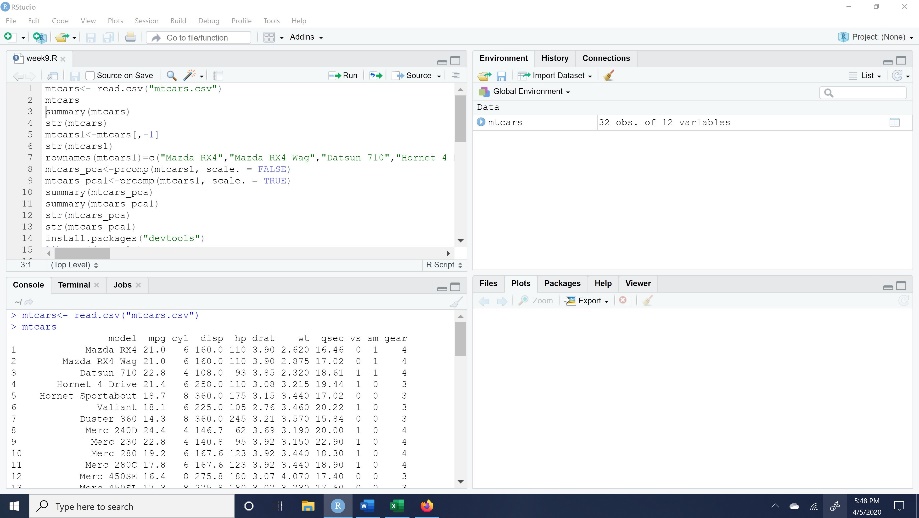
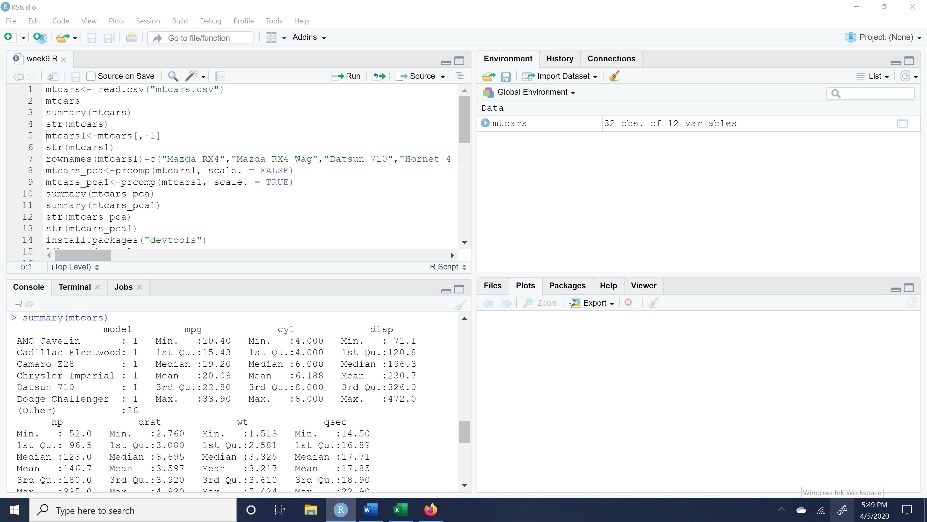
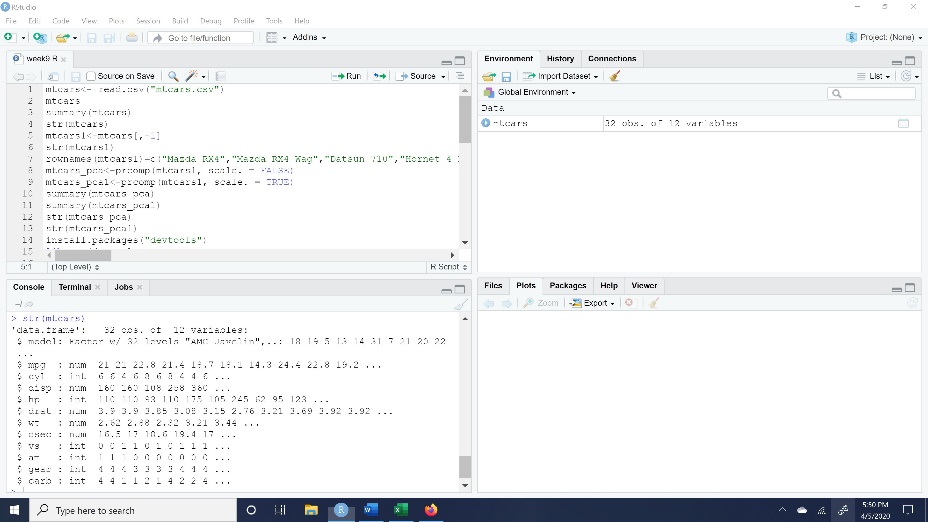
Principal Components Analysis Execution in R studio using R commands are given below with screenshots:

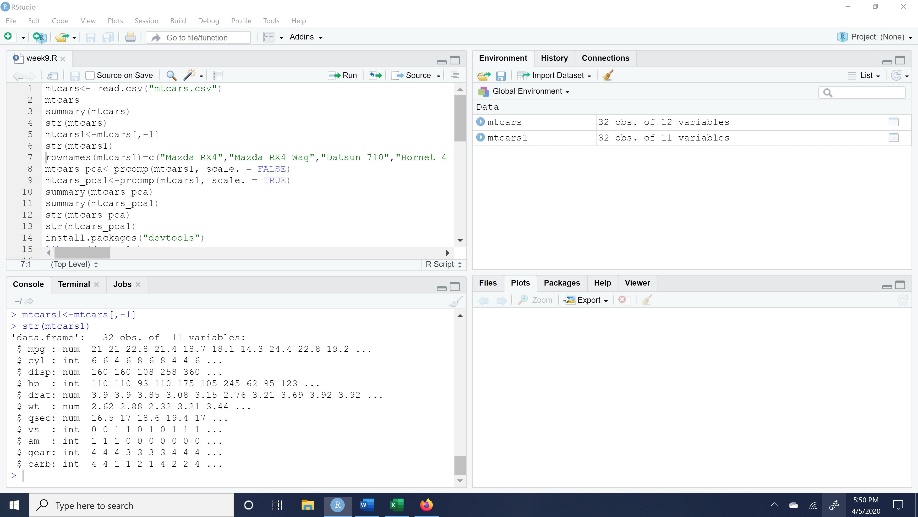
Step1: In this step reading the .csv file into R studio.



Step2: Data Exploration and data preprocessing:



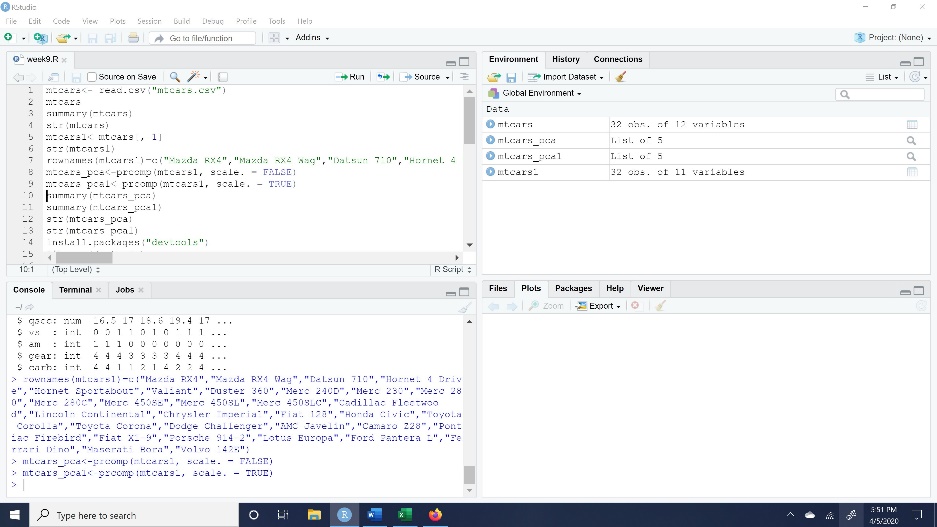


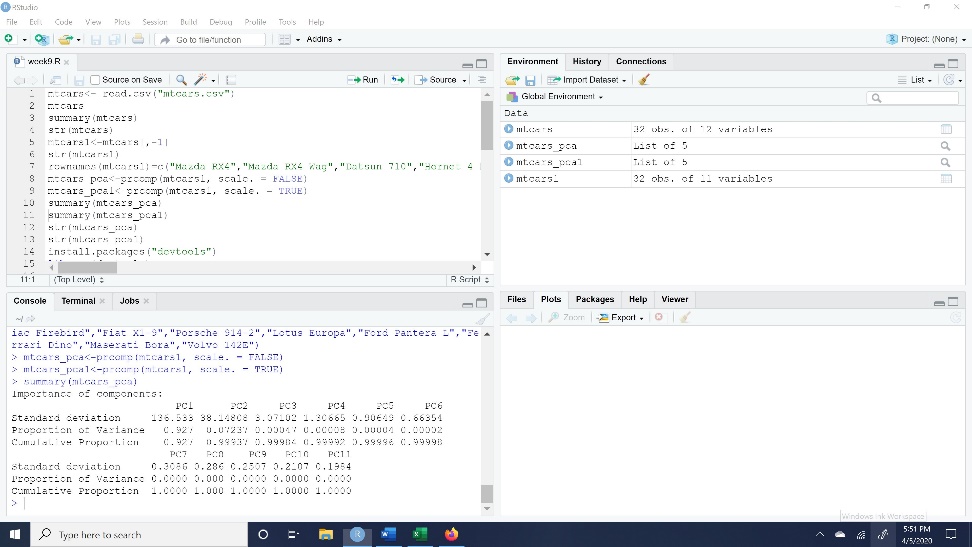


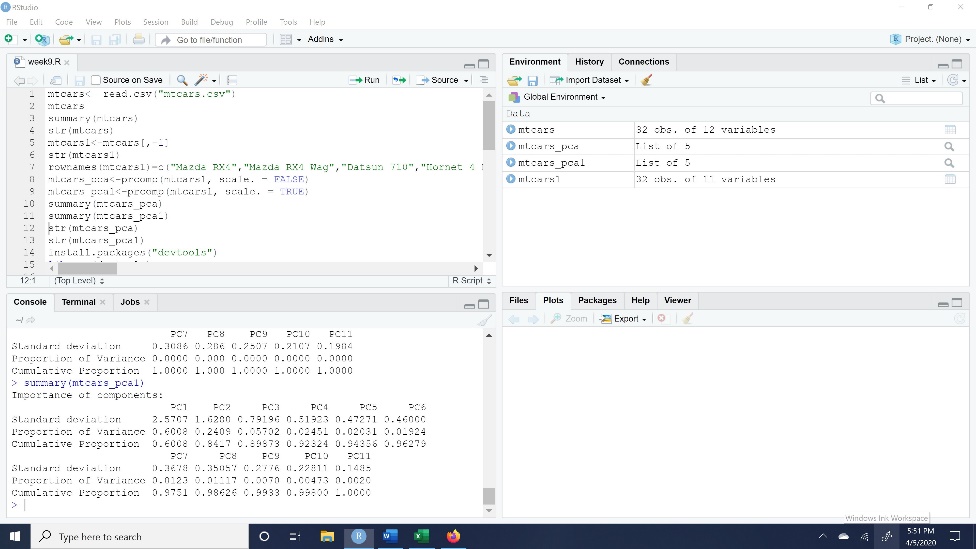
Step3:

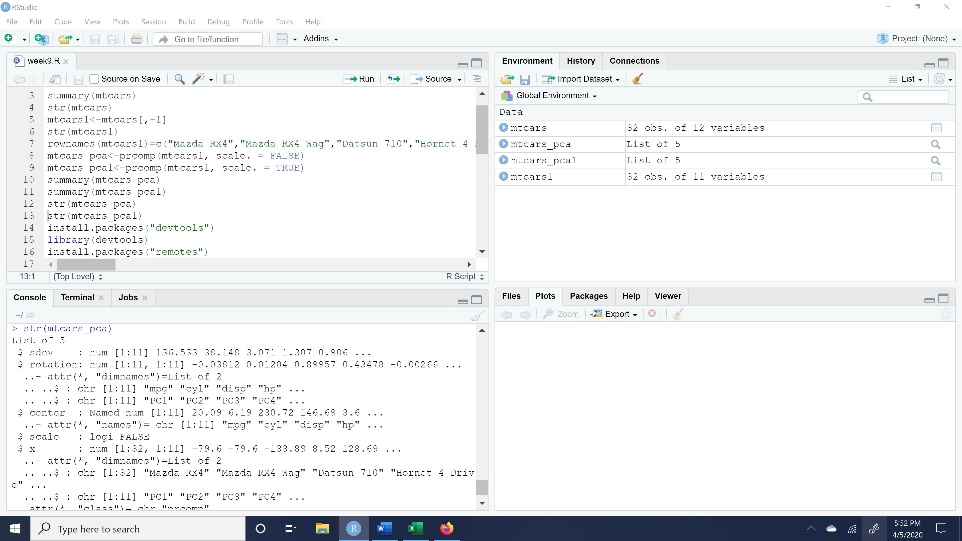
1. Should be principal components analysis of this data be based on the covariance or the correlation matrix? Explain.

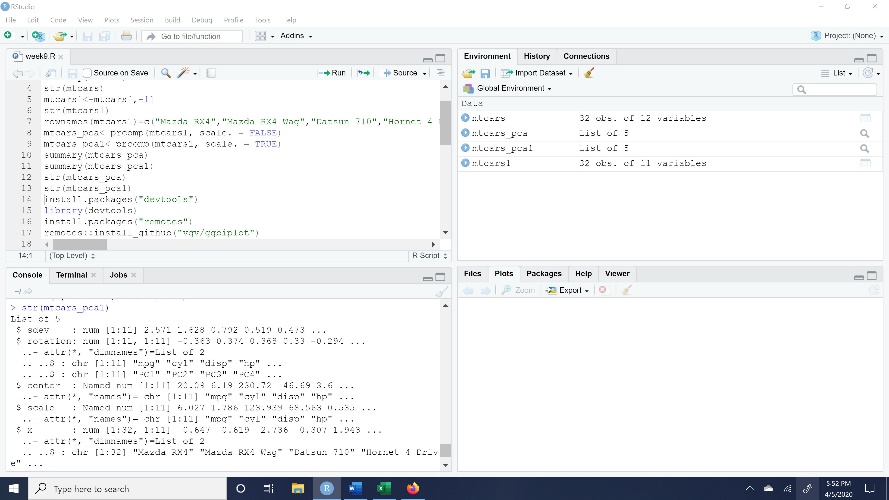
The better advice is to use covariance matrix when the variables are on similar scales and the correlation matrix, when the scales of the variables are different. PCA with covariance matrix, for that we need to set the ‘scale’ option as FALSE. As a conclusion, not a lot of significant insights can be driven from the principal component analysis based on the covariance matrix. PCA with correlation analysis, to do this we need to set the ‘scale’ as TRUE. To conclude this, we can draw relational connect between the variables and most significant difference can be seen at the standard deviation values in both covariance and correlation. As we can see the values of PCA done by using correlation matrix are closer to each other and more uniform as compared to the analysis done with the covariance matrix.

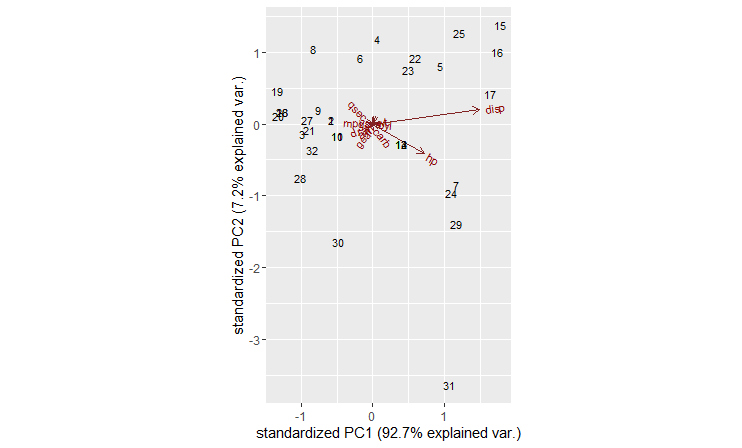
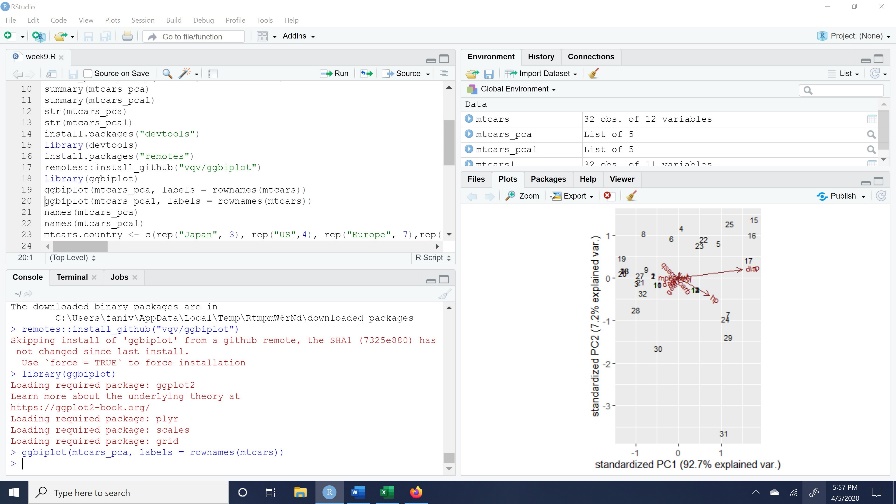


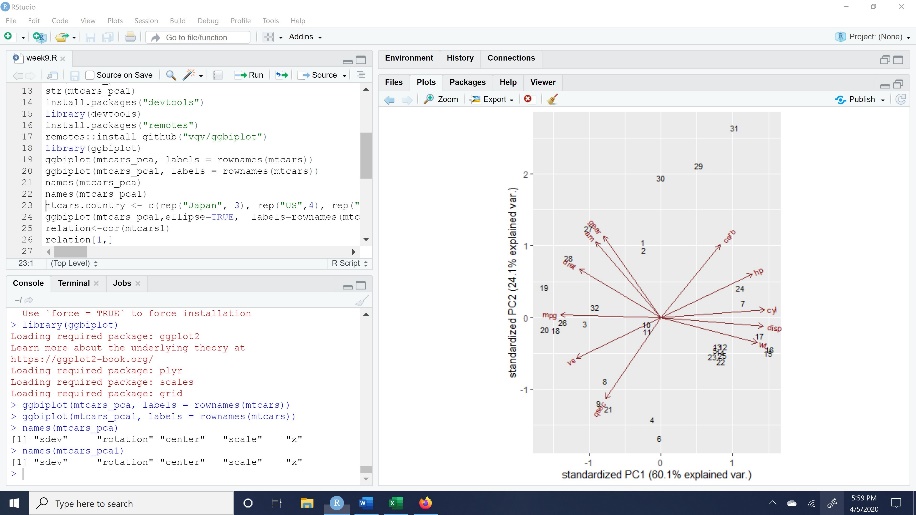


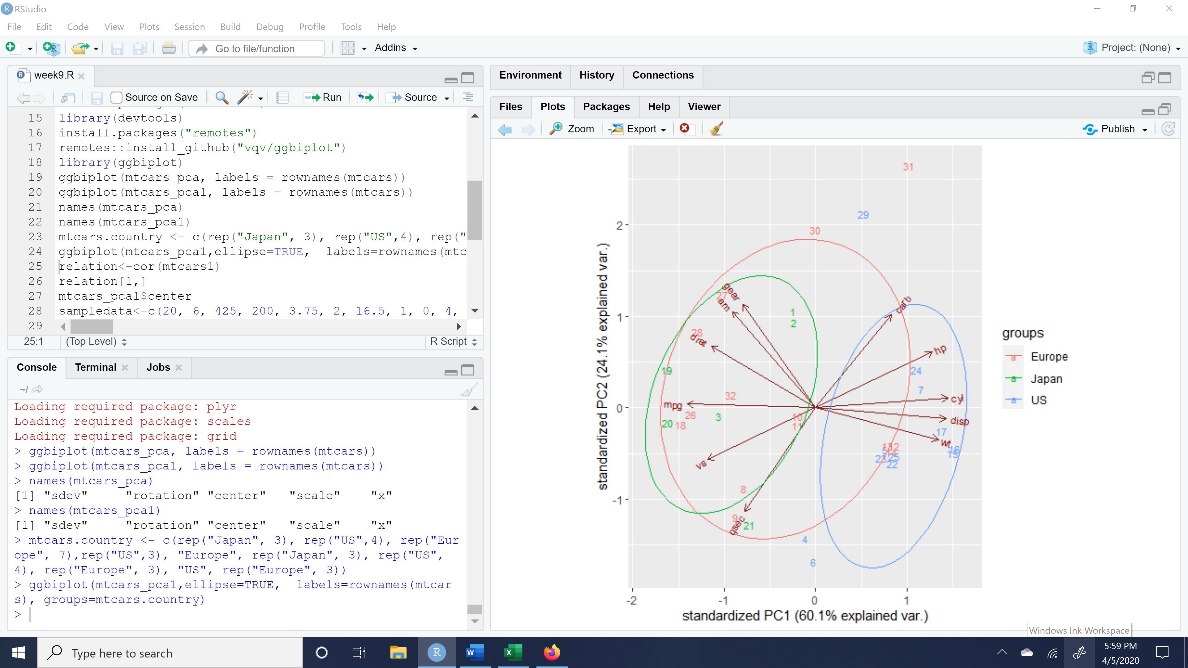
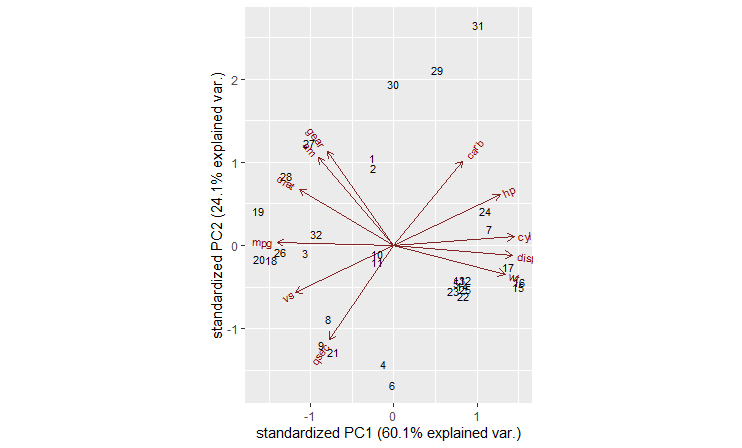






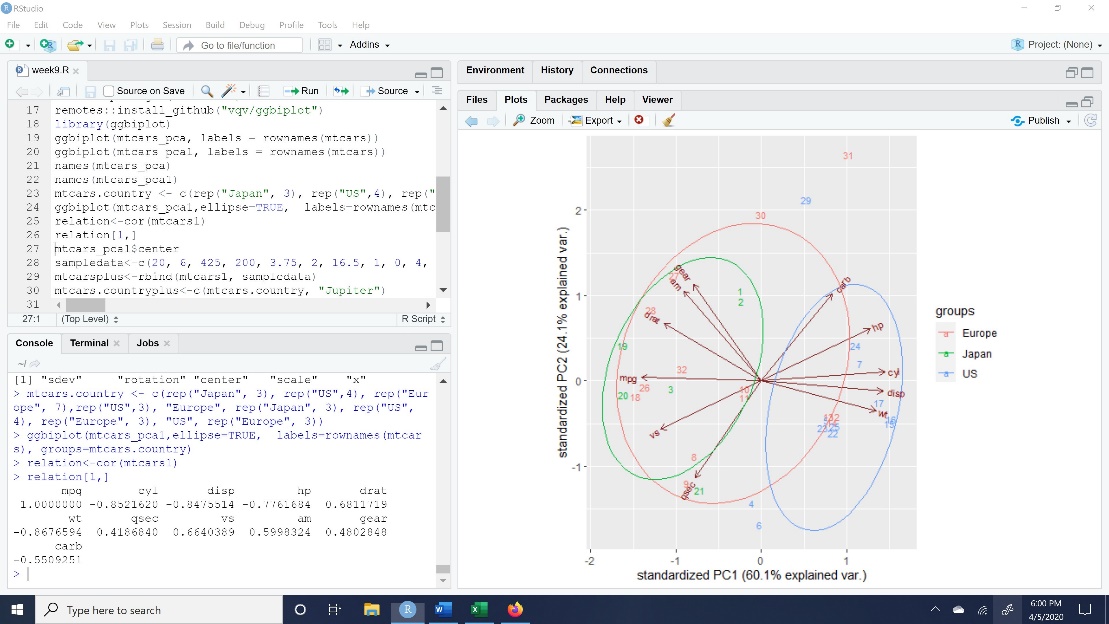




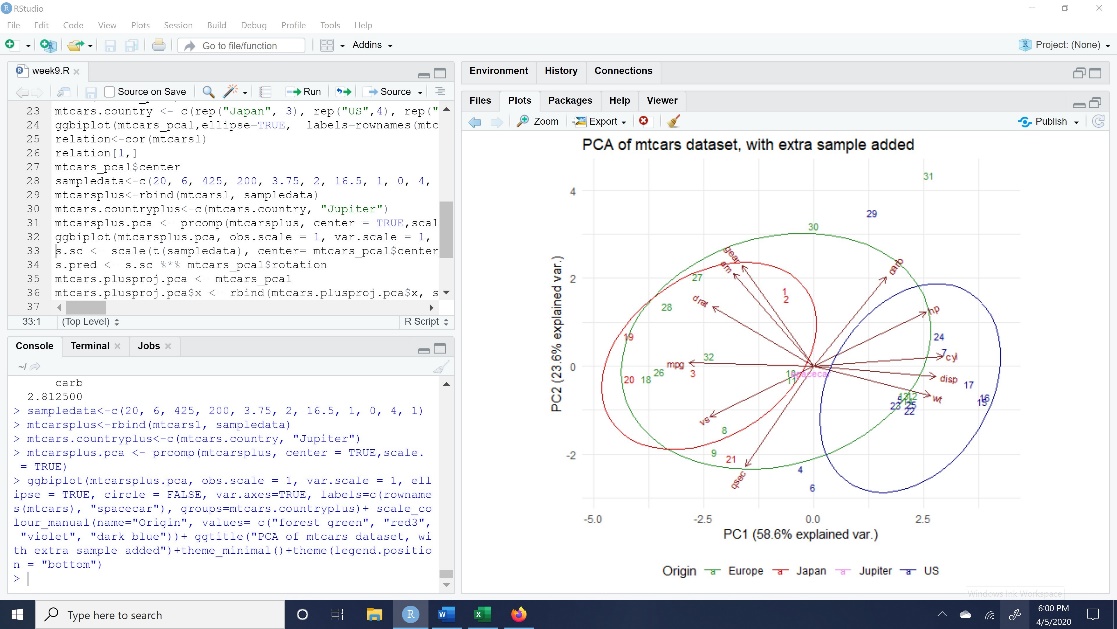


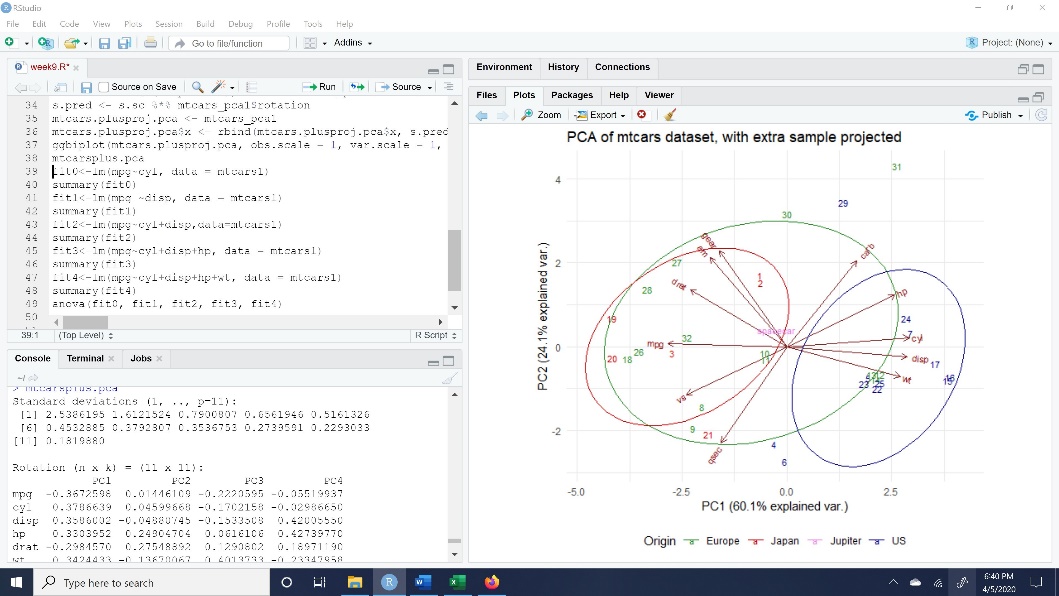
1. Which variables seem to have the strongest relation to the mileage?

Mpg has a very strong correlation with cyl, disp, hp, wt and carb.

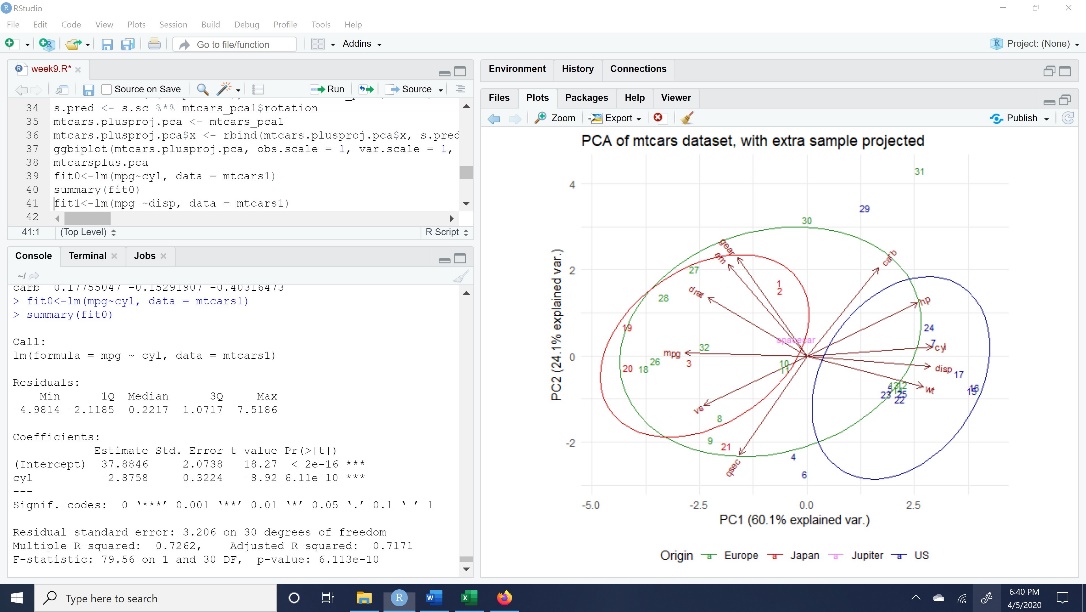


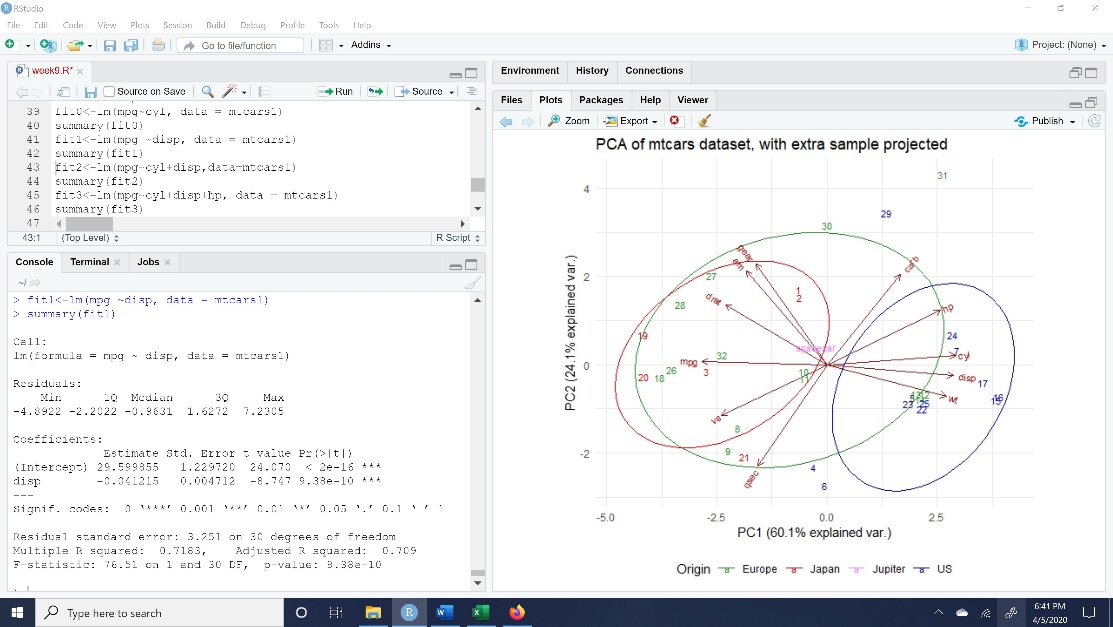
1. Suppose my car gets 20 mpg, has 6 cylinders, a displacement of 425, 200 horsepower, a rear axle ratio of 3.75, weighs2000 pounds, can go a quarter mile in 16.5 seconds, has v/s (vertical steering?), automatic transmission, 4 gears and 1 carburetor. What are its scores on the first and second principal components? What sort of car, if any, is it most like?

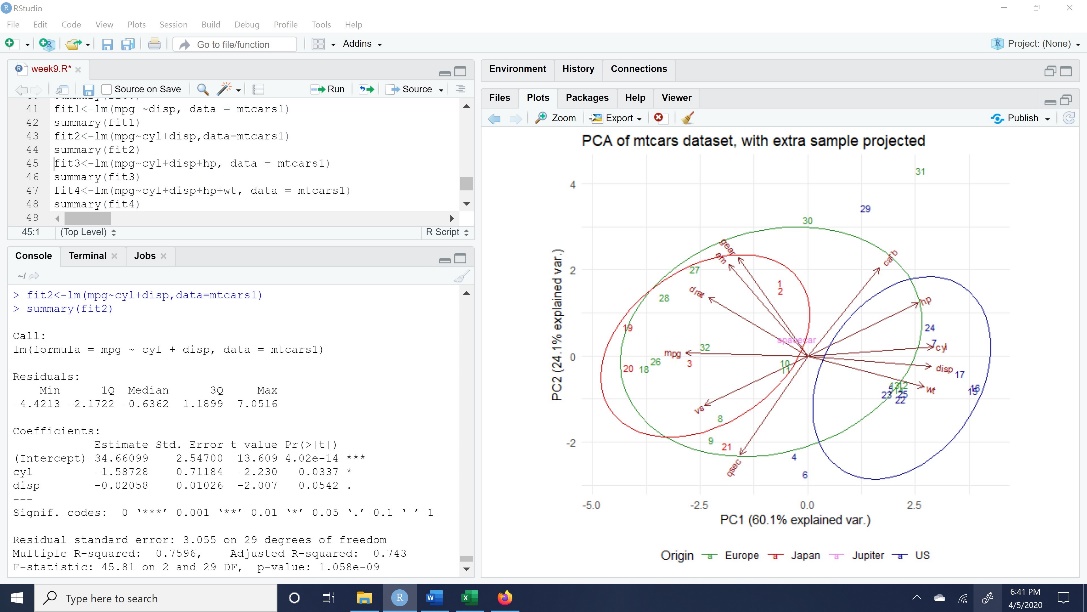


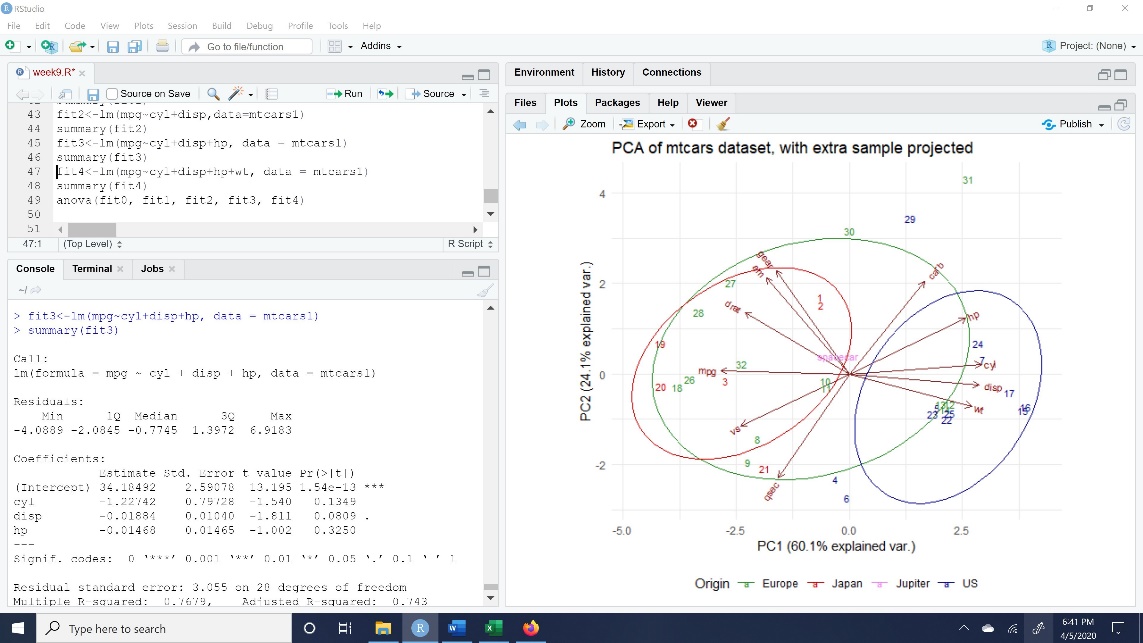


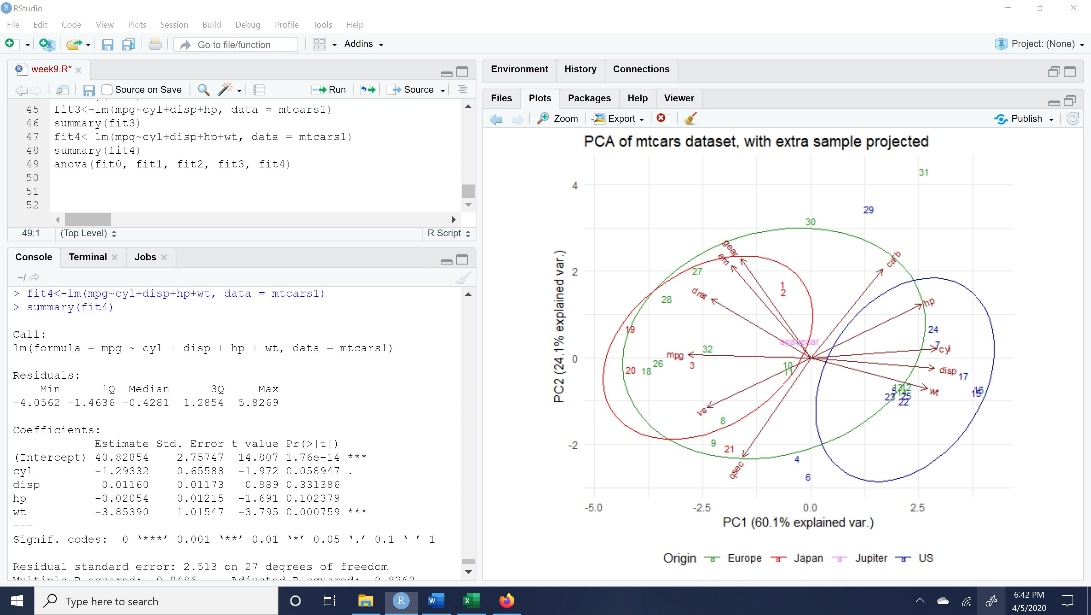
1. Fit a regression to predict mpg. Evaluate the fit. What can be done to improve it?





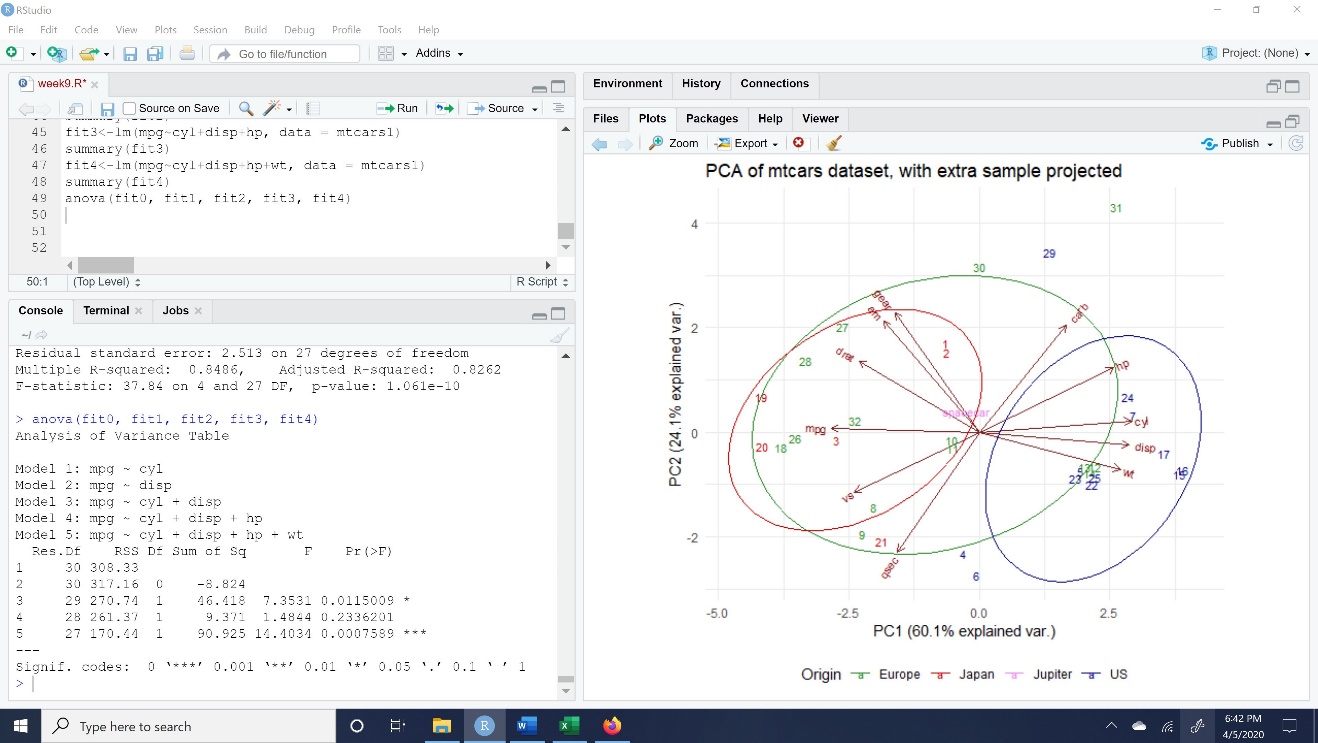






1. Try your suggestion. Did it help? What’s the best predictor of mileage?

As per the anova R command the best fit is fit4 which is the combination of cyl+disp+hp+wt so if we add the variables the predicted values are as close as the actual values.



1. Are Mercedes different from other cars? If so, what characteristic would you say they share?

Mercedes lies on qsec line in the graph. When I observe the dataset Mercedes has this qsec high as compared to other cars.

1. What characteristics separate sports cars from the others?

Gear and crab characteristics separate sports cars from others.

1. Suppose your car gets good mileage. What else is likely to be true about it?

If a car has good mileage, the number of cylinders (cyl) should be less in number.

References:

1. <https://towardsdatascience.com/let-us-understand-the-correlation-matrix-and-covariance-matrix-d42e6b643c22>
2. <https://www.datacamp.com/community/tutorials/pca-analysis-r>
3. https://rstudio-pubs static.s3.amazonaws.com/36290\_2793b3e2583146e2b78f533592488972.html