Assignments for Lesson 8

1. Refer to MANATEES data. Plot the total number of manatee deaths in each year on the vertical axis and the year on the horizontal axis. Mark these points on the plot with a T. On the same plot, show the total number of deaths attributed to humans (from all watercraft [WATERC], flood gates [PERIN], canal locks [LOCK], and other human-related deaths [RELAT]) plotted against the year. Mark the human-related fatalities with the character H.

**CODE**

**data** manatees;

infile '\\Client\C$\Users\Ultron\Desktop\ISM6930 SAS\Data\manatees.txt' dlm='09'x firstobs = **2**;

input Year Waterc Lock Relat Perin Natur Undet;

Label Waterc = 'Watercraft'

Perin = 'Flood Gates'

Lock = 'Canal Locks'

Relat = 'Human Related';

Total= Waterc + Lock + Relat + Perin + Natur + Undet;

Human= Waterc + Perin + Lock + Relat;

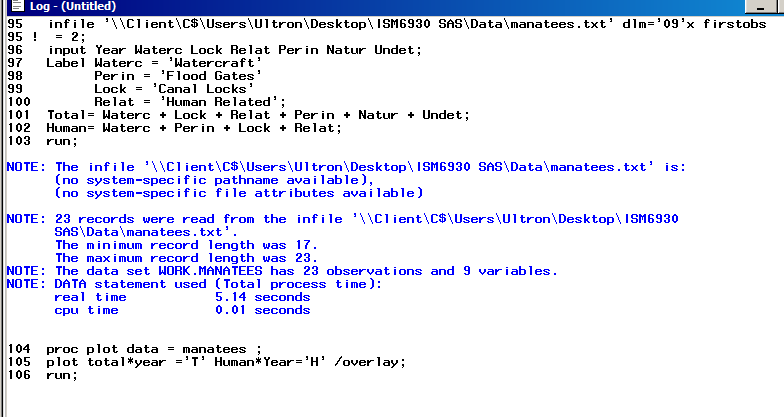
**run**;

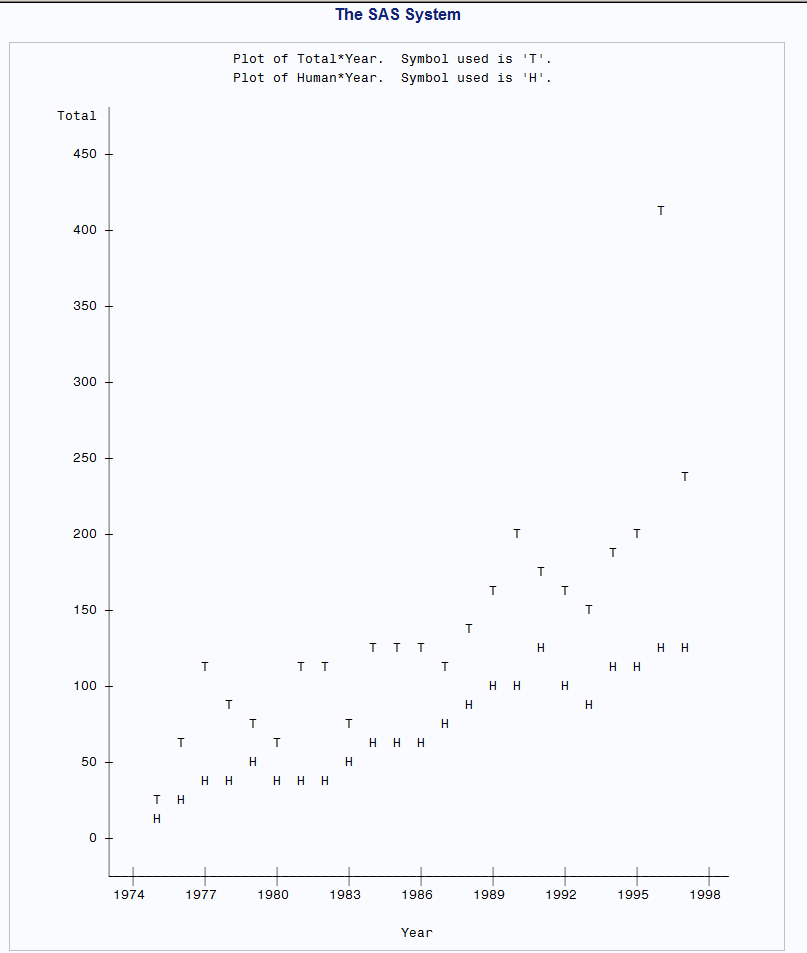
**proc** **plot** data = manatees ;

plot total\*year ='T' Human\*Year='H' /overlay;

**run**;

**quit**; **LOG**

 **OUTPUT**



1. 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 Upper

Quart 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.)

**CODE-**

**data** tribune;

input Year LowerQuart UpperQuart;

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= tribune;

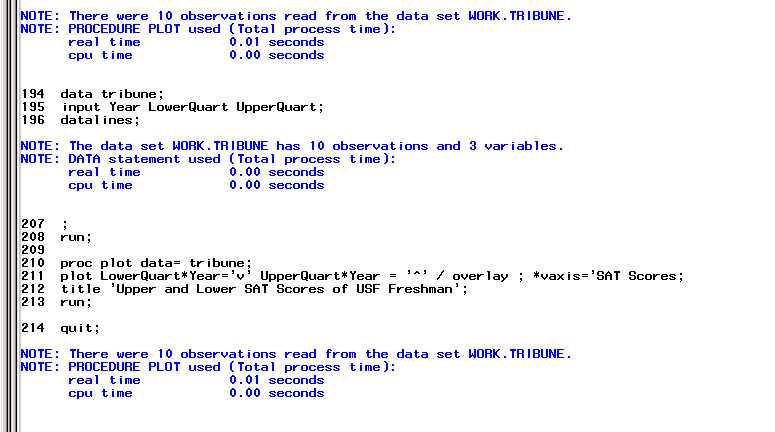
plot LowerQuart\*Year='v' UpperQuart\*Year = '^' / overlay ; \*vaxis='SAT Scores;

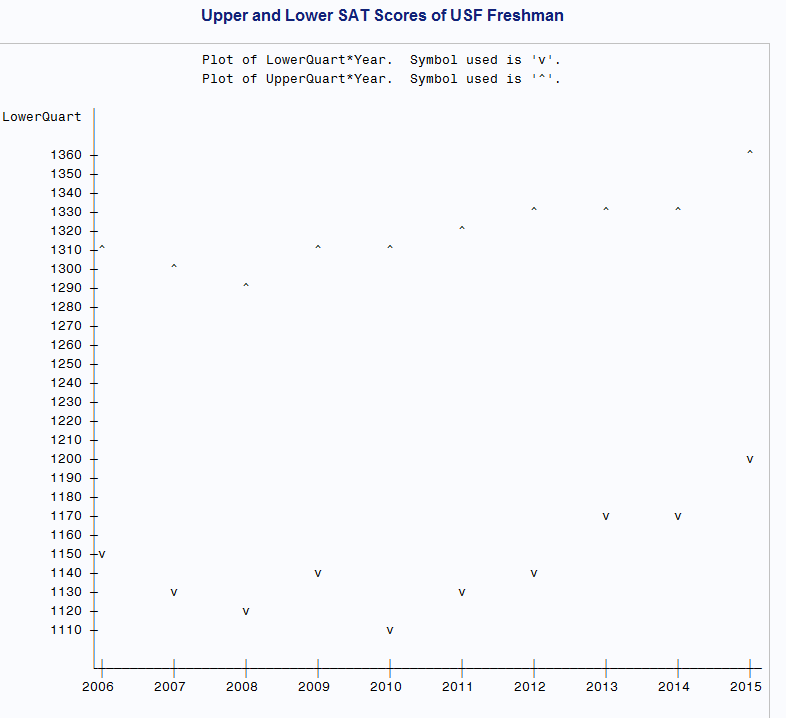
title 'Upper and Lower SAT Scores of USF Freshman';

**run**;

**quit**;

**LOG**

 **OUTPUT**



1. Refer to the IRIS data. Create labels for the variables in the dataset. Prepare three scatterplots (one for each iris species) to illustrate the relationship between sepal length (SL) and sepal width (SW) within each species. Use the UNIFORM option so that all 3 plots have the same dimensions for the X- and Y-axes. Add an appropriate descriptive title to the plots.

**CODE**

**data** iris;

infile '\\Client\C$\Users\Ultron\Desktop\ISM6930 SAS\Lesson 4\Homework\iris.txt' firstobs=**2**;

input class $ SL SW PL PW;

label SL = 'Sepal Length'

SW = 'Sepal Width'

PL = 'Pupil Length'

PW = 'Pupil Width';

**run**;

\*descending goes before the var;

**proc** **sort** data=iris;

by class;

**run**;

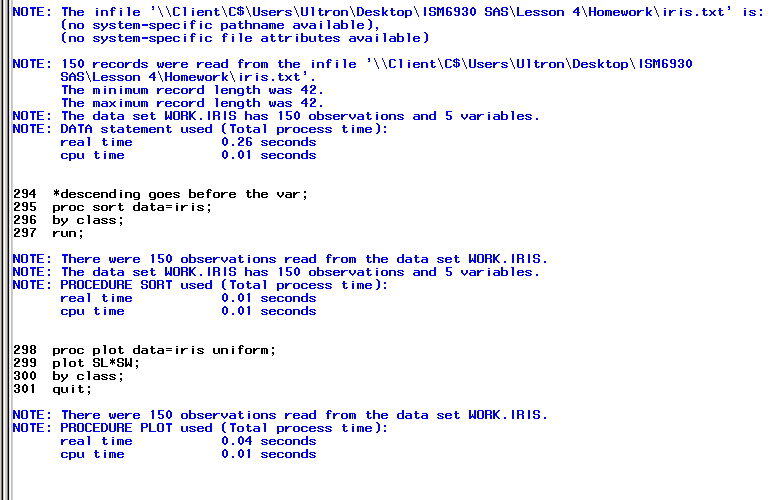
**proc** **plot** data=iris uniform;

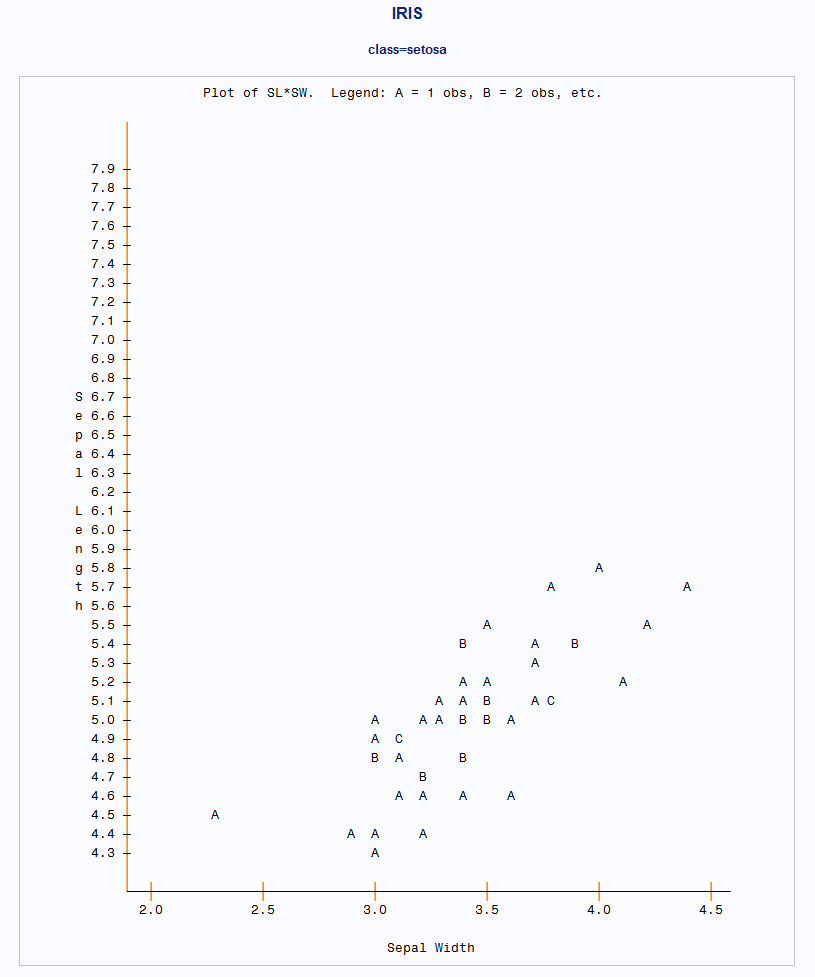
plot SL\*SW;

by class;

**quit**;

**LOG**

 **OUTPUT**







1. Refer to the CLINTON data. Calculate the average percent approval rating for President Clinton in each year from 1992-1998, and plot the average approval rating on the vertical axis versus the year on the horizontal axis.

Question 4 – having trouble with the horizontal axis spacing by one year. When I add an haxis ref the year starts in 1965. Without it there are multiple instances for the year.

**data** clinton;

infile '\\Client\C$\Users\Leland\Desktop\ISM6930 SAS\Lesson 5\Homework\Clinton.txt' firstobs=**3**;

input Date date18. Approval Disapproval No\_Opinion;

format date year4.;

**run**;

**proc** **sort** data=clinton;

by date;

**run**;

**data** app\_per;

set clinton;

app\_percent = Approval / ( Approval + Disapproval + No\_opinion);

format app\_percent percent10.;

if date=**.** then delete;

**run**;

**proc** **means** data= app\_per ;

var app\_percent;

class date;

output out= clint\_mean

mean= rating;

**run**;

**proc** **print** data=clint\_mean;

title 'Clinton Mean';

**run**;

**proc** **gplot** data=clint\_mean;

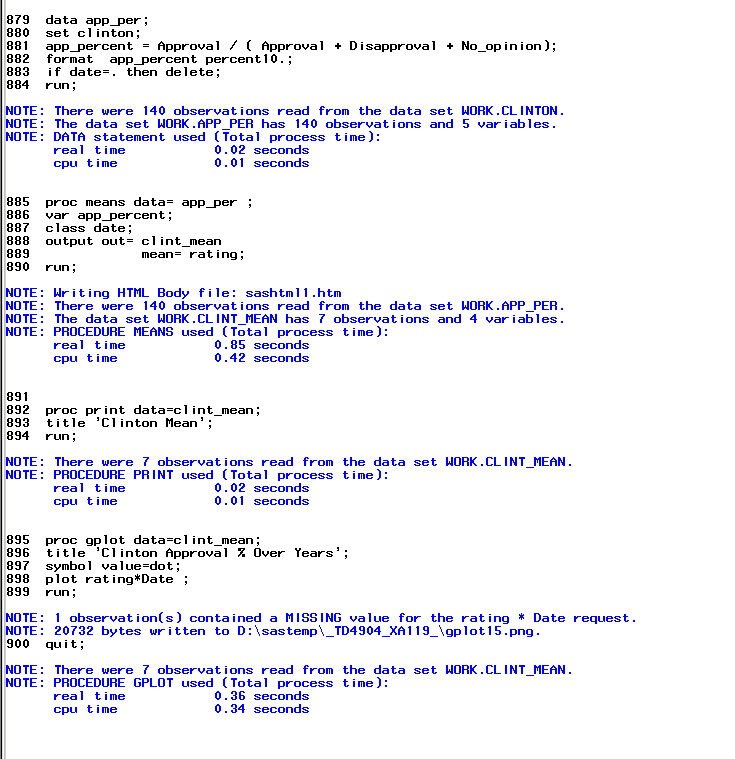
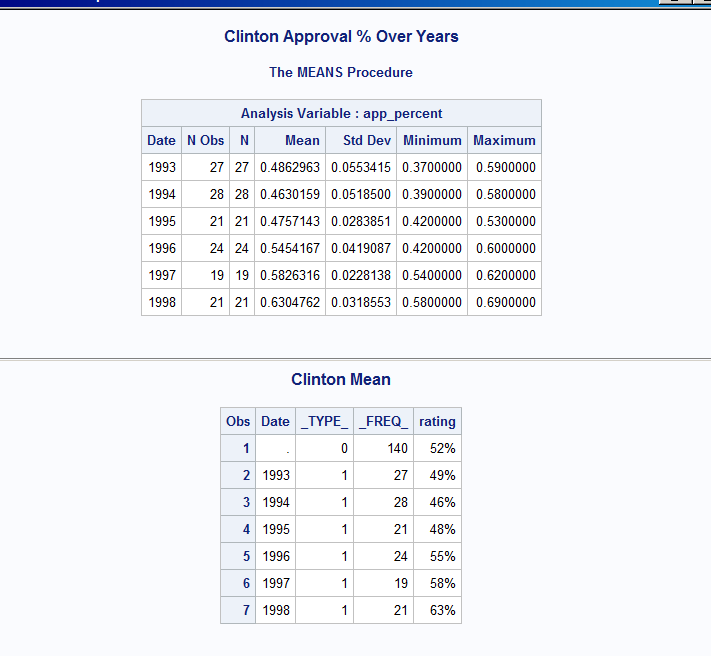
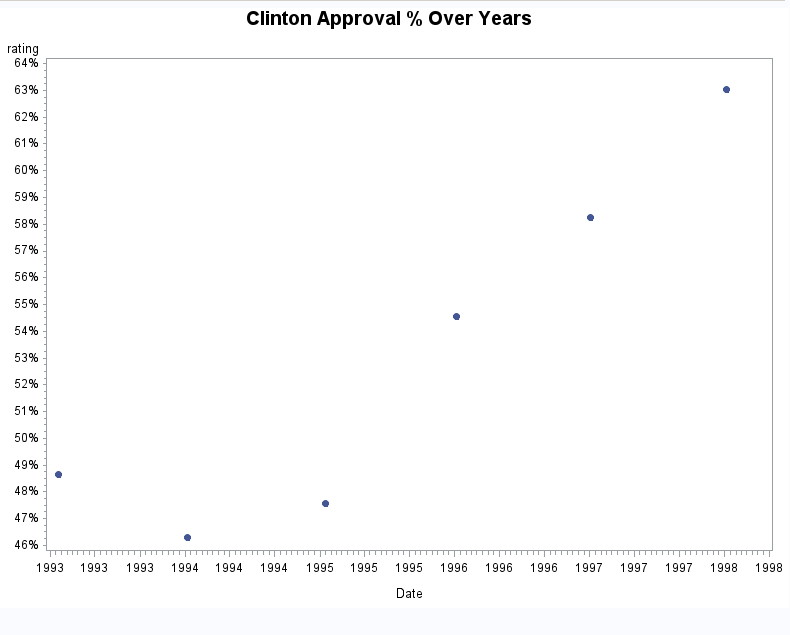
title 'Clinton Approval % Over Years';

symbol value=dot;

plot rating\*Date ;

**run**;

**quit**;

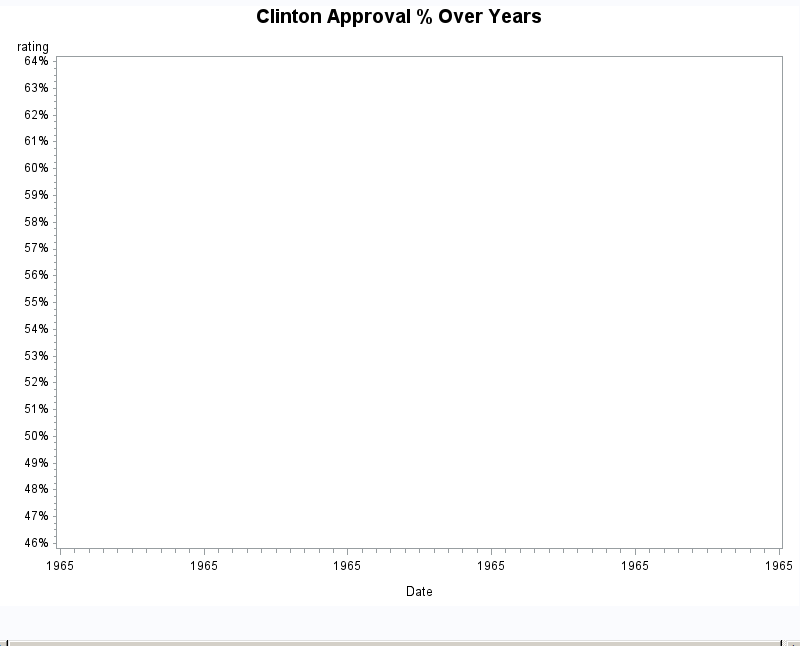
   **proc** **gplot** data=clint\_mean;

title 'Clinton Approval % Over Years';

symbol value=dot;

plot rating\*Date / haxis=**1993** to **1998** by **1**;

**run**;

**quit**; 

1. Refer to the CATS (MERGE CATS1 CATS2 and CATS3) data. Plot the GFR of the treated kidney at Week 1 on the vertical axis versus the GFR of the untreated kidney at Week 1 on the horizontal axis. Label the points on the plot with the first letter of each cat’s name. Then, on the same graph, plot the GFR of the untreated kidney at Week 1 versus the GFR of the untreated kidney at week 1 and label the points with +. Then, on the same graph, plot the GFR of the treated kidney at Week 1 versus the GFR of the treated kidney at week 1 and label the points with +.

This scatterplot shows a line (+) that would be followed by the data if the surgically-treated kidney and the control kidney were identical after 1 week. The letters indicate how individual cats responded to the treatment.

**CODE**

libname learn '\\Client\C$\Users\Ultron\Desktop\ISM6930 SAS\Data';

**data** cats1;

infile '\\Client\C$\Users\Ultron\Desktop\ISM6930 SAS\cats1.txt';

input Cat $ Dir $ @@;

**run**;

**data** cats2;

infile '\\Client\C$\Users\Ultron\Desktop\ISM6930 SAS\cats2.txt' firstobs=**2** ;

input Cat $ @**9** Side $ Week0 Week1 Week2;

**run**;

**data** cats3;

infile '\\Client\C$\Users\Ultron\Desktop\ISM6930 SAS\cats3.txt' firstobs = **2**;

input Cat $ Side $ Week0 Week1 Week2;

**run**;

**proc** **sort** data=cats2;

by Cat;

**proc** **sort** data=cats3;

by Cat;

**data** cats23;

set cats2 cats3;

by cat;

drop week0 week2;\*edited week1 to week0;

**run**;

**proc** **sort** data=cats1;\*(rename(dir=side));

by cat;

**run**;

**data** all\_cats;

merge cats1 cats23;

by cat;

format treated **4.2** Untreated **4.2**;

if dir=Side then treated=Week1; \*changed from week0;

else if dir^=side then untreated=Week1;\*changed from week0;

drop dir side week1;\*changed from week0;

**run**;

**data** treated untreated;

set all\_cats;

by cat;

if treated ^=**.** then output treated;

else if untreated^=**.** then output untreated;

**run**;

**data** all\_done;

merge treated(drop=untreated) untreated(drop=treated);

by cat;

**run**;

**proc** **plot** data=all\_done uniform;

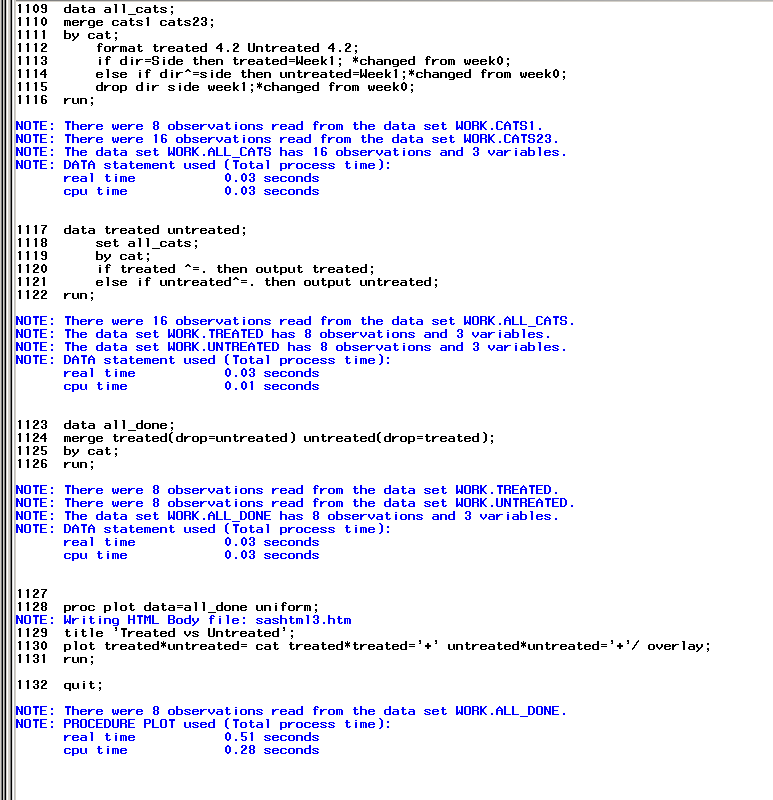
title 'Treated vs Untreated';

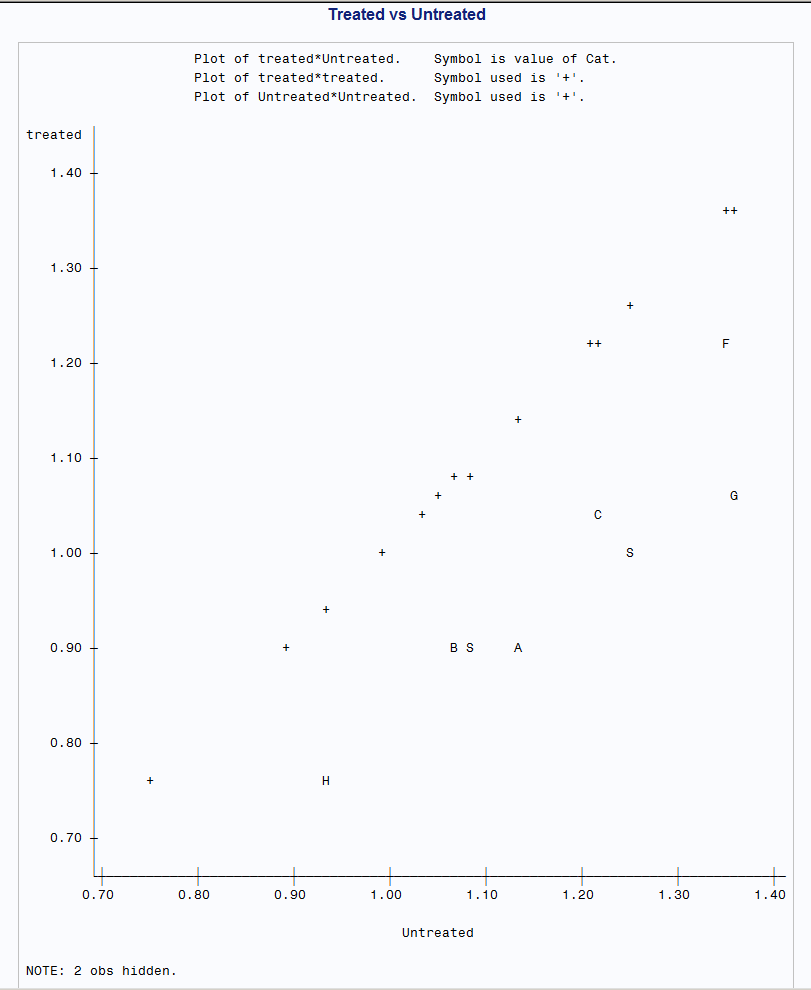
plot treated\*untreated= cat treated\*treated='+' untreated\*untreated='+'/ overlay;

**run**;

**quit**;

**LOG**

 **OUTPUT**



1. Refer to the AIRPORTS data. Create a SAS dataset. Within the DATA step, create a new variable called PLOTCHAR with this command:

PLOTCHAR=BYTE(\_N\_+32);

Print the resulting dataset. Now, plot the 1995 passenger total versus the 1985 passenger total and label the points in the plot with the first letter of the city. On a separate graph, plot the 1995 passenger total versus the 1985 passenger total and label the points in the plot with PLOTCHAR. What advantages does the second plot have over the first plot?

The advantage would be not having any duplicate characters points for to identify which point belongs to which city.

**CODE**

**data** airport;

infile '\\Client\C$\Users\Leland\Desktop\ISM6930 SAS\Lesson 3\Homework\airport.csv' dlm = ',' firstobs=**2** dsd ;

input city :$30. state $ abrev $ pass95 pass85;

PLOTCHAR=BYTE(\_N\_+**32**);

**run**;

**proc** **print** data=airport;

title 'airport';

**run**;

**proc** **plot** data=airport;

plot pass95\*pass85= city;

**run**;

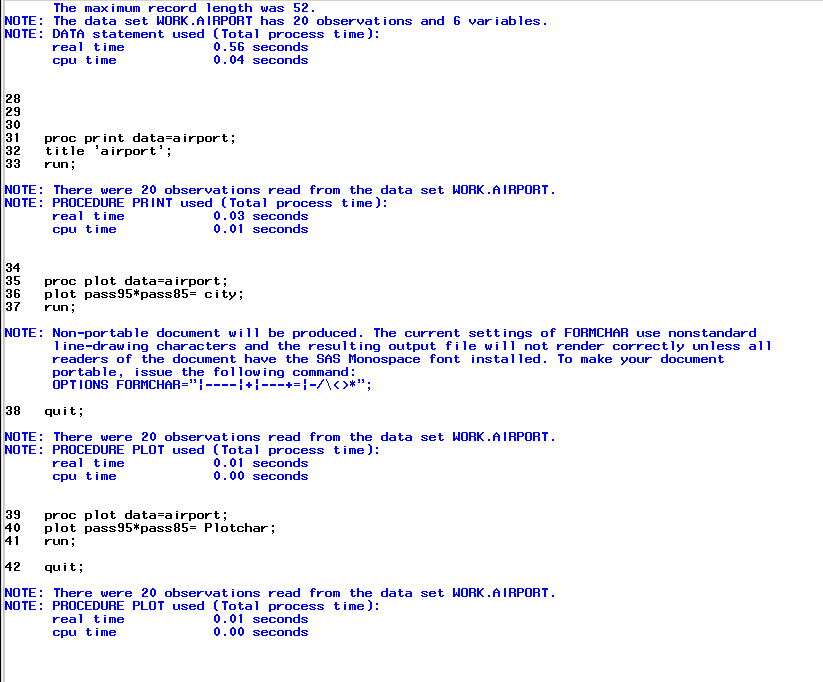
**quit**;

**proc** **plot** data=airport;

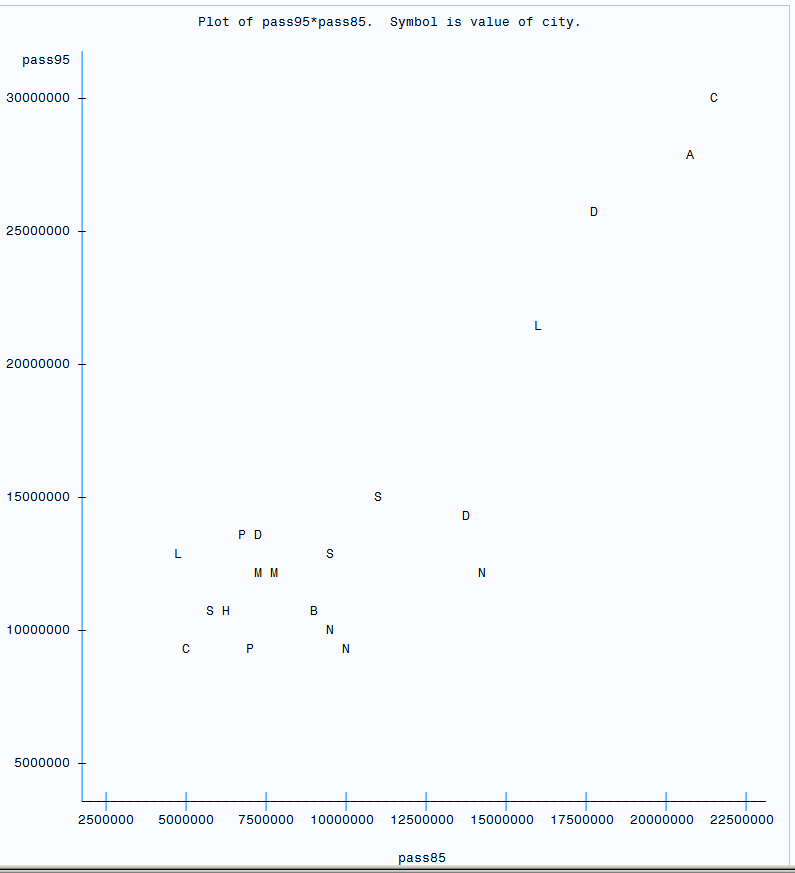
plot pass95\*pass85= Plotchar;

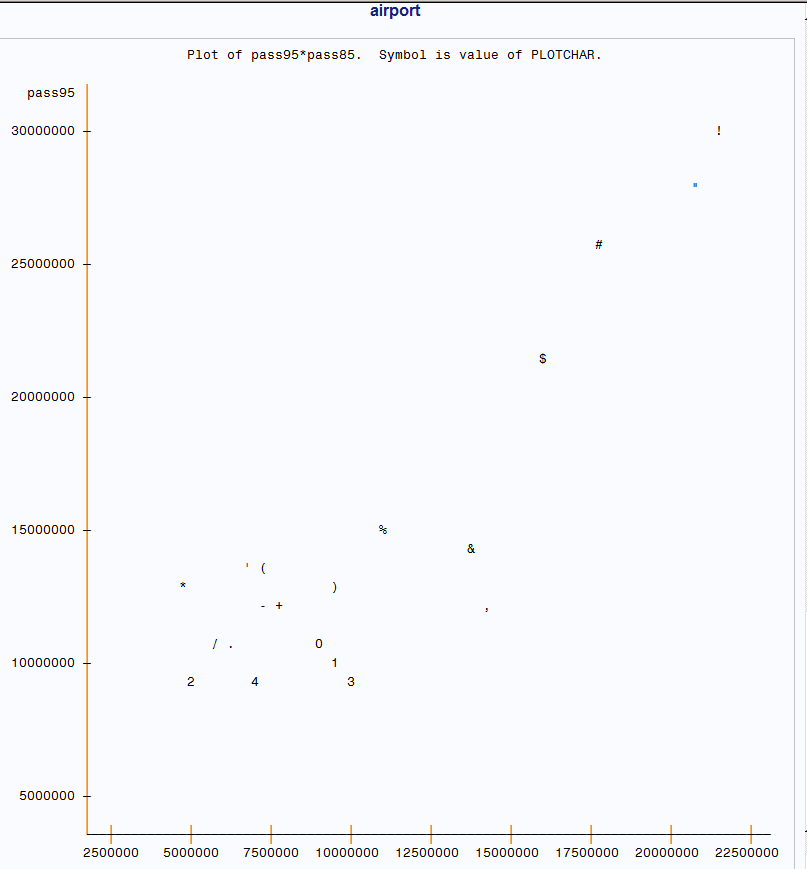
**run**;

**quit**; **LOG**

 **OUTPUT**







1. Refer to the DOGS1 dataset. To fairly assess the effects of the three shampoo treatments, we want to see if roughly equal numbers of male and female dogs received each treatment. (For example, if only males have severe allergic reactions to the drug, we want enough males in the study to be able to see the adverse reaction occur.) Prepare a side-by-side vertical bar chart which shows the frequencies of males and females within each treatment group. Your chart should have 6 bars (2 genders x 3 treatments = 6 bars).

**CODE**

**data** dogs1;

infile '\\Client\C$\Users\Leland\Desktop\ISM6930 SAS\data\dogs1.txt' firstobs = **2**;

input Dog $ Conc **16** Sex $ **17** Age **31**-**32** Haircoat $ @**32** Weight **43**-**48**;

**run**;

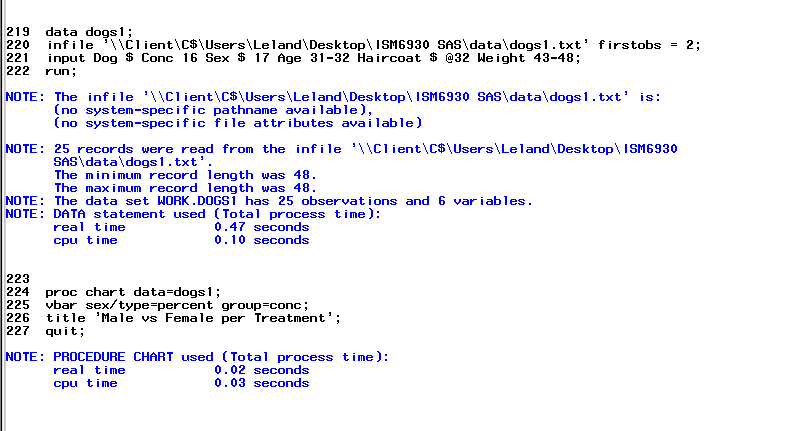
**proc** **chart** data=dogs1;

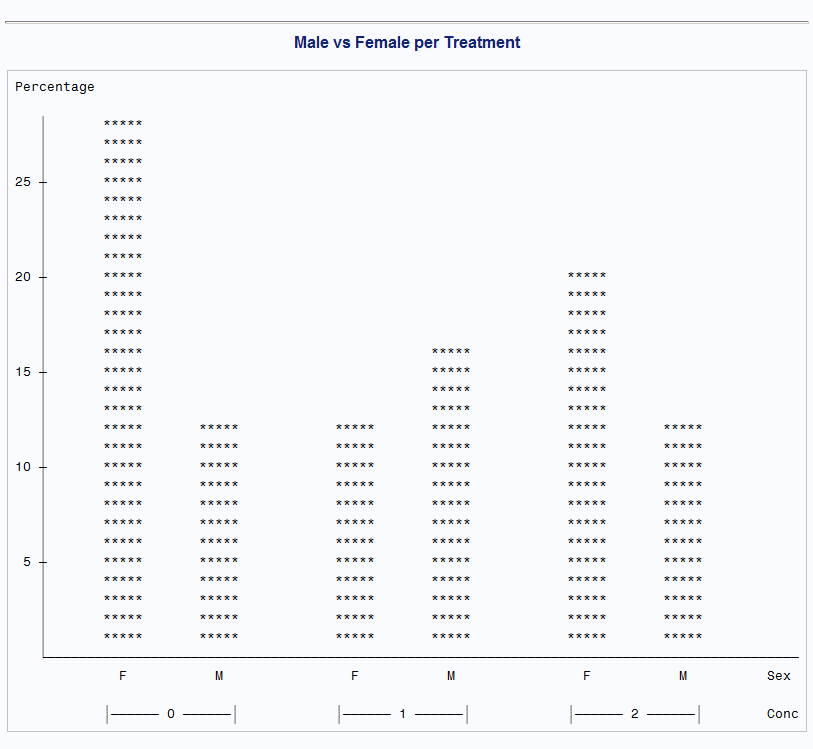
vbar sex/type=percent group=conc;

title 'Male vs Female per Treatment';

**quit**;

**LOG**

 **OUTPUT**



1. Refer to the LIMES dataset. Fruits and vegetables are sometimes classified into groups by size for sale on the market. Suppose that “extra-fancy” limes are those with diameters of at least 6 centimeters and lengths of at least 7 centimeters. Plot the fruit diameters versus the fruit lengths, and insert horizontal and vertical reference lines to indicate the 6-cm mark for diameters and the 7-cm mark for lengths. Use the plot to estimate the number of “extra-fancy” limes in the sample.

Approximately 13 fancy limes

**CODE**

**data** limes;

infile '\\Client\C$\Users\Leland\Desktop\ISM6930 SAS\Lesson 3\Homework\limes.txt'

dlm=',' firstobs=**2** dsd;

input Date :MMDDYY10. FruitDia FruitLength FruitWt FruitVol JuiceVol JuiceWt PeelWt ;

format date MMDDYY8.;

**run**;

**proc** **plot** data= limes;

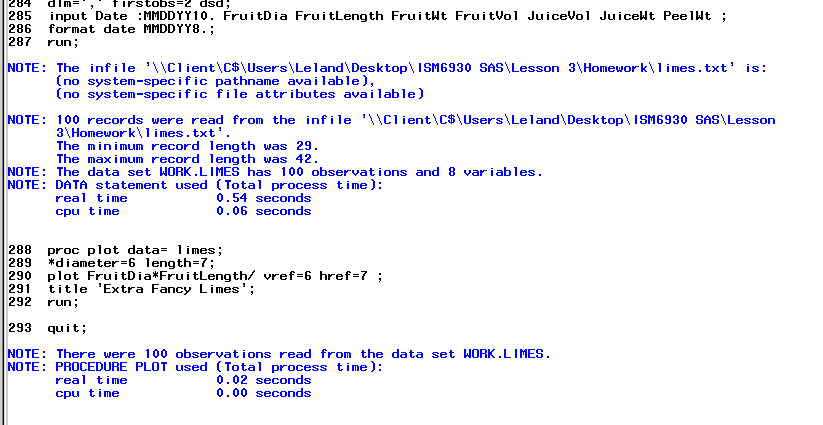
\*diameter=6 length=7;

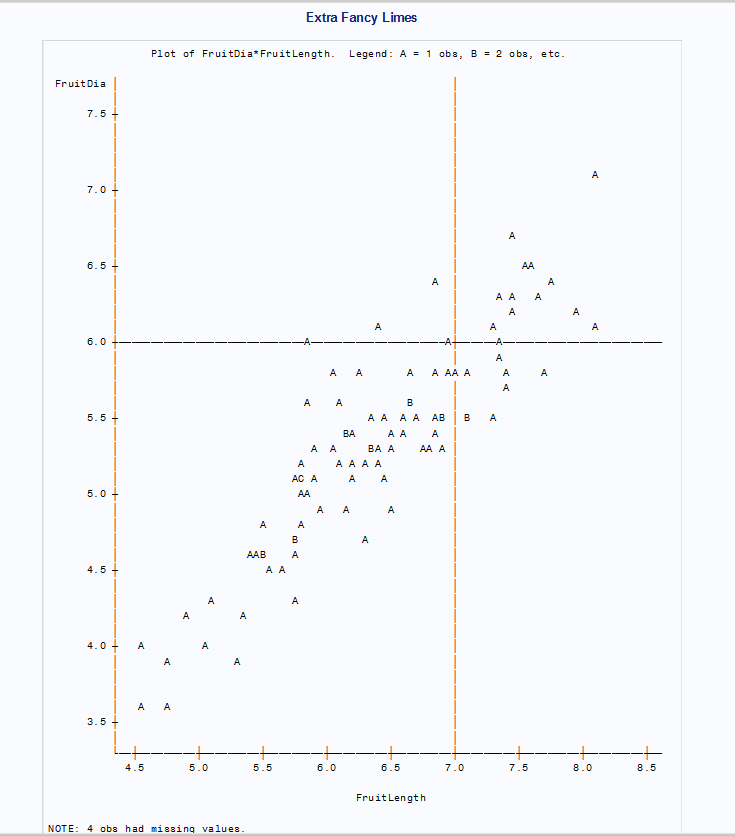
plot FruitDia\*FruitLength/ vref=**6** href=**7** ;

title 'Extra Fancy Limes';

**run**;

**quit**; **LOG**

 **OUTPUT**



1. 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.

**CODE**

libname learn v9'\\Client\C$\Users\Leland\Desktop\ISM6930 SAS\Lesson 5\Homework';

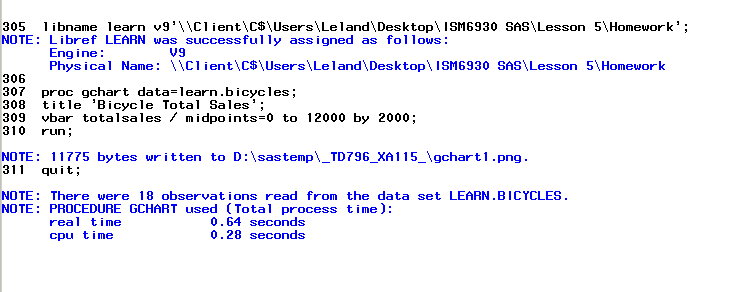
**proc** **gchart** data=learn.bicycles;

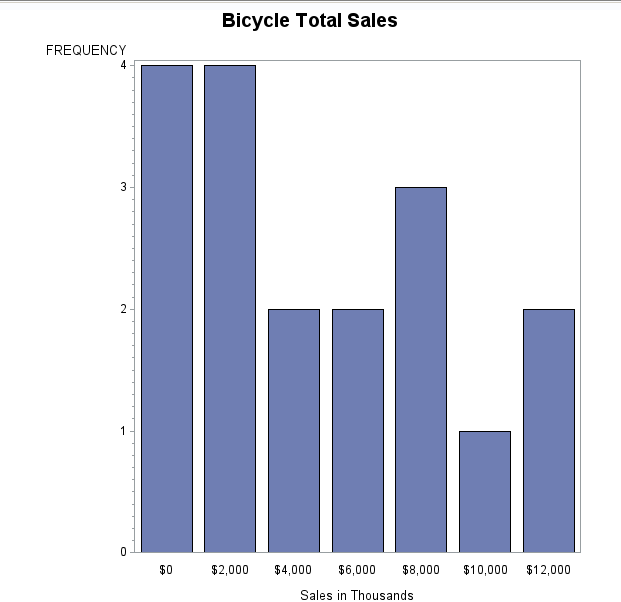
title 'Bicycle Total Sales';

vbar totalsales / midpoints=**0** to **12000** by **2000**;

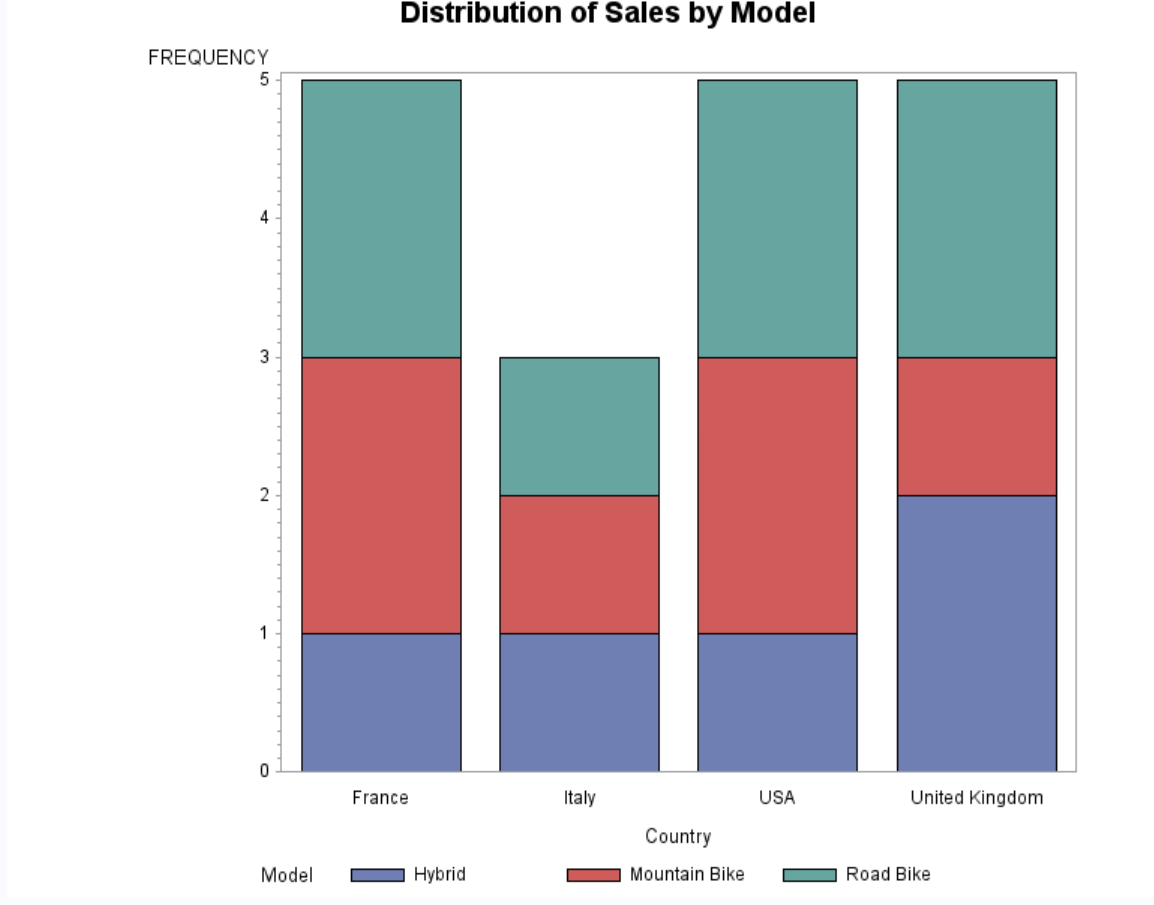
**run**;

**quit**; **LOG**

 **OUTPUT**



1. 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:



**CODE**

libname learn v9'\\Client\C$\Users\Leland\Desktop\ISM6930 SAS\Lesson 5\Homework';

**proc** **print** data=learn.bicycles;

**run**;

**proc** **gchart** data=learn.bicycles;

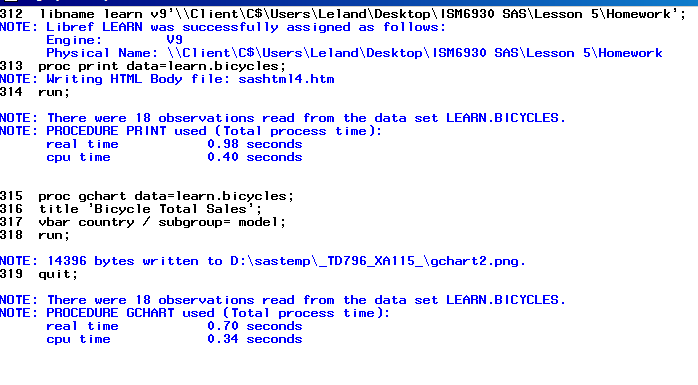
title 'Bicycle Total Sales';

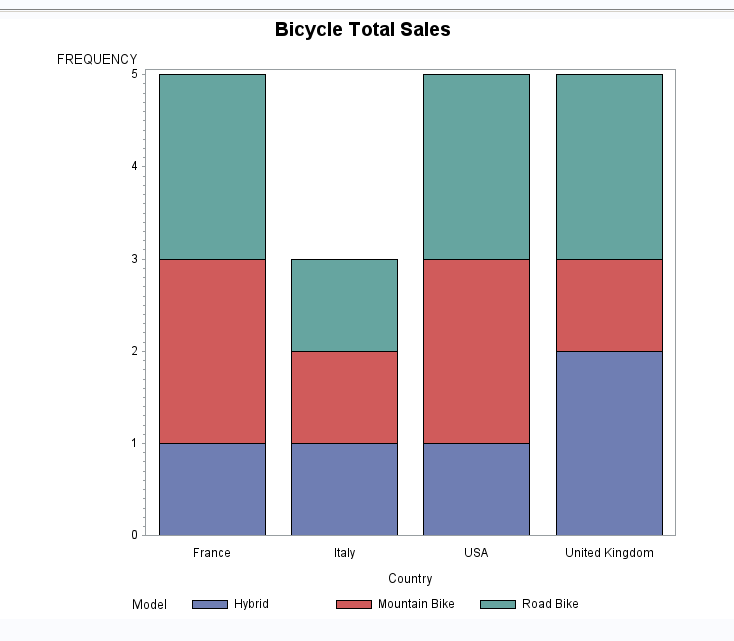
vbar country / subgroup= model;

**run**;

**quit**;

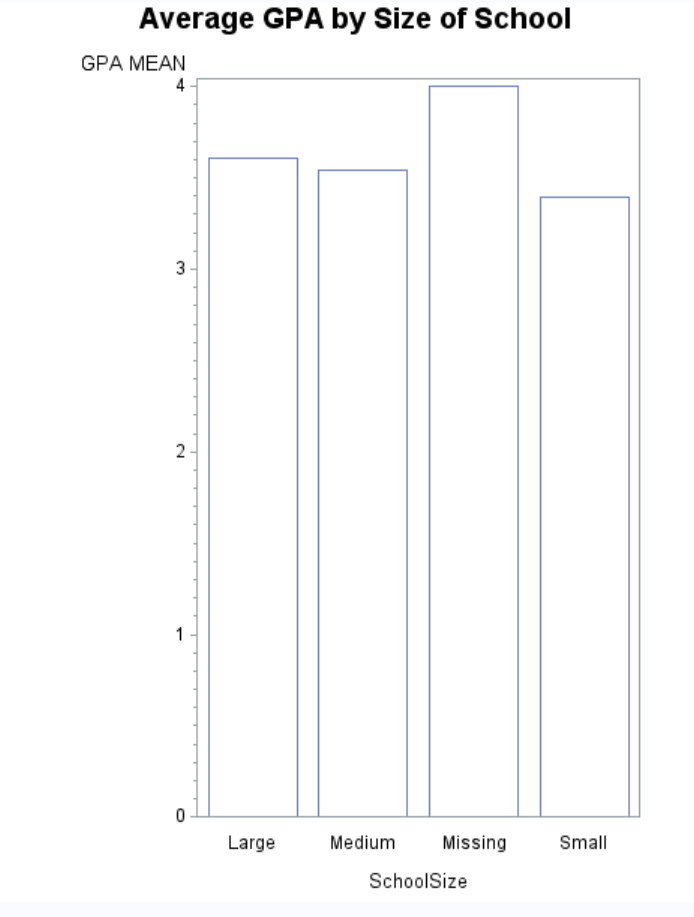
**LOG**

 **OUTPUT**



1. Using the SAS data set College, produce a vertical bar chart where the mean GPA shown for each value of school size (SchoolSize). Remember to include a FMTSEARCH option, use the system option NOFMTERR, or write a format of your own. Your chart should look like this:

(Do NOT forget you need the formats file to be placed in the same folder as the data so you can map to it)



**CODE**

option nofmterr;

libname HW '\\Client\C$\Users\Leland\Desktop\ISM6930 SAS\Lesson 7';

options fmtsearch=(HW);

title "Statistics on the college data set";

**proc** **print** data=hw.college;

**run**;

**proc** **gchart** data= hw.college;

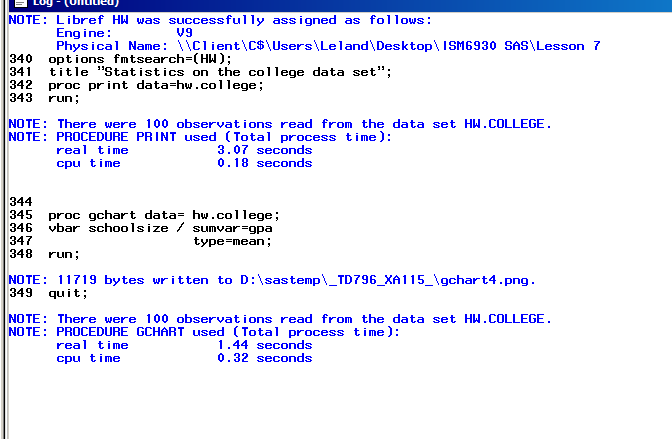
vbar schoolsize / sumvar=gpa

type=mean;

**run**;

**quit**;

**LOG**



**OUTPUT**

