Homework 2

R Crash Course (20 points)

Instruction:

- This HW must be done in **Rmarkdown!**
- Please submit both the .rmd and the Microsoft word files.
- Please do not submit PDF files or image formats! the TAs are going to give you feedback in your word document.
- All the HW assignments are group work. However, you need to submit it individually.
- Late homework assignments will not be accepted under any circumstances.
- The answer key will be uploaded on Canvas a couple of days after the due date.

Question 1 Replicate the following vector of numbers using seq() function and call it **v1**.

```
[1]  1  5  9 13 17
```

Question 2 (I) Replicate the following matrix in R and call it A.

```
     [,1] [,2] [,3] [,4] [,5]
[1,]   20   21   22   23   24
[2,]   25   26   27   28   29
[3,]   30   31   32   33   34
[4,]   35   36   37   38   39
```

(II) Rename the columns as A,B,C,D, and E.

```
      A  B  C  D  E
[1,] 20 21 22 23 24
[2,] 25 26 27 28 29
[3,] 30 31 32 33 34
[4,] 35 36 37 38 39
```

(III) Extract the following matrix from A and call it B.

```
     B  D
[1,] 26 28
[2,] 36 38
```

(IV) Find the transpose of matrix B.

```
  [,1] [,2]
B   26   36
D   28   38
```

(V) Find the inverse of Matrix B, and call it B_inverse.

```
  [,1] [,2]
B -1.9  1.4
D  1.8 -1.3
```

(VI) Multiply B and B_inverse. What is the name of the new matrix?

Question 3 (I) Assign the dataframe *mtcars* from the built-in datasets in R to a new dataframe and call it **df**. Show the first 10 rows of your dataframs:

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4

For the following questions, you must call the *dplyr* package first and then use the functions *select, filter, mutate* and *arrange* when appropriate

(II) Make a new dataframe based on *df* and call it *df1*. Show the first 5 rows of **df1**.

	cyl	hp	wt	vs	am	gear	mpg
Datsun 710	4	93	2.320	1	1	4	22.8
Merc 240D	4	62	3.190	1	0	4	24.4
Merc 230	4	95	3.150	1	0	4	22.8
Fiat 128	4	66	2.200	1	1	4	32.4
Honda Civic	4	52	1.615	1	1	4	30.4

(III) In **df1** make a new column called **gpm**. Show the first 5 rows again.

```
  cyl hp    wt vs am gear  mpg   gpm
1   4 93 2.320  1  1    4 22.8 0.044
2   4 62 3.190  1  0    4 24.4 0.041
3   4 95 3.150  1  0    4 22.8 0.044
4   4 66 2.200  1  1    4 32.4 0.031
5   4 52 1.615  1  1    4 30.4 0.033
```

(IV) Show the structure of **df1**.

```
'data.frame':   32 obs. of  8 variables:
 $ cyl : num  4 4 4 4 4 4 4 4 4 4 ...
 $ hp  : num  93 62 95 66 52 65 97 66 91 113 ...
 $ wt  : num  2.32 3.19 3.15 2.2 1.61 ...
 $ vs  : num  1 1 1 1 1 1 1 1 0 1 ...
 $ am  : num  1 0 0 1 1 1 0 1 1 1 ...
 $ gear: num  4 4 4 4 4 4 3 4 5 5 ...
 $ mpg : num  22.8 24.4 22.8 32.4 30.4 33.9 21.5 27.3 26 30.4 ...
 $ gpm : num  0.044 0.041 0.044 0.031 0.033 0.029 0.047 0.037 0.038 0.033 ...
```

(V) Change the type of the following variables **(vs, am, cyl, gear)** from **num** to **factor**. Show the structure of **df1** again.

```
'data.frame':   32 obs. of  8 variables:
 $ cyl : Factor w/ 3 levels "4","6","8": 1 1 1 1 1 1 1 1 1 1 ...
 $ hp  : num  93 62 95 66 52 65 97 66 91 113 ...
 $ wt  : num  2.32 3.19 3.15 2.2 1.61 ...
 $ vs  : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 1 2 ...
 $ am  : Factor w/ 2 levels "Automatic","Mannual": 2 1 1 2 2 2 1 2 2 2 ...
 $ gear: Factor w/ 3 levels "3","4","5": 2 2 2 2 2 2 1 2 3 3 ...
 $ mpg : num  22.8 24.4 22.8 32.4 30.4 33.9 21.5 27.3 26 30.4 ...
 $ gpm : num  0.044 0.041 0.044 0.031 0.033 0.029 0.047 0.037 0.038 0.033 ...
```

Computer Exercises from Wooldridge Textbook Ch1.

Question 4 The data set in *ALCOHOL* contains information on a sample of men in the United States. Two key variables are self-reported employment status and alcohol abuse (along with many other variables). The variables *employ* and *abuse* are both binary, or indicator, variables: they take on only the values zero and one.

Hint 1: To answer this question, you should use *table()* and *prop.table()* functions.

Hint 2: Unemployment rate is defined as $Unemplomentrate = \frac{unemployed}{unemployed+employed}$.

Hint 3; For unemployment rate, you only need to work with the status variable from the dataset.We do not include the *"out of workforce"* in the denominator..

(I) What is percentage of the men in the sample report abusing alcohol?

(II) What is the employment rate?

(III) Consider the group of men who abuse alcohol. What is the employment rate?

(IV) What is the employment rate for the group of men who do not abuse alcohol?

(V) Discuss the difference in your answers to parts (III) and (IV). Does this allow you to conclude that alcohol abuse causes unemployment?