***“The work contained and presented here is my work and my work alone.” – Haardik Sharma***

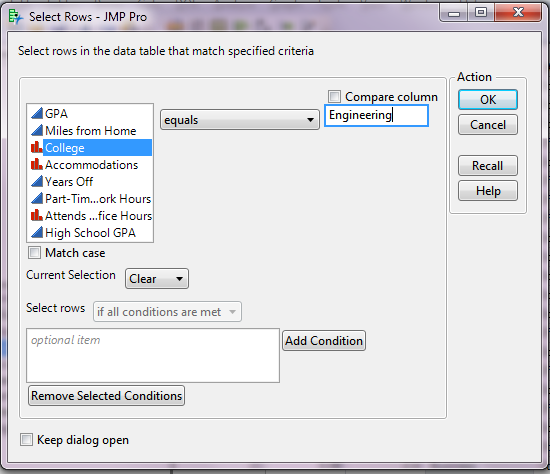
**Question 1**- Using the “freshmen” data set, provide screen shots of the following:

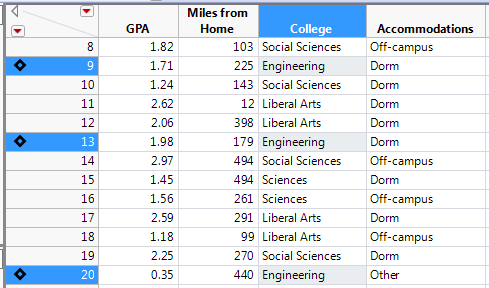
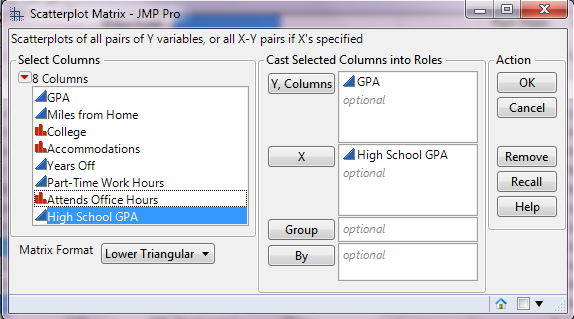
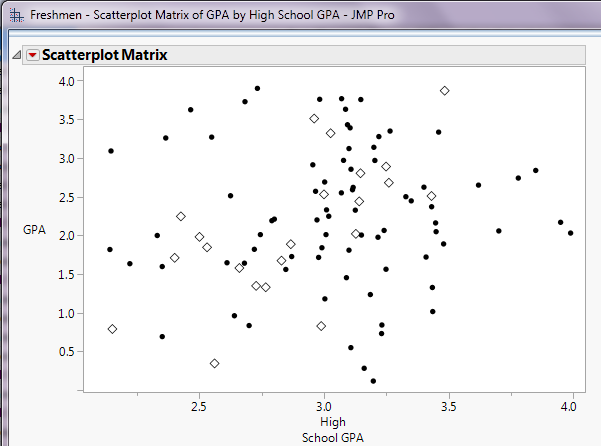
A) a scatterplot of GPA with Engineering students indicated by a diamond marker; repeat the process but this time have business students indicated by the color red;

B) a scatterplot of GPA with the data filtered so only students living dorms are shown

**Solution1A**-

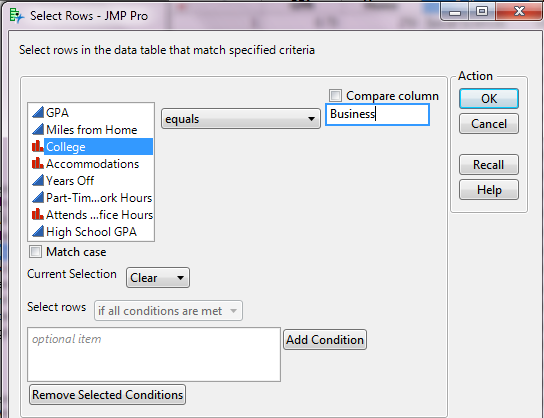
1. We have loaded the Freshmen.jmp file into the JMP Pro Application.
2. As per ask of the question we need to have a scatterplot of GPA with Engineering students indicated by a diamond marker. So we need to select the rows where College is Engineering and mark them with a Diamond icon.

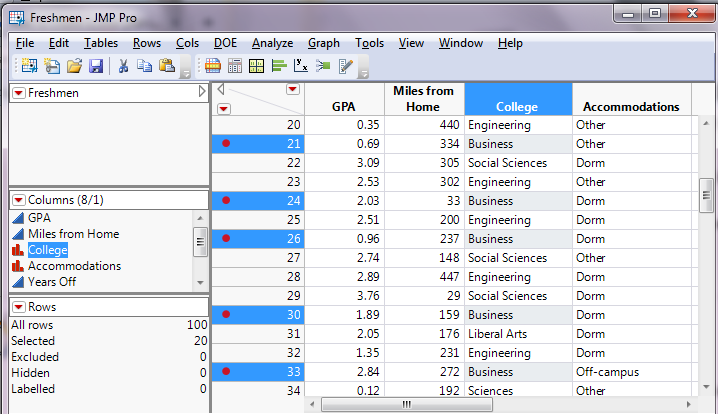
Rows > Row Selection>Select Where > Provide the following query

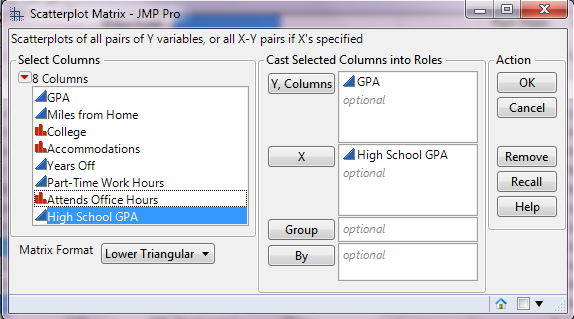
1. After this you will have all the rows selected in the table where College is Engineering.
2. Right click on any of the selected rows and select a diamond marker from the Markers option. Now the screen should look like this:
3. Now goto Graph > Scatterplot Matrix. Provide the inputs as follows:
4. You should get this screen where the GPA with Engineering students is indicated by a diamond marker:

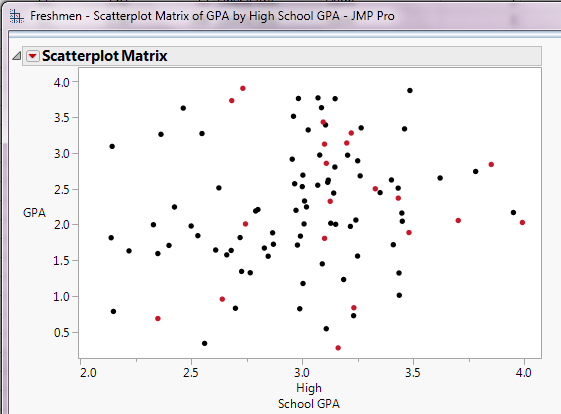
**Solution1A(Second Part)**-

1. We have loaded the Freshmen.jmp file into the JMP Pro Application.
2. As per ask of the question we need to have a scatterplot of GPA with business students indicated by the colour red. So we need to select the rows where College is ‘Business’ and mark them with red colour.

Rows > Row Selection>Select Where > Provide the following query

1. After this you will have all the rows selected in the table where College is Business.
2. Right click on any of the selected rows and select a red colour from the Colours option. Now the screen should look like this:
3. Now goto Graph > Scatterplot Matrix. Provide the inputs as follows:

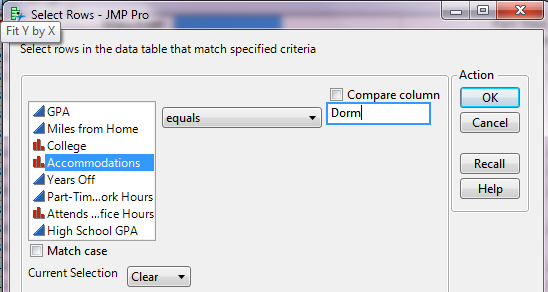


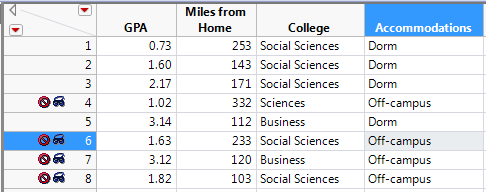
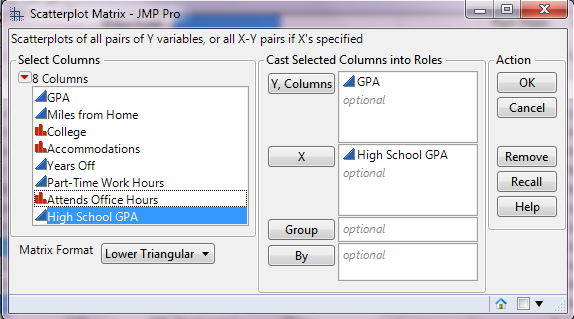
1. You should get this screen where the GPA with Business is indicated by the colour red:

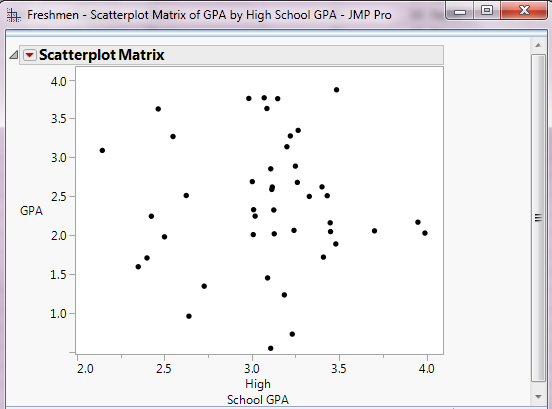
**Solution 1 B**

1. We have loaded the Freshmen.jmp file into the JMP Pro Application.
2. As per ask of the question we need to have a scatterplot of GPA with the data filtered so only students living dorms are shown. So we need to select the rows where Accomodations is ‘Dorm’.

Rows > Row Selection>Select Where > Provide the following query:

****

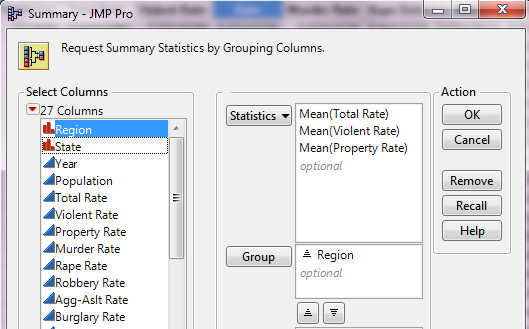
1. ****Right click on the Row and select ‘Invert Selection’ and ‘Hide and Exclude’ the selection. Now we are left with only the rows where Accommodations = ‘Dorm’
2. Now goto Graph > Scatterplot Matrix. Provide the inputs as follows:

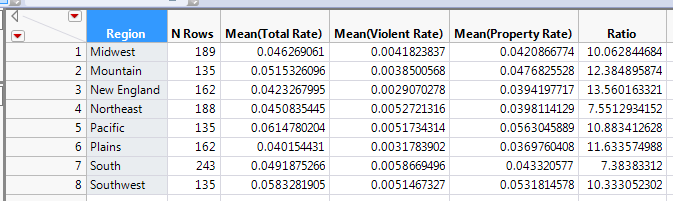
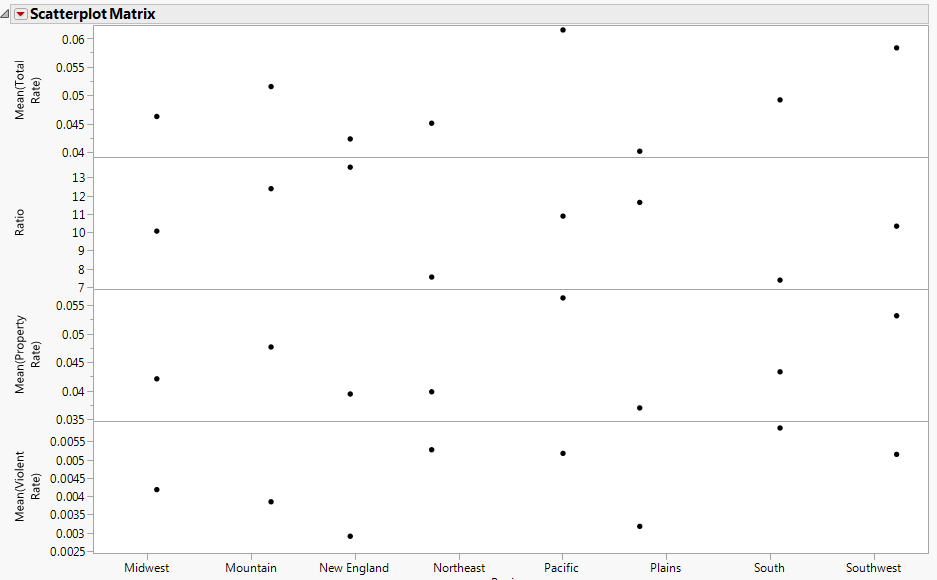
1. Now we have the Scatterplot where only students living in dorms are shown.

**Question 2** - Using Regional Crime Data, complete a summary table that provides the following information by region: average Total Rate, average Violent Rate, and average Property Rate.  Briefly summarize any large or surprising differences in the average crime rates across regions. Show your screen shot of the summary table.

**Solution 2** –

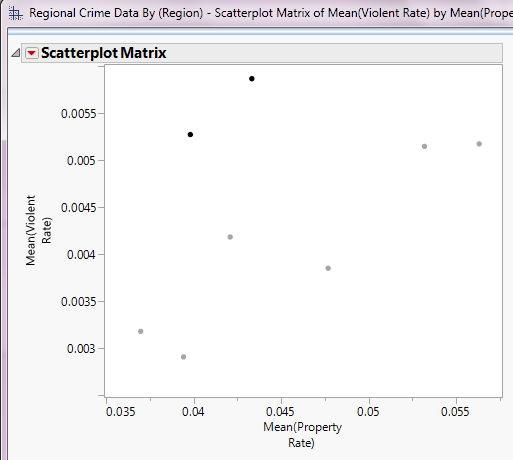
1. We have loaded the ‘Regional Crime Data’ file into the JMP Pro Application.
2. For the required summary table(considering ‘Group’ by ‘Region’ and ‘Statistics’ as ‘Mean’). We will create this summary table with the help of following path:

Tables > Summary> refer this data:

1. We get this Summary table:-
2. To further analyse these Crime rate (consisting of Violent Rate and Property Rate). I have created a new column ‘Ratio’ where I have taken ratio of Property Rate to Violent Rate. This helped me in finding out the Regions where the trend is different between Violent Rate and Property Rate.
3.  I have used a scatterplot, where I am analyzing the relationship between the means of Total Rate, Violent Rate and Property Rate.

Here the important observations are as follows:

1. Crime rate is highest for Pacific followed by Southwest.
2. Crime rate is lowest for Plains followed by New England.
3. For most of the regions, Property rate is correlated with Violent rate.
4. The only exception being the regions of ‘Northeast’ and ‘South’, where in spite of less Property rate, Crime rate is substantially high.
5. Below is the scatterplot between our 2 continuous variables ‘Property Rate’ vs ‘Violent Rate’. The dark points in the graph are ‘Northeast’ and ‘South’ regions.



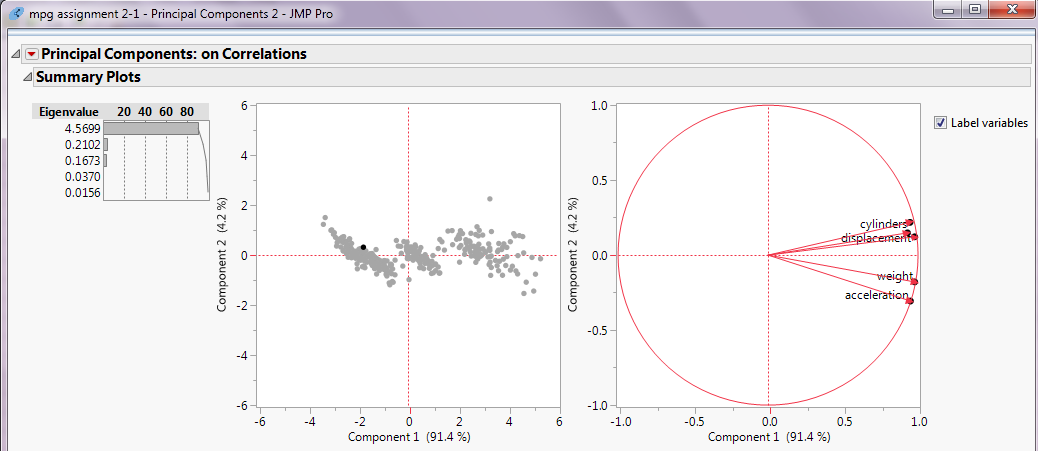
**Question 3** - (use the data set mpg assignment 2) mpg is the target variable; the other variables are explanatory variables. Complete a principal component analysis to determine if you can reduce the dimensionality.

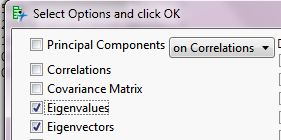
a) If you want to maintain 95% of the variation in the original explanatory variables, how many principal component variables would you use? Show the appropriate screen shot (what you are using to make the decision.)

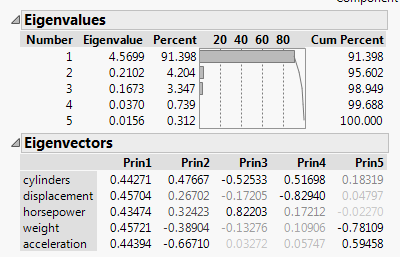
b) Based on correlation coefficients (rather than principal component analysis) what variables might be deleted? Briefly explain and show appropriate screen shot.

**Solution 3** -

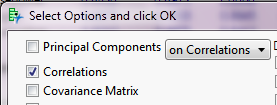
1. Load the data set into jmp.
2. Select ‘Principal Component Analysis’ under the Analyze > MultiVariate Methods.
3. Since PCA need to be done for the explanatory variables. We will select them and will get the following chart after pressing OK.

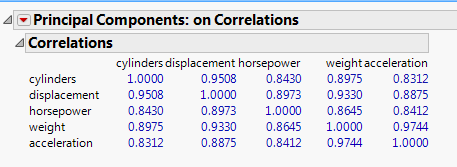


**Solution 3 A)** - In order to maintain 95% of the variation in the original explanatory variables, we need to know the cumulative variation coverage of every eigenvalue. Here is how to do this.

We will get the chart as follows:

As we can see from the graph above that in order to achieve **95% of the variation in the original explanatory variables, we need to keep first two Eigenvalues**.

**Solution 3 B)** To reduce the dimensions based on correlation coefficients, we need to enable ‘Correlation’ from the red drop down menu:

 We will get the following table:

On analyzing the data, we can find high correlation between cylinders & displacement and Weight & acceleration. Since we need to reduce the dimensionality we will remove the ones with maximum correlation. So we can remove **displacement** and **weight** variables.