Math 189 HW 1

Code

Authors: David Thai, Shir Levin, Stanley Park.

Math 189 Section B

Hide

```
cars <- read.csv("mtcars.csv")
cars
```

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

1-10 of 32 rows | 1-10 of 12 columns   Previous **1** 2 3 4 Next

1. Calculate Sample mean and Variance

Hide

```
cars <- subset(cars, select= -c(model))
#View(cars)
colMeans(cars)
```

```
       mpg        cyl       disp         hp       drat         wt       qsec         vs 
am 
 20.090625   6.187500 230.721875 146.687500   3.596563   3.217250  17.848750   0.437500 
0.406250 
      gear       carb 
  3.687500   2.812500 
```

Hide

```
sapply(cars, var)
```

```
         mpg          cyl          disp            hp          drat            wt          qs
ec           vs
3.632410e+01 3.189516e+00 1.536080e+04 4.700867e+03 2.858814e-01 9.573790e-01 3.193166e+
00 2.540323e-01
          am         gear          carb
2.489919e-01 5.443548e-01 2.608871e+00
```

2. The diagonals of the variance-covariance matrix is equal to the variance of its corresponding variable. Therefore, when comparing the variances calculated from the first problem with the variances found along the diagonals of the variance-covariance matrix, we find that they are in fact the same for each variable.

Furthermore, by looking at the variance-covariance matrix, it is evident that it is symmetric.

Beyond the diagonals of the variance-covariance matrix, the other values (i.e. the covariances between two variables) suggest to us the direction of their correlation; however, it does not tell us the strength of their correlation, but it does tell us something else: if the covariance is greater than zero, less than zero, or equal to zero - we can expect the two variables to be positively correlated, negatively correlated, or uncorrelated, respectively. This idea is supported when we compare the variance-covariance matrix with the correlation matrix. For example, when the ij-th covariance in the variance-covariance matrix has a positive covariance, the ij-th correlation (i.e. the strength of the association between two variables) in the correlation matrix usually is also positive. The same idea also applies when the ij-th covariance in the variance-covariance matrix has a negative covariance. This shows that there is some relationship between the variance-covariance matrix and correlation matrix.

The diagonals of the correlation matrix is one because a variable is directly correlated to itself. Once again, we can see that the correlation matrix is symmetric. Using the correlation matrix, we can also find evidence that supports our intuition. For example, the variables mpg and wt are negatively correlated. This makes sense because we should expect heavier cars to be less gas efficient. Furthermore, the variables cyl and hp are positively correlated, which makes sense because we expect cars with more cylinders in their engines to deliver more power to the car.

Hide

```
cov(cars)
```

```
             mpg        cyl       disp         hp        drat         wt         qs
ec           vs
mpg    36.324103  -9.1723790  -633.09721 -320.732056   2.19506351  -5.1166847   4.509149
19   2.01713710
cyl    -9.172379   3.1895161   199.66028  101.931452  -0.66836694   1.3673710  -1.886854
84  -0.72983871
disp -633.097208 199.6602823 15360.79983 6721.158669 -47.06401915 107.6842040 -96.051681
45 -44.37762097
hp   -320.732056 101.9314516  6721.15867 4700.866935 -16.45110887  44.1926613 -86.770080
65 -24.98790323
drat    2.195064  -0.6683669   -47.06402  -16.451109   0.28588135  -0.3727207   0.087140
73   0.11864919
wt     -5.116685   1.3673710   107.68420   44.192661  -0.37272073   0.9573790  -0.305481
61  -0.27366129
qsec    4.509149  -1.8868548   -96.05168  -86.770081   0.08714073  -0.3054816   3.193166
13   0.67056452
vs      2.017137  -0.7298387   -44.37762  -24.987903   0.11864919  -0.2736613   0.670564
52   0.25403226
am      1.803931  -0.4657258   -36.56401   -8.320565   0.19015121  -0.3381048  -0.204959
68   0.04233871
gear    2.135685  -0.6491935   -50.80262   -6.358871   0.27598790  -0.4210806  -0.280403
23   0.07661290
carb   -5.363105   1.5201613    79.06875   83.036290  -0.07840726   0.6757903  -1.894112
90  -0.46370968
               am        gear        carb
mpg    1.80393145   2.1356855 -5.36310484
cyl   -0.46572581  -0.6491935  1.52016129
disp -36.56401210 -50.8026210 79.06875000
hp    -8.32056452  -6.3588710 83.03629032
drat   0.19015121   0.2759879 -0.07840726
wt    -0.33810484  -0.4210806  0.67579032
qsec  -0.20495968  -0.2804032 -1.89411290
vs     0.04233871   0.0766129 -0.46370968
am     0.24899194   0.2923387  0.04637097
gear   0.29233871   0.5443548  0.32661290
carb   0.04637097   0.3266129  2.60887097
```

Hide

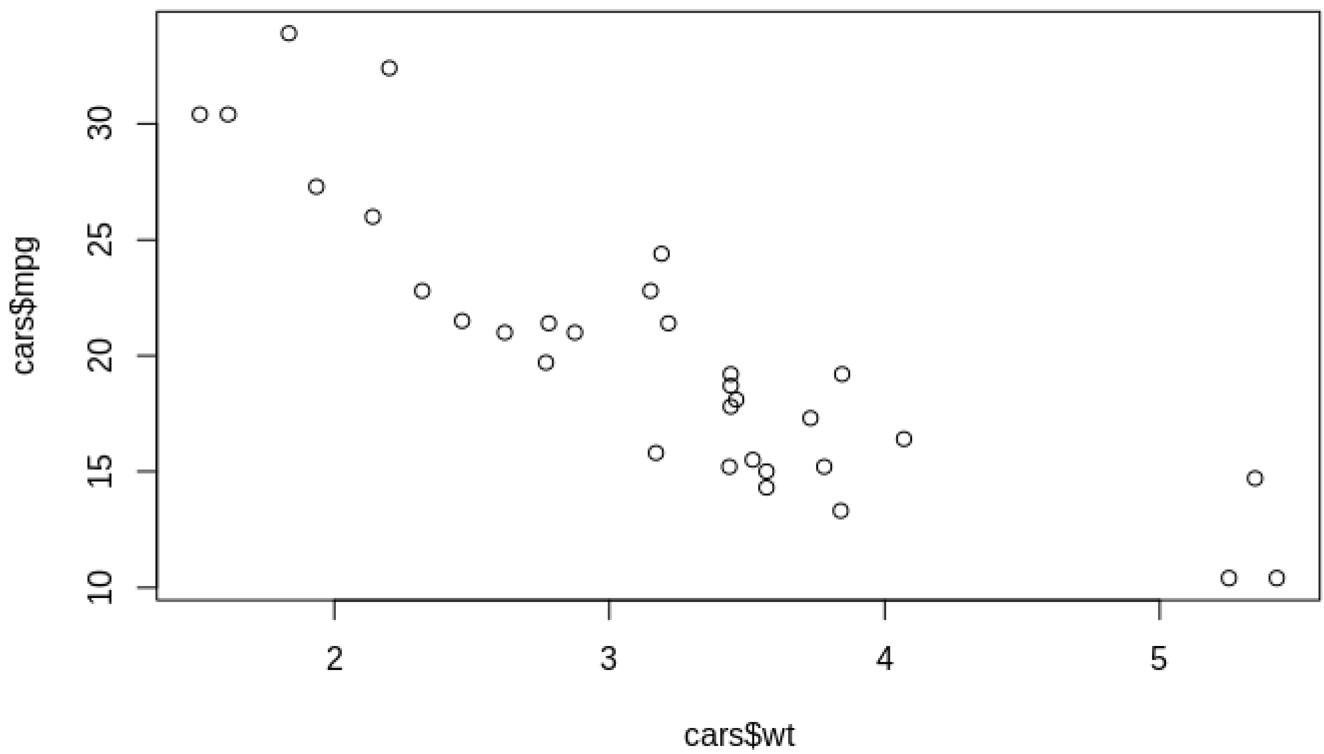
```
cor(cars)
```

```
            mpg        cyl       disp         hp        drat         wt        qsec
vs          am
mpg   1.0000000 -0.8521620 -0.8475514 -0.7761684  0.68117191 -0.8676594  0.41868403  0.6
640389  0.59983243
cyl  -0.8521620  1.0000000  0.9020329  0.8324475 -0.69993811  0.7824958 -0.59124207 -0.8
108118 -0.52260705
disp -0.8475514  0.9020329  1.0000000  0.7909486 -0.71021393  0.8879799 -0.43369788 -0.7
104159 -0.59122704
hp   -0.7761684  0.8324475  0.7909486  1.0000000 -0.44875912  0.6587479 -0.70822339 -0.7
230967 -0.24320426
drat  0.6811719 -0.6999381 -0.7102139 -0.4487591  1.00000000 -0.7124406  0.09120476  0.4
402785  0.71271113
wt   -0.8676594  0.7824958  0.8879799  0.6587479 -0.71244065  1.0000000 -0.17471588 -0.5
549157 -0.69249526
qsec  0.4186840 -0.5912421 -0.4336979 -0.7082234  0.09120476 -0.1747159  1.00000000  0.7
445354 -0.22986086
vs    0.6640389 -0.8108118 -0.7104159 -0.7230967  0.44027846 -0.5549157  0.74453544  1.0
000000  0.16834512
am    0.5998324 -0.5226070 -0.5912270 -0.2432043  0.71271113 -0.6924953 -0.22986086  0.1
683451  1.00000000
gear  0.4802848 -0.4926866 -0.5555692 -0.1257043  0.69961013 -0.5832870 -0.21268223  0.2
060233  0.79405876
carb -0.5509251  0.5269883  0.3949769  0.7498125 -0.09078980  0.4276059 -0.65624923 -0.5
696071  0.05753435
           gear        carb
mpg   0.4802848 -0.55092507
cyl  -0.4926866  0.52698829
disp -0.5555692  0.39497686
hp   -0.1257043  0.74981247
drat  0.6996101 -0.09078980
wt   -0.5832870  0.42760594
qsec -0.2126822 -0.65624923
vs    0.2060233 -0.56960714
am    0.7940588  0.05753435
gear  1.0000000  0.27407284
carb  0.2740728  1.00000000
```

3. Scatterplot between Wt and Mpg

Hide

```
plot(cars$wt, cars$mpg)
```



4. Drawing 3D scatterplot using columns of mtcars

Hide

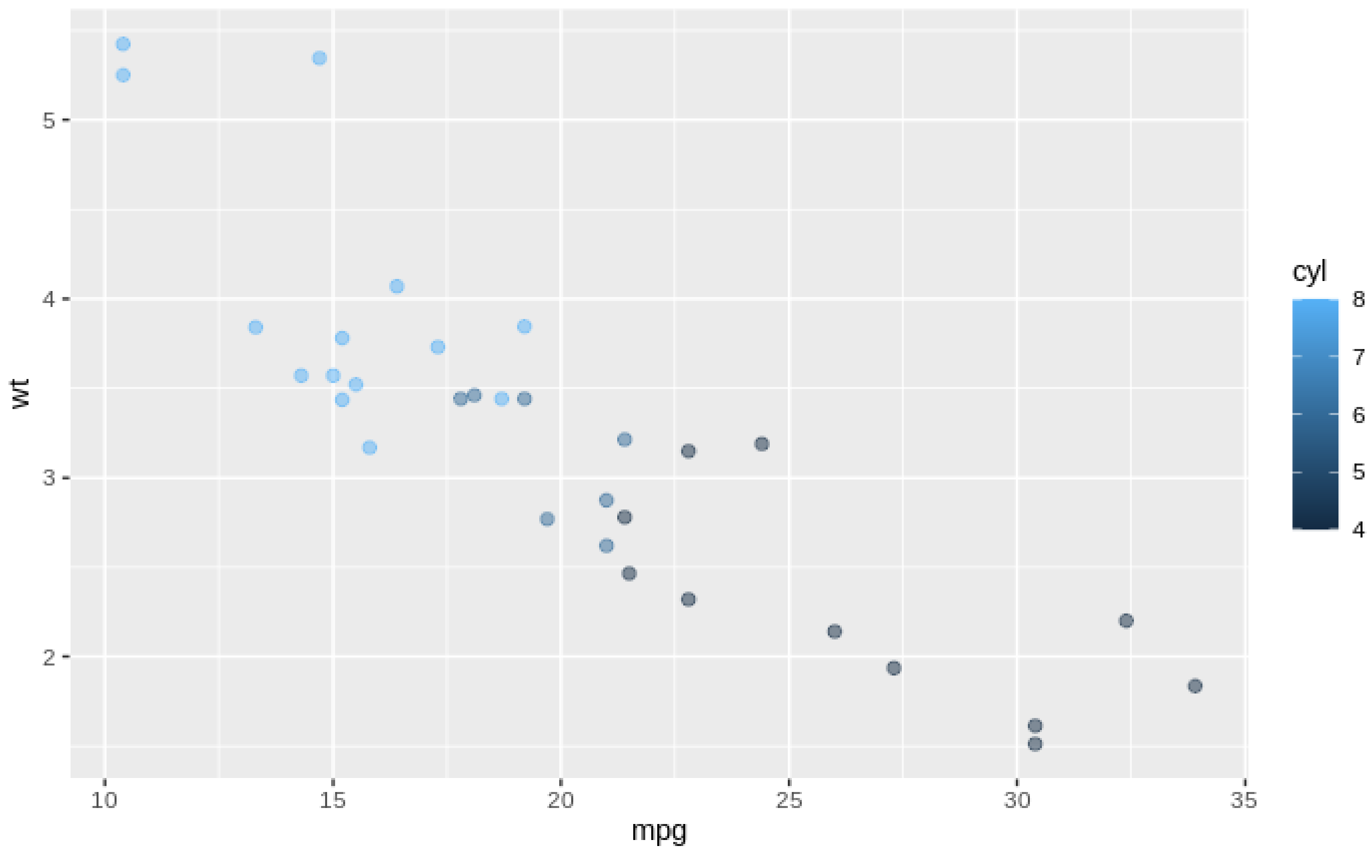
```
#install.packages("tidyverse")
library(tidyverse)
```

```
Registered S3 methods overwritten by 'dbplyr':
  method         from
  print.tbl_lazy
  print.tbl_sql
```

```
── Attaching packages ─────────────────────────────────────────────────
─ tidyverse 1.3.2 ──
✔ ggplot2 3.3.5     ✔ purrr   0.3.4
✔ tibble  3.1.8     ✔ dplyr   1.0.7
✔ tidyr   1.2.1     ✔ stringr 1.4.0
✔ readr   2.1.3     ✔ forcats 0.5.1
── Conflicts ────────────────────────────────────────────────── tidy
verse_conflicts() ──
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
```

Hide

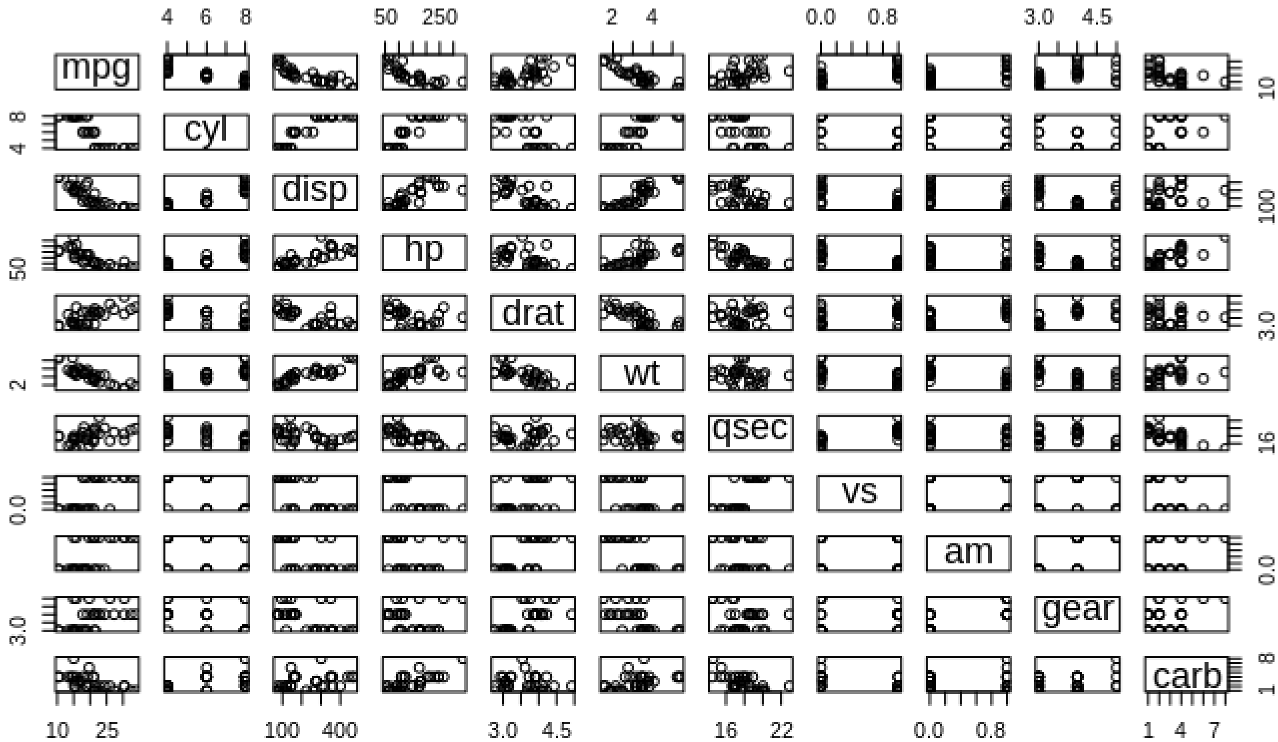
```
#scatterplot3d(x = cars$mpg, y=cars$wt, z=cars$cyl)
#plot_ly(x=cars$mpg, y=cars$wt, z=cars$cyl, type="scatter3d", mode="markers", color=cars
$cyl)
#lot3d(cars$wt, cars$disp, cars$mpg, type = "s", size = 0.75, lit = FALSE)
cars |>
  ggplot(aes(mpg, wt)) + geom_point(alpha=0.5, size=2, aes(color=cyl))
```



5. Pairwise scatterplot

Hide

```
pairs(cars)
```



6. Yes it looks like cylinders has an impact on the relationship between weight and MPG. The lighter the shade of blue of an observation , the more cylinders it has. From the scatterplot in 4, we can clearly see that there is a linear relationship between the shades of blue and points with similar weight and MPG. Cars with heavier weight and lower mpg have lighter shades of blue than those with lighter weights and higher mpg (when considering the same number of cyl per observation).