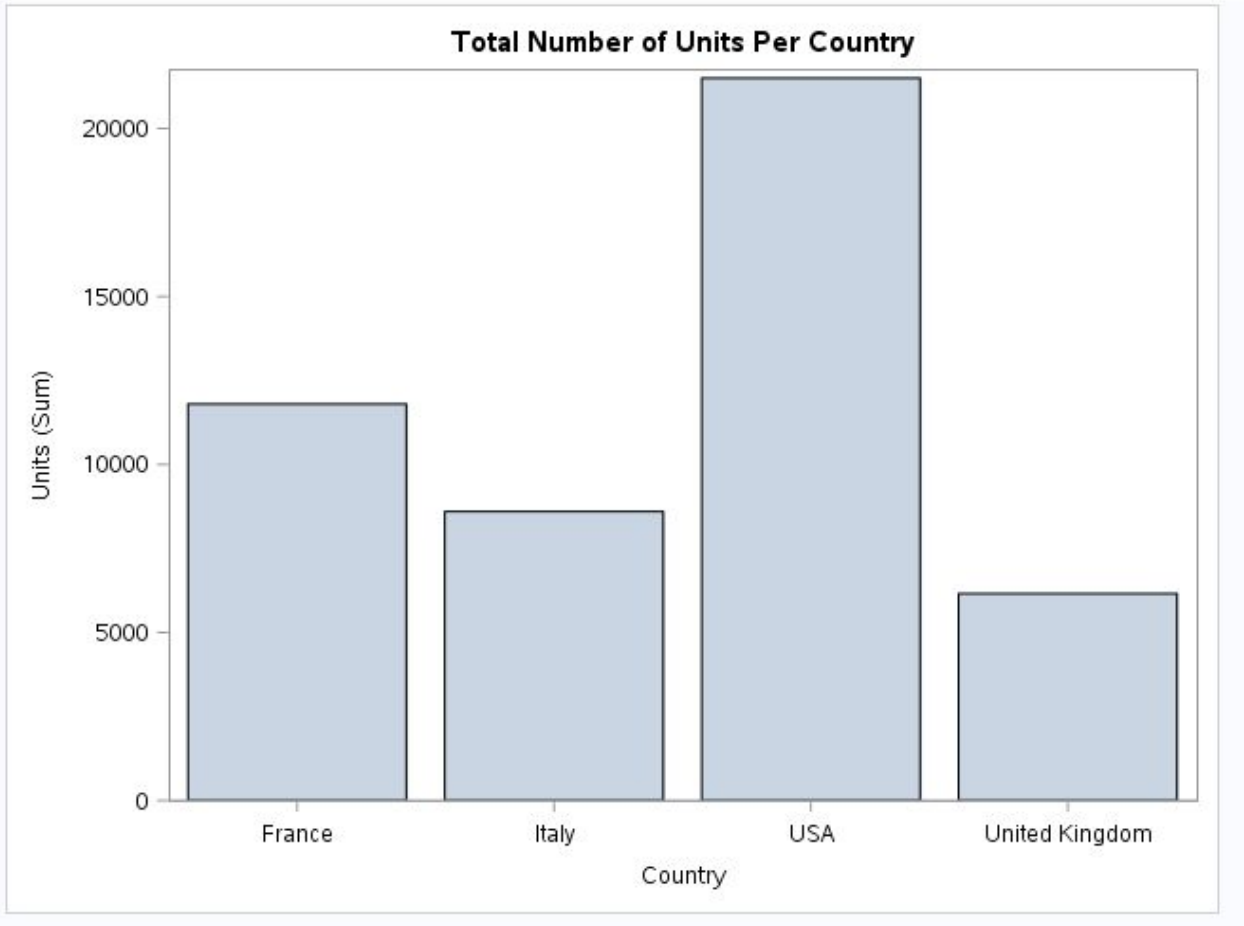
Assignments for Lesson 8 -Undergraduate

1. Write a PROC to produce a vertical bar chart and make sure you give it a title. Choose one of the homework data sets that you already have and run this procedure on that data.

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
61
62         title "Total Number of Units Per Country";
63         proc sgplot data=_temp1.bicycles;
NOTE: Data file _TEMP1.BICYCLES.DATA is in a format that is native to another host, or the file encoding does not match the session
      encoding. Cross Environment Data Access will be used, which might require additional CPU resources and might reduce
      performance.
64         vbar Country / response=Units;
65         run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.41 seconds
      cpu time            0.15 seconds

NOTE: There were 18 observations read from the data set _TEMP1.BICYCLES.

66         quit;
67
68
69
70         OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
83
```

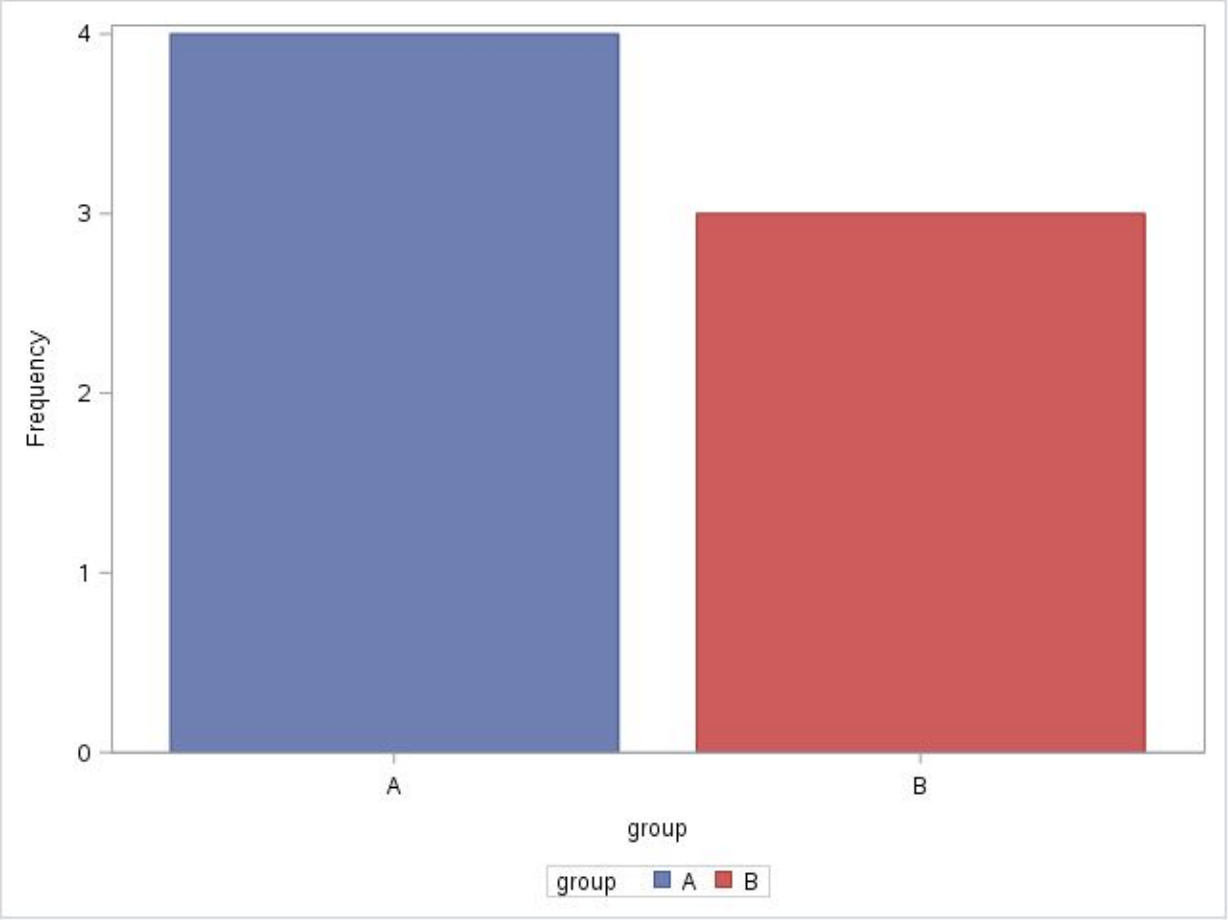


```
title "Total Number of Units Per Country";
proc sgplot data=_temp1.bicycles;
vbar Country / response=Units;
run;
quit;
```

2. Run the following program to create SAS data set called PROB2, containing variables X Y Z and Group:

```
data prob2;
 length group $ 1;
 input X Y Z Group $;
 datalines;
 2 4 6 A
 3 3 3 B
 1 3 7 A
 7 5 3 B
 1 1 5 B
 2 2 4 A
 5 5 6 A
 ;
run;
```

(a) Write SAS statements to generate a frequency bar chart (histogram) for GROUP. You may use either SAS procedures or the SAS Graphic procedures to answer questions.

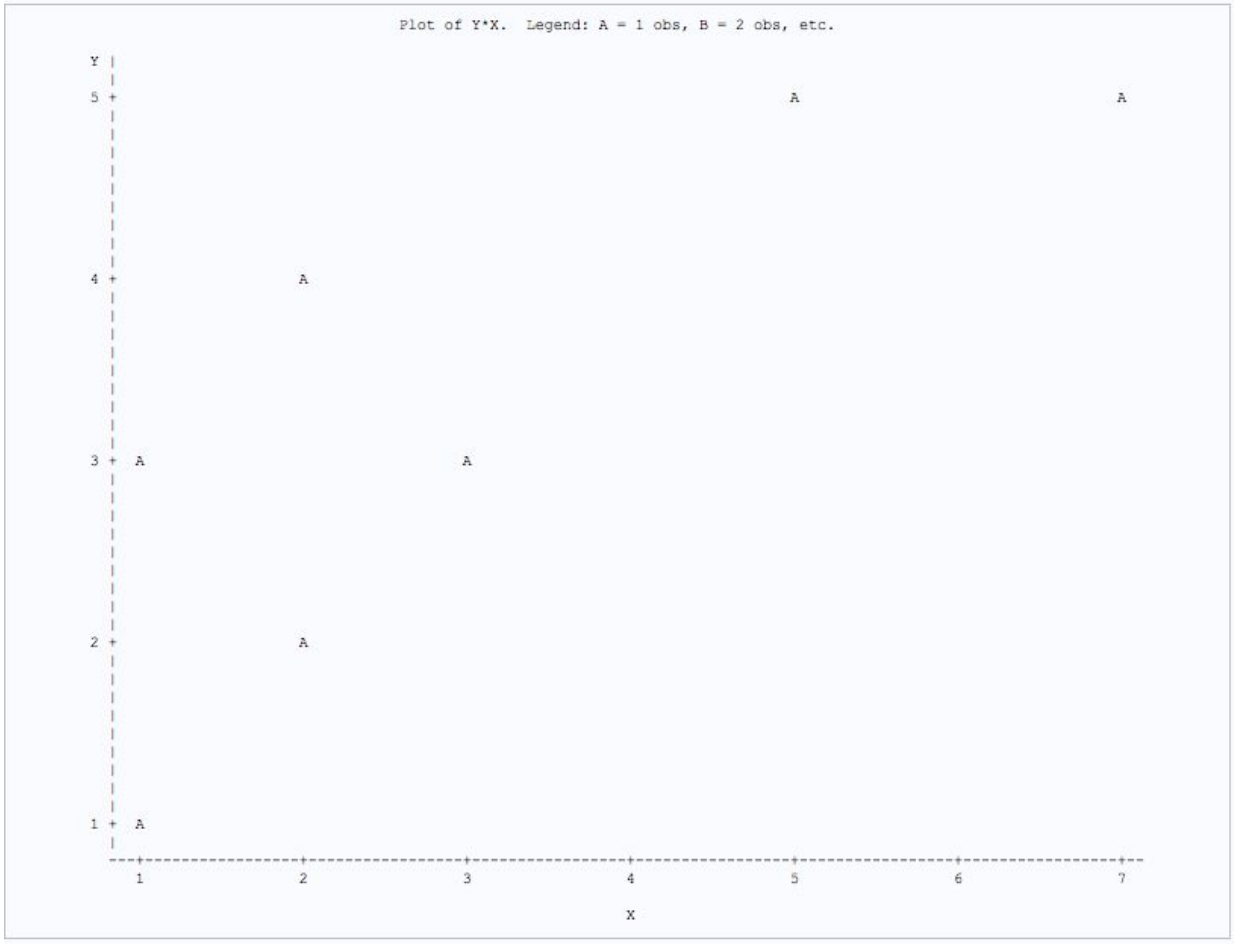

4
3
2
1
0
Frequency
A
B
group
group A B

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
61
62         data prob2;
63          length group $ 1;
64          input X Y Z Group $;
65          datalines;

NOTE: The data set WORK.PROB2 has 7 observations and 4 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      cpu time            0.00 seconds

73          ;

74         run;
75
76
77
78         proc sgplot data=prob2;
79         vbar Group / group=Group;
80         run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.47 seconds
      cpu time            0.15 seconds

NOTE: There were 7 observations read from the data set WORK.PROB2.

81         quit;
82
83         OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
96
```

```
data prob2;
 length group $ 1;
 input X Y Z Group $;
 datalines;
 2 4 6 A
 3 3 3 B
 1 3 7 A
 7 5 3 B
 1 1 5 B
 2 2 4 A
 5 5 6 A
 ;
run;
```

```
proc sgplot data=prob2;
vbar Group / group=Group;
run;
quit;
```

(b) Write the SAS statements to generate a plot of Y versus X (with “Y” on the vertical axis and “X” on the horizontal).




```
OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;

proc plot data=prob2;
plot Y * X;
run;
```


```
OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
```

```
proc plot data=prob2;
    plot Y * X;
run;
```

(c) Write SAS statements to generate a separate plot of Y versus X for each value of the GROUP variables.

group=A

group=B

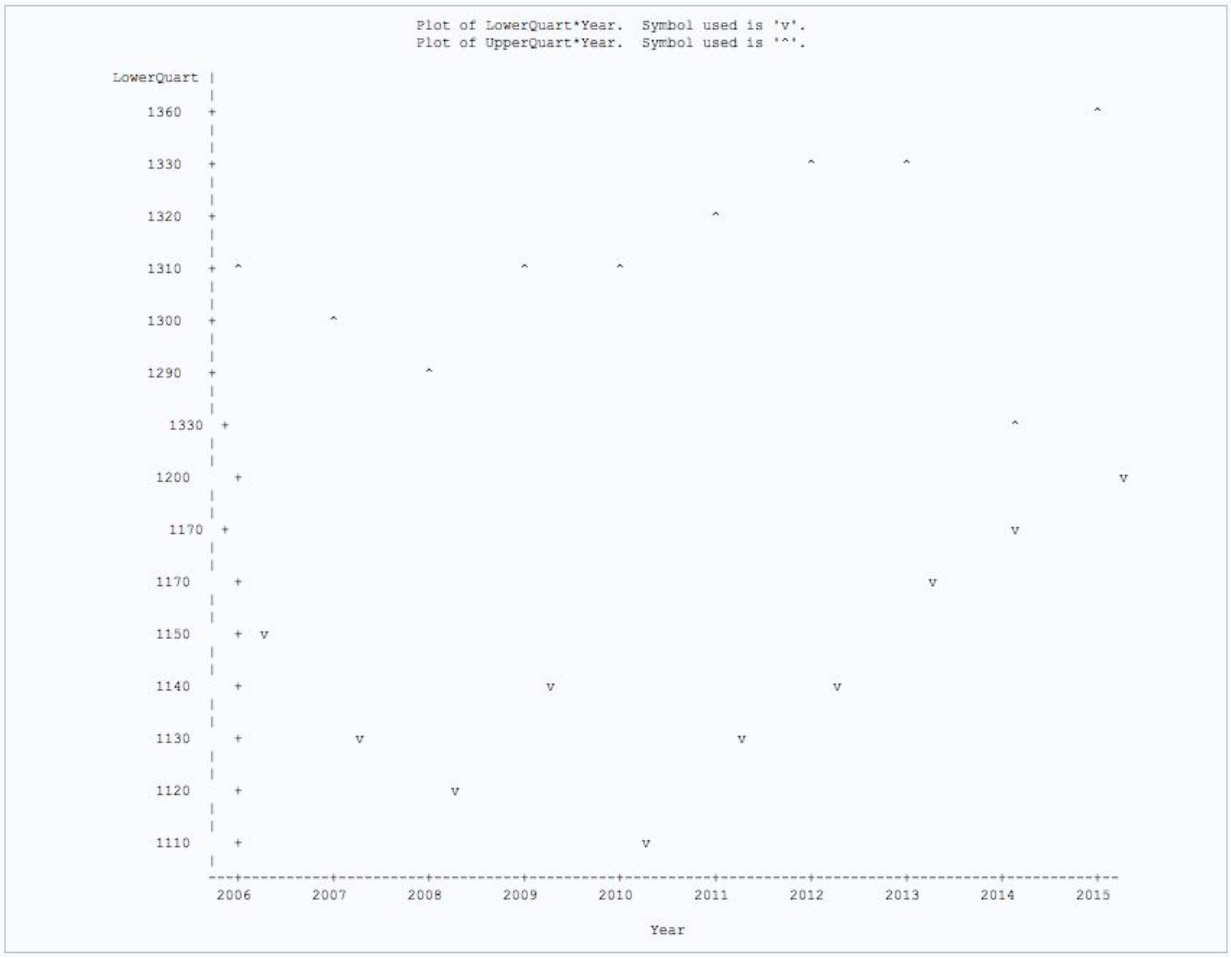
```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
61
62         data prob2;
63          length group $ 1;
64          input X Y Z Group $;
65          datalines;

NOTE: The data set WORK.PROB2 has 7 observations and 4 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds

73          ;

74         run;
75
76
77
78
79         proc sort data=prob2;
80         by Group;
81

NOTE: There were 7 observations read from the data set WORK.PROB2.
NOTE: The data set WORK.PROB2 has 7 observations and 4 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds

82         proc plot data=prob2;

83         plot Y * X;
84         by Group;
85         quit;

NOTE: There were 7 observations read from the data set WORK.PROB2.
NOTE: PROCEDURE PLOT used (Total process time):
      real time           0.07 seconds
      cpu time            0.07 seconds


86
87
88
89         OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
102
```

```
proc sort data=prob2;
by Group;

proc plot data=prob2;
    plot Y * X;
    by Group;
quit;
```

3. The Tampa Tribune recently published the following list of the 1st and 3rd quartiles of the distributions of SAT scores for freshmen admitted to the University of South Florida.

Year	Lower Quart	Upper Quart
2015	1200	1360
2014	1170	1330
2013	1170	1330
2012	1140	1330
2011	1130	1320
2010	1110	1310
2009	1140	1310
2008	1120	1290
2007	1130	1300
2006	1150	1310

Prepare a scatterplot with the lower quartile on the vertical axis versus the year on the horizontal axis. Mark these observations on the plot with the character v. On the same plot, show the upper quartile plotted against the year. Mark these points with a ^. (You may want to use this idea to show confidence intervals or prediction intervals on a plot.)


Plot of LowerQuart*Year. Symbol used is 'v'.
Plot of UpperQuart*Year. Symbol used is '^'.
LowerQuart
1360
1330
1320
1310
1300
1290
1330
1200
1170
1170
1150
1140
1130
1120
1110
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
Year

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
61
62
63         proc plot data=satscores;
64         plot LowerQuart * Year='v' UpperQuart * Year='^' / overlay;
65         run;

66         quit;

NOTE: There were 10 observations read from the data set WORK.SATSCORES.
NOTE: PROCEDURE PLOT used (Total process time):
      real time           0.03 seconds
      cpu time            0.04 seconds


67
68
69         OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
82
```

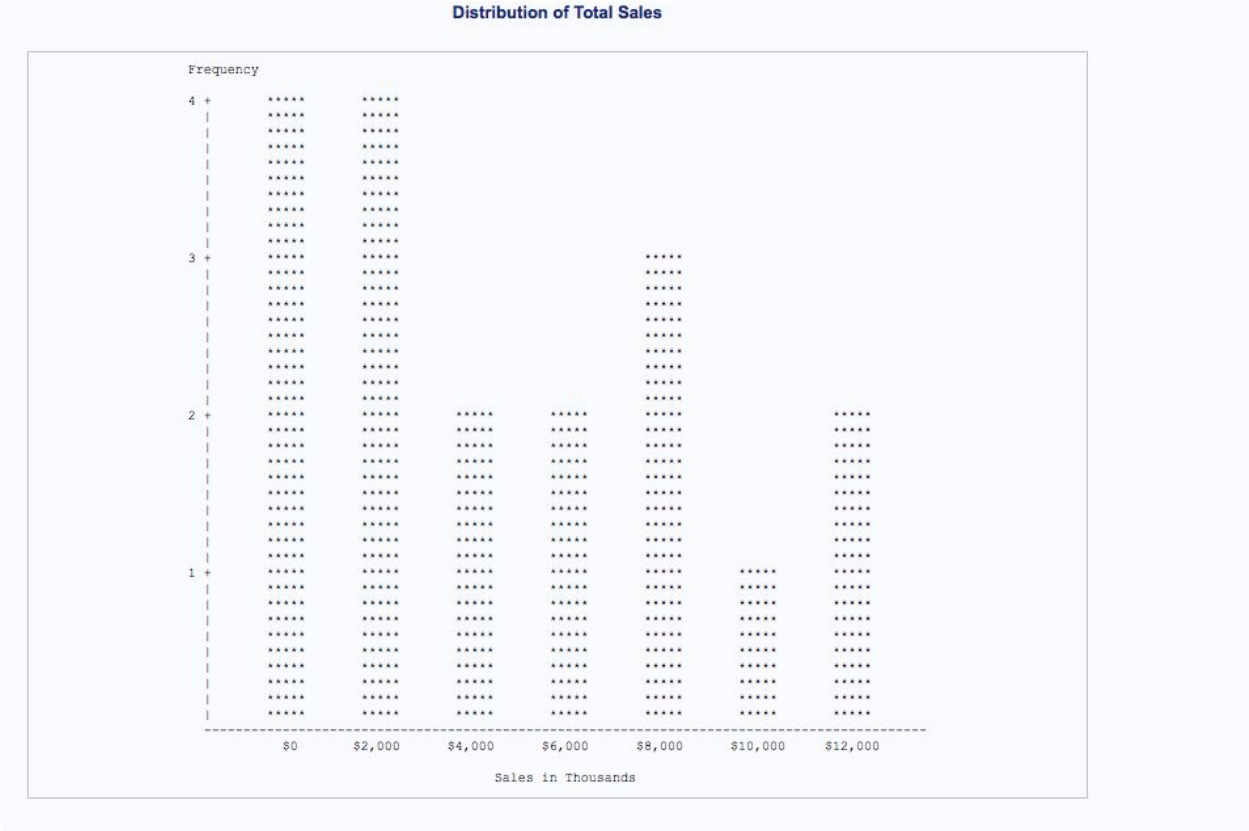
```
data satscores;
input @1 Year $4. @5 LowerQuart $6.  @11 UpperQuart $5.;
datalines;
2015    1200    1360
2014    1170    1330
2013    1170    1330
2012    1140    1330
2011    1130    1320
2010    1110    1310
2009    1140    1310
2008    1120    1290
2007    1130    1300
2006    1150    1310
;
run;
```

```
proc plot data=satscores;
    plot LowerQuart * Year='v' UpperQuart * Year='^' / overlay;
run;
quit;
```

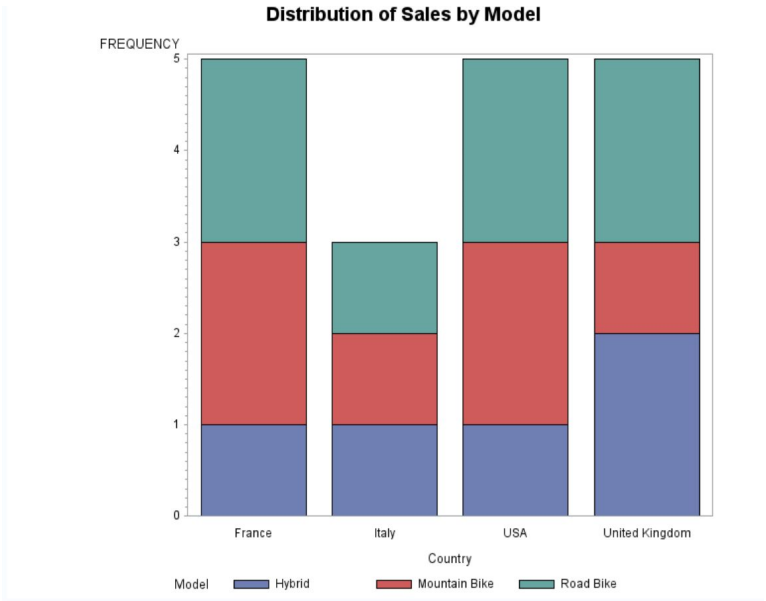
4. Using the SAS dataset BICYCLES, produce a vertical bar chart showing the distribution of Total Sales (TotalSales). Use midpoints of 0 to 12,000, with intervals of 2,000.

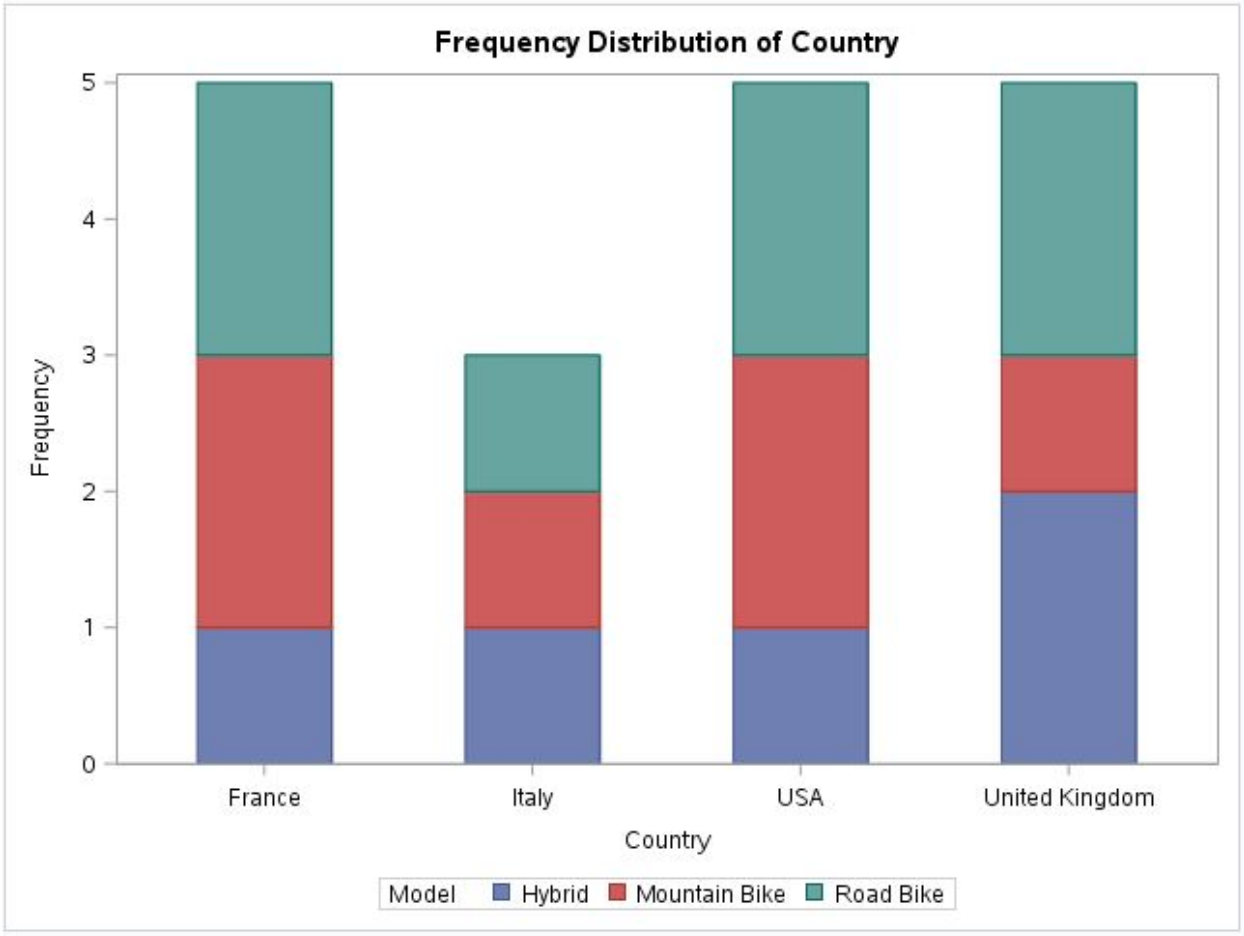
```
title 'Distribution of Total Sales';
proc chart data=_temp0.bicycles;
vbar TotalSales / midpoints=0 to 12000 by 2000 ;
run;
```

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
61
62         title 'Distribution of Total Sales';
63         proc chart data=_temp0.bicycles;
NOTE: Data file _TEMP0.BICYCLES.DATA is in a format that is native to another host, or the file encoding does not match the session
      encoding. Cross Environment Data Access will be used, which might require additional CPU resources and might reduce
      performance.
64         vbar TotalSales / midpoints=0 to 12000 by 2000 ;
65         run;

NOTE: PROCEDURE CHART used (Total process time):
      real time           0.04 seconds
      cpu time            0.04 seconds


66
67
68
69
70         OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
83
```



5. Using the SAS data set BICYCLES, produce a bar chart showing a frequency distribution of Country. Within each bar, show the distribution of Model. Your chart should look like this:






```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
61
62
63         proc sgplot data=_temp0.bicycles;
NOTE: Data file _TEMP0.BICYCLES.DATA is in a format that is native to another host, or the file encoding does not match the session
      encoding. Cross Environment Data Access will be used, which might require additional CPU resources and might reduce
      performance.
64         vbar Country / barwidth=0.5 GROUP=Model;
65         title 'Frequency Distribution of Country';
66         run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.63 seconds
      cpu time            0.17 seconds

NOTE: There were 18 observations read from the data set _TEMP0.BICYCLES.

67         quit;
68
69
70
71         OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
84
```

```
proc sgplot data=_temp0.bicycles;
    vbar Country / barwidth=0.5 GROUP=Model;
    title 'Frequency Distribution of Country';
run;
quit;
```