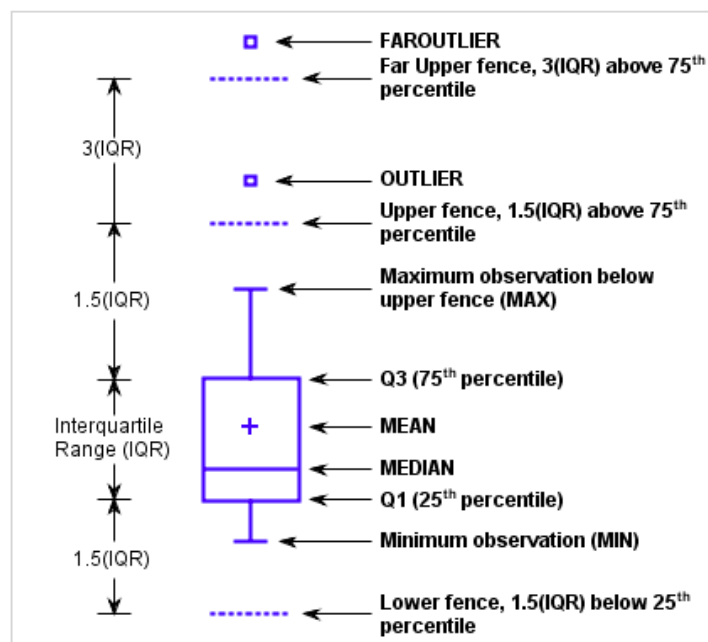
- HBOX statement: specify the continuous variable
- CATEGORY option: identify the groups or categories

Results:



Supplementary material: how to read a Boxplot?

Parts of a Box Plot



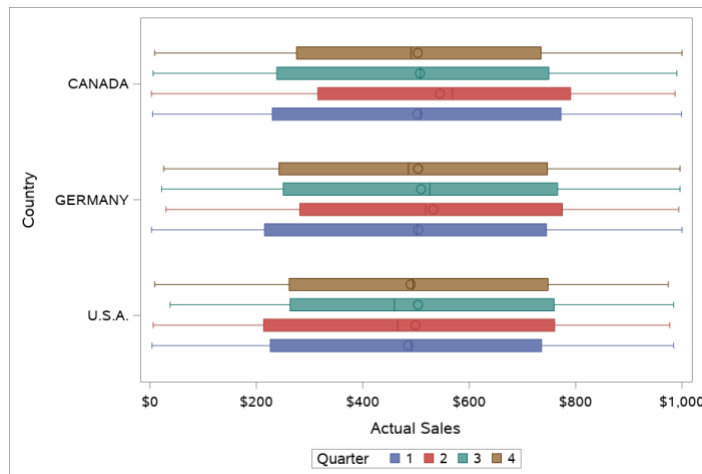
Note: Although it shows a vertical box plot, the basic concepts apply to horizontal box plots.

- GROUP option: add another categorical variable.

Let's look at a different dataset (sashelp.prdsale) showing furniture sales data. Display actual sale by country in each quarter. Code:

```
*Boxplot -- actual sale by country per quarter*;
proc sgplot data=sashelp.prdsale;
  hbox actual / category=country group=quarter;
run;
```

Results:



3.3 Bar chart – Frequency distribution of categorical variable

Using sashelp.prdsale to create two bar charts for two continuous variables overlapping on a single graph.

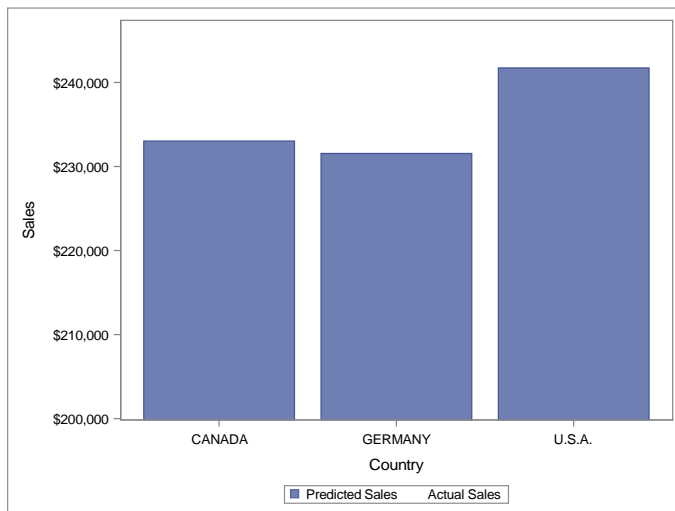
Code:

```
*Bar chart -- two charts overlapped*;
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;
```

- YAXIS statement: label – defines the label for the axis; MIN and MAX defines the range of the y-scale.
- VBAR statement: specify groups or categories for the bars. Can have multiple VBAR statements, but each one needs to be clearly described.
- RESPONSE option: define the continuous variable to be displayed by groups.

- Other options: barwidth – width of each bar (as a ratio of the maximum possible width); transparency – degree of transparency for the plot ranging from 0 (not transparent) to 1 (totally transparent).

Results:

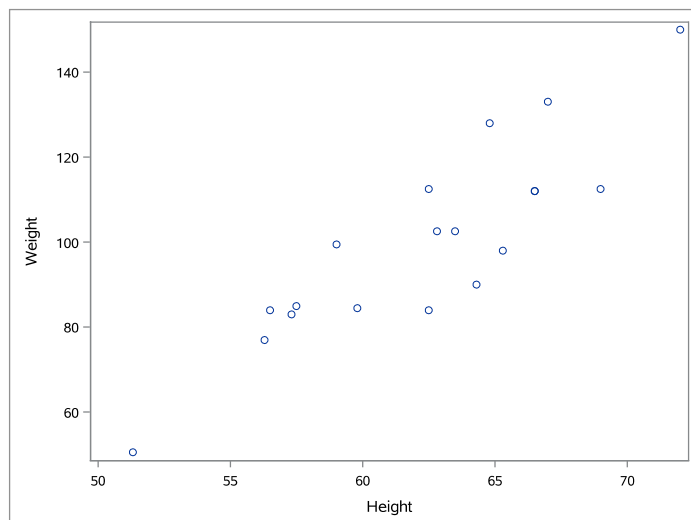


3.4 Scatterplot – Relationship between two continuous variables

a. Using SCATTER statement: specify variables on the x-axis and y-axis separately. Code:

```
*Scatterplot -- Height by Weight*;
proc sgplot data=sashelp.class;
  scatter x=height y=weight;
run;
```

Results:



b. MARKERATTRS Option for either REG or SCATTER statement.

Marker options specify the appearance of the markers in the plot. Usually we specify color, size, and symbols.

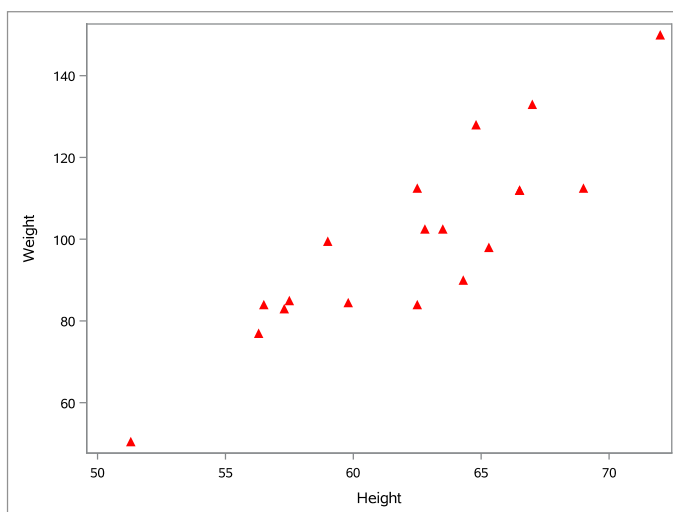
Code:

```
*Scatterplot with user defined markers -- Height by Weight*;
proc sgplot data=sashelp.class;
  scatter x=height y=weight / markerattrs=(color=red size=10 symbol=trianglefilled);
run;
```

- **Color:** Common colors are all available. Here is a complete list of colors supported by SAS Graph: https://support.sas.com/content/dam/SAS/support/en/books/pro-template-made-easy-a-guide-for-sas-users/62007_Appendix.pdf.
- **Size:** specifies the size of the markers.
- **Symbols:** Here are the symbol markers supported by the PROC SGPLOT.

↓ ArrowDown	I Ibeam	◁ TriangleLeft	▼ HomeDownFilled
* Asterisk	+ Plus	▷ TriangleRight	■ SquareFilled
○ Circle	□ Square	∪ Union	★ StarFilled
◇ Diamond	☆ Star	× X	▲ TriangleFilled
> GreaterThan	┐ Tack	Y Y	▼ TriangleDownFilled
< LessThan	⌋ Tilde	Z Z	◀ TriangleLeftFilled
# Hash	△ Triangle	● CircleFilled	▶ TriangleRightFilled
▽ HomeDown	▽ TriangleDown	◆ DiamondFilled	

Results:



c. Display Height-Weight relationship by gender using GROUP option. Code:

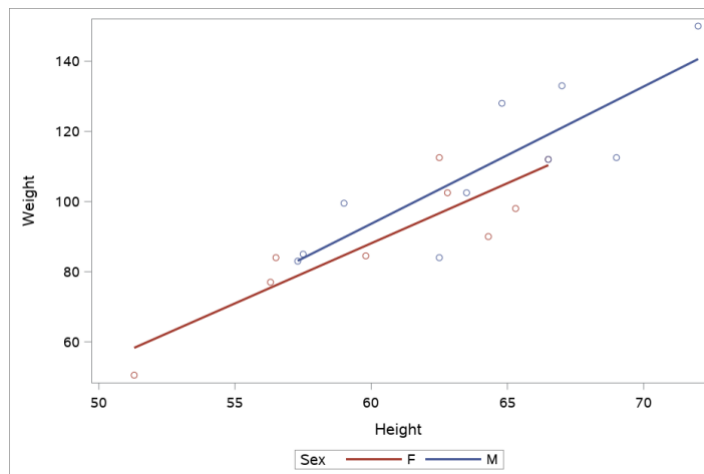
```
*Scatterplot -- Height by Weight separated by gender*;
proc sgplot data=sashelp.class;
  scatter x=height y=weight / group=sex;
run;
```

Check out the graph by yourself.

d. Using REG statement: specify dependent (y) and independent variable (x), and they will be displayed on the x-axis and y-axis correspondingly. Provide separate lines for female and male using GROUP option. Code:

```
*Scatterplot with regression line -- Height by Weight*;
proc sgplot data=sashelp.class;
  reg y=weight x=height / group=sex;
run;
```

Results:



3.5 Line plot – Changes over time

Let's look at another data set (sashelp.stocks) – Performance of Three Stocks from 1996 to 2005. The data set contains 699 observations.

Selected Variables	Type	Description
Stock	Character	IBM, Intel, and Microsoft
Date	Number	Date
Close	Number	Closing price on a specific date
High	Number	Highest price on a specific date
Low	Number	Lowest price on a specific date

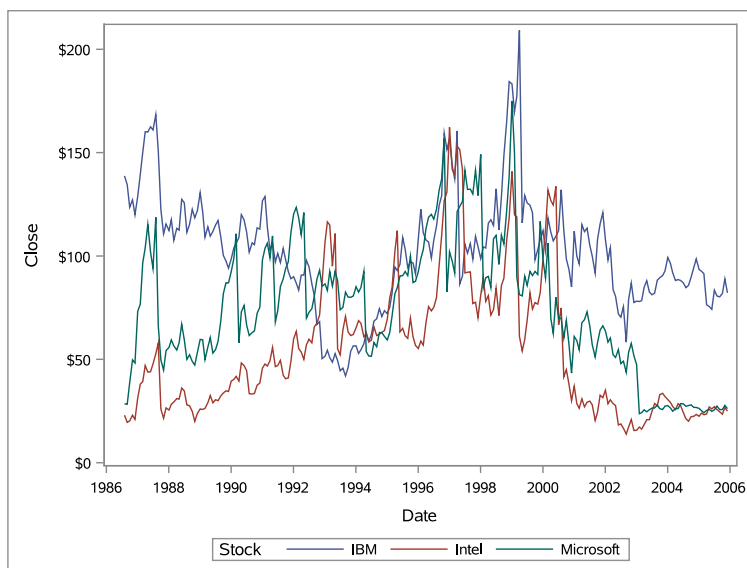
Let's see the closing price between 1996 and 2006 for the three stocks.

Example I -- Code:

```
*Line plot -- Three stocks over years*;
proc sgplot data=sashelp.stocks;
  series x=date y=close / group=stock;
run;
```

- SERIES statement: specify the x-axis and y-axis. Usually time variable is on the x-axis.
- GROUP option: define groups or categories for separate lines.

Results:

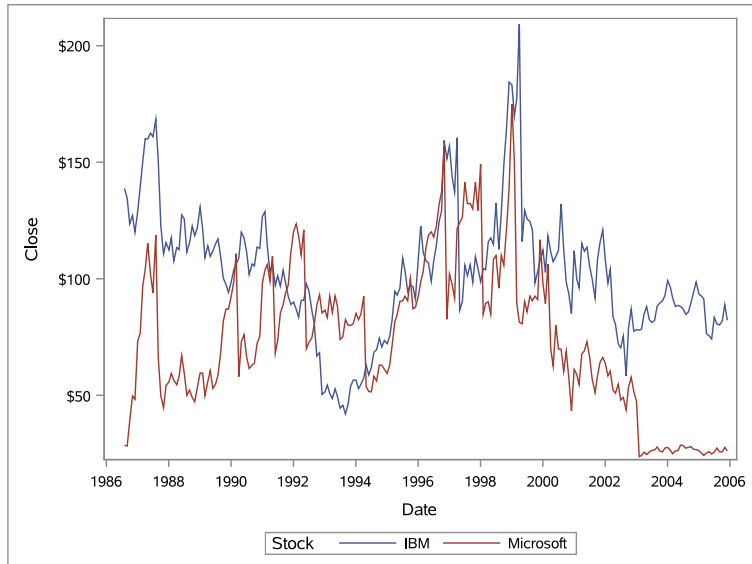


Example II -- Code:

```
*Line plot -- Display only two stocks over years*;
proc sgplot data=sashelp.stocks;
  where stock in ('IBM', 'Microsoft');
  series x=date y=close / group=stock;
run;
```

- WHERE statement: create the graph using selected data responses, e.g., only display the stocks for IBM and Microsoft.

Results:

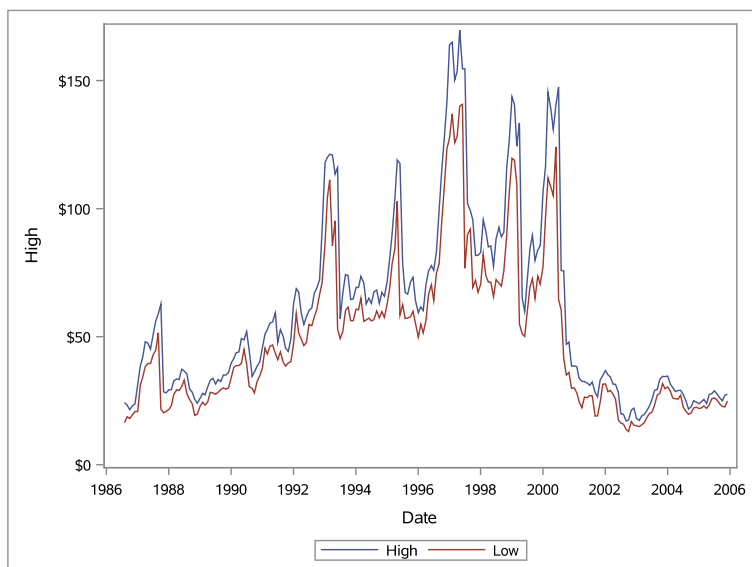


Example III -- Code:

```
*Line plot -- Multiple SERIES statements*;
proc sgplot data=sashelp.stocks;
  where stock in ('Intel');
  series x=date y=high;
  series x=date y=low;
run;
```

- Use multiple SERIES statements to draw separate lines (same x-axis).

Results:

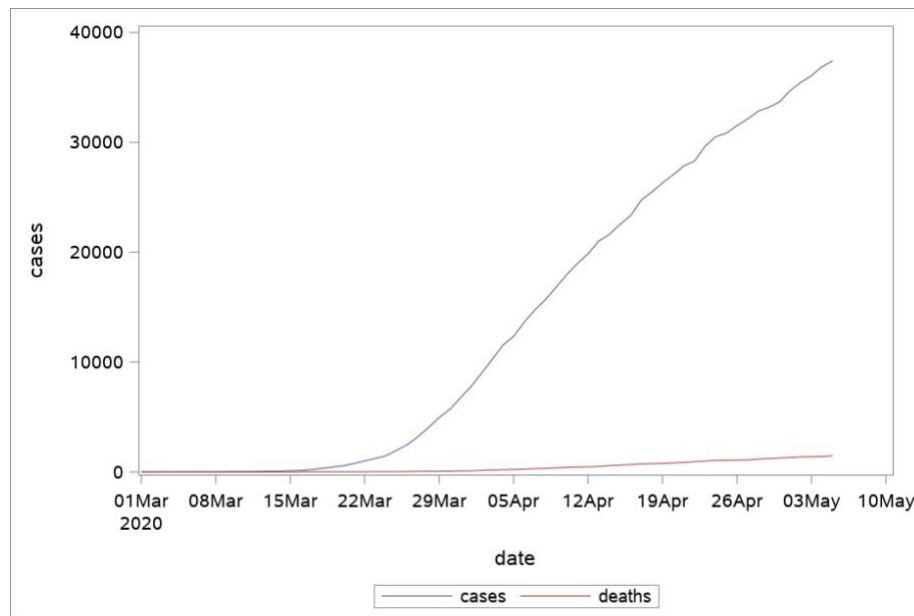


Two more examples based on COVID-19 confirmed cases (up to 5/5/2020).

Example I – Florida (confirmed cases and death). Code:

```
*Line plot -- COVID-19 cases of Florida*;
proc sgplot data=import;
  where state in ('Florida');
  series x=date y=cases;
  series x=date y=deaths;
run;
```

Results:



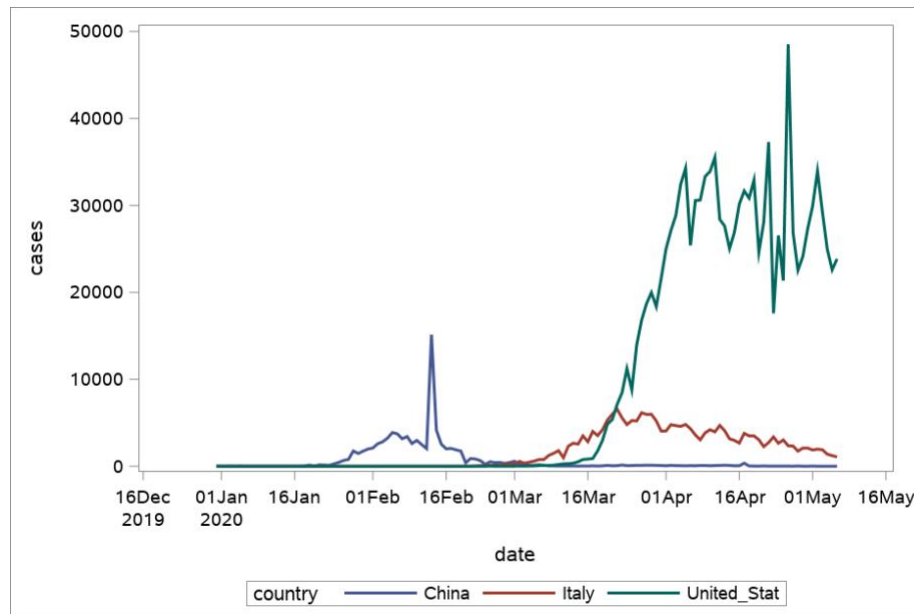
Note: We can also specify a different y-axis for deaths using Y2AXIS in SERIES statement.

```
proc sgplot data=import;
  where state in ('Florida');
  series x=date y=cases;
  series x=date y=deaths / y2axis;
run;
```

Example II – Confirmed cases of three countries (China, Italy, and USA). Code:

```
*Line plot -- COVID-19 cases of China, Italy, and USA*;
proc sgplot data=world;
  where countrycode in ('CHN','ITA','USA');
  series x=date y=cases / group=country Lineattrs = (thickness = 2.5);
run;
```

Results:



Final Comments: There are many statements in SGPLOT procedure. Within each statement, there are also many options available which creates lots of flexibilities for creating graphical displays in SAS. Play with it! For more information, please check out SAS documentation: <https://documentation.sas.com/?docsetId=grstatproc&docsetTarget=n0yjjdd910dh59zn1toodgupa.j4v9.htm&docsetVersion=9.4&locale=en>