Homework 1

Problem 1

The *Motor Trend Car Road Tests* dataset (mtcars.csv) contains the data extracted from the 1974 Motor Trend US magazine and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models). The dataset contains 32 observations on 11 variables. Please analyze the dataset according to the following steps:

1. Calculate sample mean and sample variance of each variable.

   1.1 Mean

mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
20.0906	6.1875	230.7219	146.6875	3.5966	3.2172	17.8487	0.4375	0.4062	3.6875	2.8125

   1.2 Variance

mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
36.3241	3.1895	15360.7998	4700.8669	0.2859	0.9574	3.1932	0.2540	0.2490	0.5444	2.6089

2. Calculate sample variance-covariance matrix and correlation matrix. What can you say about the variance-covariance matrix and correlation matrix?

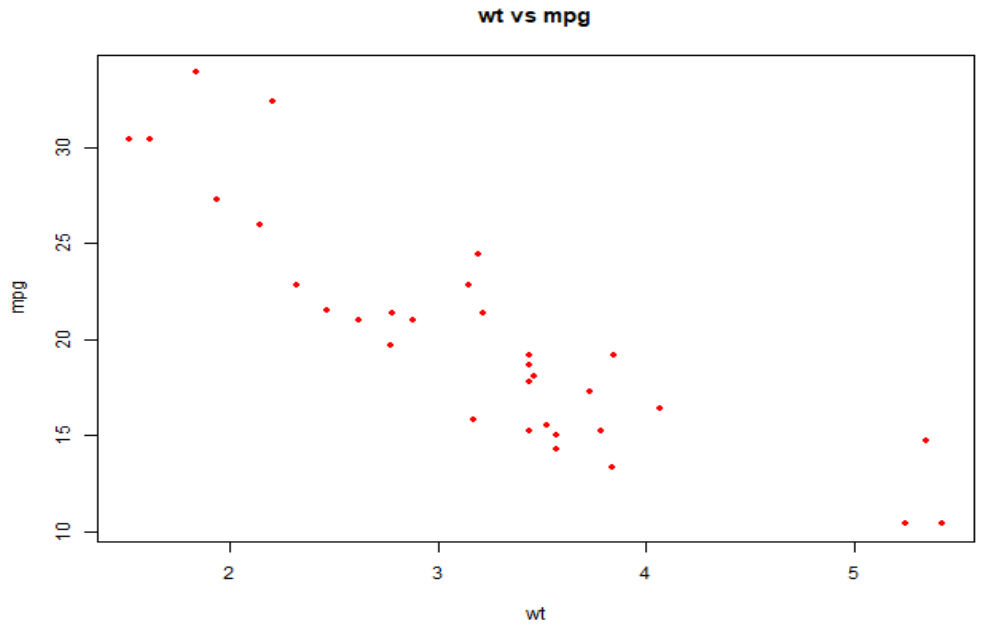
   2.1 Covariance matrix

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
mpg	36.3241	-9.1724	-633.0972	-320.7321	2.1951	-5.1167	4.5091	2.0171	1.8039	2.1357	-5.3631
cyl	-9.1724	3.1895	199.6603	101.9315	-0.6684	1.3674	-1.8869	-0.7298	-0.4657	-0.6492	1.5202
disp	-633.0972	199.6603	15360.7998	6721.1587	-47.0640	107.6842	-96.0517	-44.3776	-36.5640	-50.8026	79.0687
hp	-320.7321	101.9315	6721.1587	4700.8669	-16.4511	44.1927	-86.7701	-24.9879	-8.3206	-6.3589	83.0363
drat	2.1951	-0.6684	-47.0640	-16.4511	0.2859	-0.3727	0.0871	0.1186	0.1902	0.2760	-0.0784
wt	-5.1167	1.3674	107.6842	44.1927	-0.3727	0.9574	-0.3055	-0.2737	-0.3381	-0.4211	0.6758
qsec	4.5091	-1.8869	-96.0517	-86.7701	0.0871	-0.3055	3.1932	0.6706	-0.2050	-0.2804	-1.8941
vs	2.0171	-0.7298	-44.3776	-24.9879	0.1186	-0.2737	0.6706	0.2540	0.0423	0.0766	-0.4637
am	1.8039	-0.4657	-36.5640	-8.3206	0.1902	-0.3381	-0.2050	0.0423	0.2490	0.2923	0.0464
gear	2.1357	-0.6492	-50.8026	-6.3589	0.2760	-0.4211	-0.2804	0.0766	0.2923	0.5444	0.3266
carb	-5.3631	1.5202	79.0687	83.0363	-0.0784	0.6758	-1.8941	-0.4637	0.0464	0.3266	2.6089

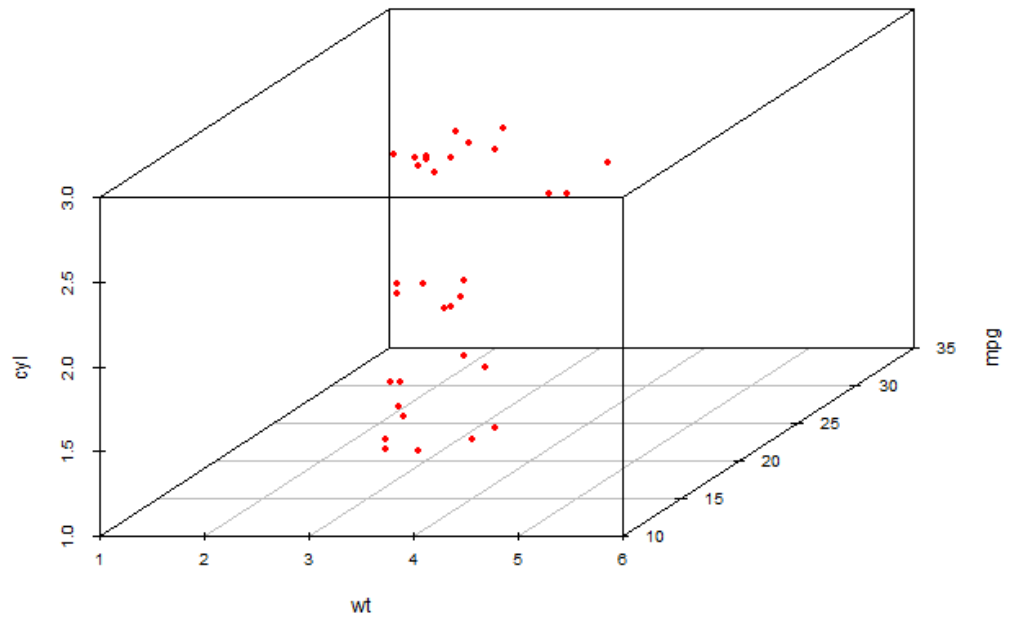
   2.2 Correlation matrix

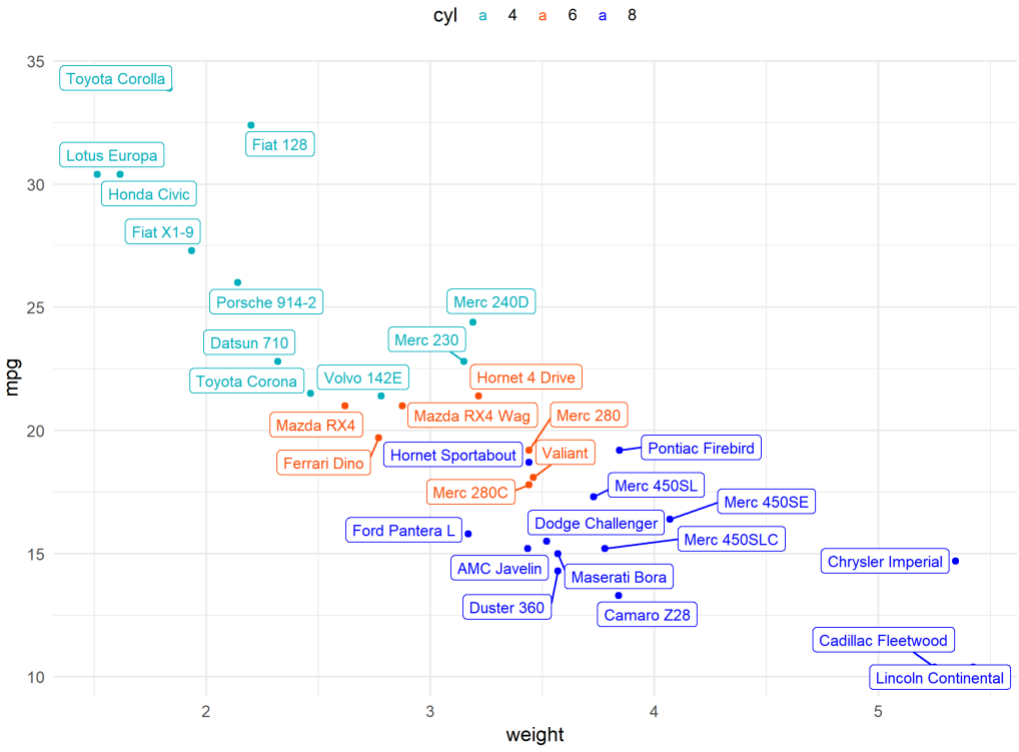
	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
mpg	1.0000	-0.8522	-0.8476	-0.7762	0.6812	-0.8677	0.4187	0.6640	0.5998	0.4803	-0.5509
cyl	-0.8522	1.0000	0.9020	0.8324	-0.6999	0.7825	-0.5912	-0.8108	-0.5226	-0.4927	0.5270
disp	-0.8476	0.9020	1.0000	0.7909	-0.7102	0.8880	-0.4337	-0.7104	-0.5912	-0.5556	0.3950
hp	-0.7762	0.8324	0.7909	1.0000	-0.4488	0.6587	-0.7082	-0.7231	-0.2432	-0.1257	0.7498
drat	0.6812	-0.6999	-0.7102	-0.4488	1.0000	-0.7124	0.0912	0.4403	0.7127	0.6996	-0.0908
wt	-0.8677	0.7825	0.8880	0.6587	-0.7124	1.0000	-0.1747	-0.5549	-0.6925	-0.5833	0.4276
qsec	0.4187	-0.5912	-0.4337	-0.7082	0.0912	-0.1747	1.0000	0.7445	-0.2299	-0.2127	-0.6562
vs	0.6640	-0.8108	-0.7104	-0.7231	0.4403	-0.5549	0.7445	1.0000	0.1683	0.2060	-0.5696
am	0.5998	-0.5226	-0.5912	-0.2432	0.7127	-0.6925	-0.2299	0.1683	1.0000	0.7941	0.0575
gear	0.4803	-0.4927	-0.5556	-0.1257	0.6996	-0.5833	-0.2127	0.2060	0.7941	1.0000	0.2741
carb	-0.5509	0.5270	0.3950	0.7498	-0.0908	0.4276	-0.6562	-0.5696	0.0575	0.2741	1.0000

3. Draw a scatter plot between wt (Weight) and mpg (Miles per gallon).

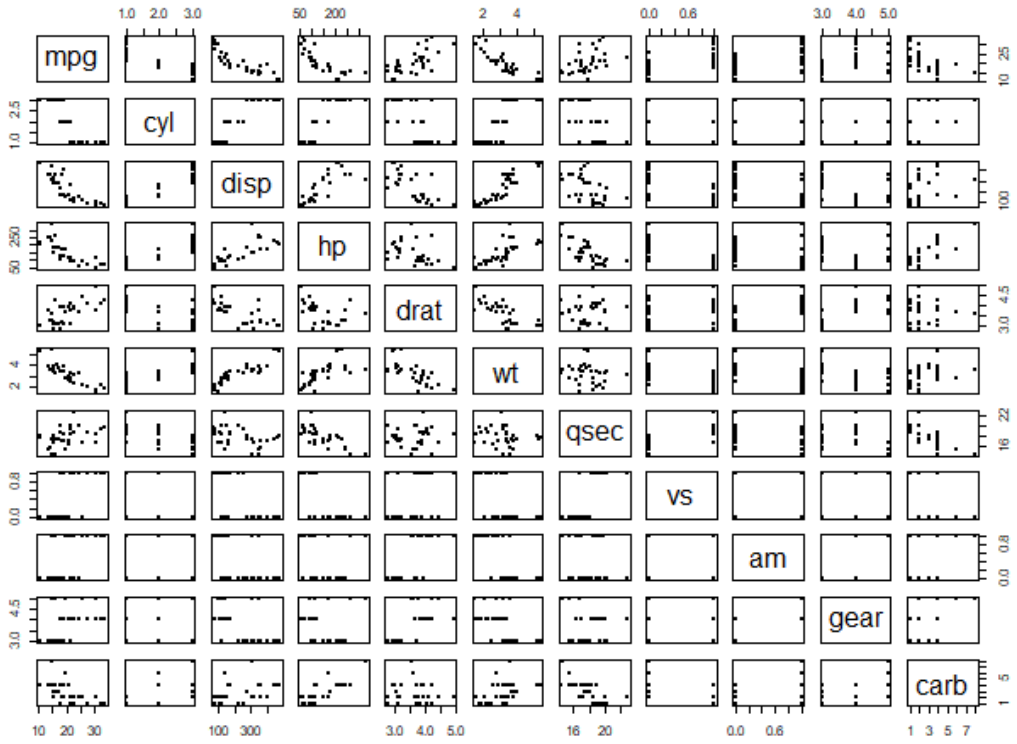


4. Draw a scatter plot to show the relationship between wt (Weight), mpg (Miles per gallon) and cyl (Number of cylinders). You can use 3D scatter plot or add cyl to your 2D scatter plot as the color of points.

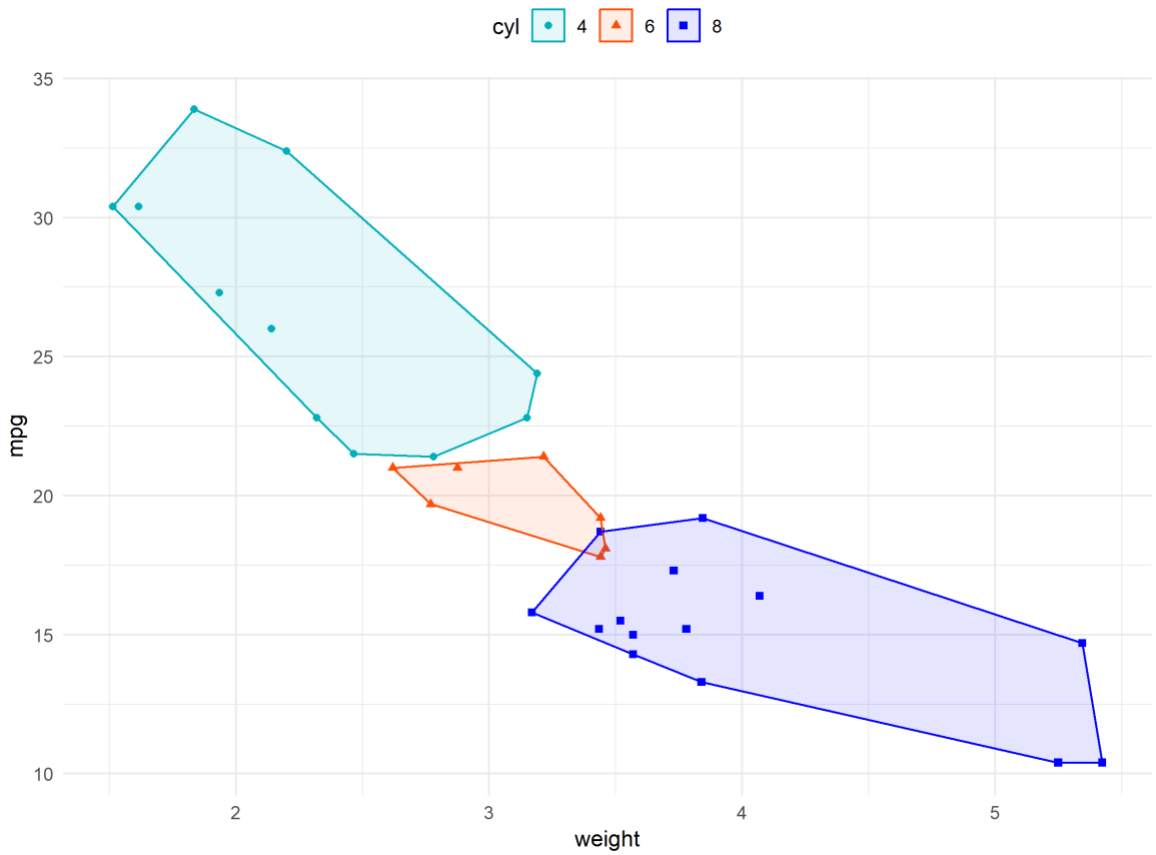




5. Draw pairwise scatter plot for all variables.



6. One engineer suggests that the relationship between wt and mpg is subject to the number of cylinders. According to the plot you draw in 4, what is your opinion towards this suggestion?



The suggestion is reasonable as the observations in the scatterplot clustered into three groups according to their number of cylinders.