WEEK-10 Assignment-Data Wrangling in R

In [ ]:
Question 1 Load the mtcars dataset using the data() function, convert it into a tibble,
and explore it using the summary() function.

In [34]:
```
library(tidyverse)
data(mtcars)
as_tibble(mtcars)
```

mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2

In [4]:
```
summary(mtcars)
```

```
      mpg             cyl             disp             hp       
 Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0  
 1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5  
 Median :19.20   Median :6.000   Median :196.3   Median :123.0  
 Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7  
 3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0  
 Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0  
      drat             wt             qsec             vs        
 Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000  
 1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000  
 Median :3.695   Median :3.325   Median :17.71   Median :0.0000  
 Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375  
 3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000  
 Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000  
       am              gear            carb      
 Min.   :0.0000   Min.   :3.000   Min.   :1.000  
 1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000  
 Median :0.0000   Median :4.000   Median :2.000  
 Mean   :0.4062   Mean   :3.688   Mean   :2.812  
 3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000  
 Max.   :1.0000   Max.   :5.000   Max.   :8.000  
```

In [ ]:
Question 2 Select all of the columns of your mtcars tibble except the qsec and vs variables.

In [9]:
```
mtcars%>%select(-c(qsec,vs))%>%as_tibble
```

mpg	cyl	disp	hp	drat	wt	am	gear	carb
21.0	6	160.0	110	3.90	2.620	1	4	4
21.0	6	160.0	110	3.90	2.875	1	4	4
22.8	4	108.0	93	3.85	2.320	1	4	1
21.4	6	258.0	110	3.08	3.215	0	3	1
18.7	8	360.0	175	3.15	3.440	0	3	2
18.1	6	225.0	105	2.76	3.460	0	3	1
14.3	8	360.0	245	3.21	3.570	0	3	4
24.4	4	146.7	62	3.69	3.190	0	4	2
22.8	4	140.8	95	3.92	3.150	0	4	2
19.2	6	167.6	123	3.92	3.440	0	4	4
17.8	6	167.6	123	3.92	3.440	0	4	4
16.4	8	275.8	180	3.07	4.070	0	3	3
17.3	8	275.8	180	3.07	3.730	0	3	3
15.2	8	275.8	180	3.07	3.780	0	3	3
10.4	8	472.0	205	2.93	5.250	0	3	4
10.4	8	460.0	215	3.00	5.424	0	3	4
14.7	8	440.0	230	3.23	5.345	0	3	4
32.4	4	78.7	66	4.08	2.200	1	4	1
30.4	4	75.7	52	4.93	1.615	1	4	2
33.9	4	71.1	65	4.22	1.835	1	4	1
21.5	4	120.1	97	3.70	2.465	0	3	1
15.5	8	318.0	150	2.76	3.520	0	3	2
15.2	8	304.0	150	3.15	3.435	0	3	2
13.3	8	350.0	245	3.73	3.840	0	3	4
19.2	8	400.0	175	3.08	3.845	0	3	2
27.3	4	79.0	66	4.08	1.935	1	4	1
26.0	4	120.3	91	4.43	2.140	1	5	2
30.4	4	95.1	113	3.77	1.513	1	5	2
15.8	8	351.0	264	4.22	3.170	1	5	4
19.7	6	145.0	175	3.62	2.770	1	5	6
15.0	8	301.0	335	3.54	3.570	1	5	8
21.4	4	121.0	109	4.11	2.780	1	4	2

In [10]:
```
mtcars%>%as_tibble%>%filter(cyl!=8)
```

mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2

In [ ]:
Question 4 Group the mtcars tibble by the gear variable, summarize the medians of the mpg and disp variables,
and mutate a new variable that is the mpg median divided by the disp median, all chained together with the %>% operator.

In [11]:
```
mtcars%>%as_tibble%>%group_by(gear)
```

mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2

In [12]:
```
mtcars%>%as_tibble%>%group_by(gear)%>%summarise("mpg/disp"=median(mpg)/median(disp))
```

gear	mpg/disp
3	0.04874214
4	0.17417876
5	0.13586207

In [ ]:
Question 5 : Generate the code to convert the following data frame to wide format.

grp	sex	meanL	sdL	meanR	sdR
A	F	0.225	0.106	0.340	0.085
A	M	0.470	0.325	0.570	0.325
B	F	0.325	0.106	0.400	0.071
B	M	0.547	0.308	0.647	0.274

The result should look like the following display.

	grp	F.meanL	F.meanR	F.sdL	F.sdR	M.meanL	M.meanR	M.sdL	M.sdR
1	A	0.22	0.34	0.11	0.08	0.47	0.57	0.33	0.33
2	B	0.33	0.40	0.11	0.07	0.55	0.65	0.31	0.27

Hint: use `pivot_longer()` in conjunction with `pivot_wider()`.

In [36]:
```
grp<-c("A","A","B","B")
sex<-c("F","M","F","M")
meanL<-c(0.225,0.470,0.325,0.547)
sdL<-c(0.106,0.325,0.106,0.308)
meanR<-c(0.340,0.570,0.400,0.647)
sdR<-c(0.085,0.325,0.071,0.274)
df<-data.frame(grp,sex,meanL,sdL,meanR,sdR)
df
```

grp	sex	meanL	sdL	meanR	sdR
A	F	0.225	0.106	0.340	0.085
A	M	0.470	0.325	0.570	0.325
B	F	0.325	0.106	0.400	0.071
B	M	0.547	0.308	0.647	0.274