Math 189 HW 1

Authors: David Thai, Shir Levin, Stanley Park.

Shir coded questions 1-3 and David coded problems 4-6. Stanley wrote the explanations out and helped find functions in R for 1-6.

Math 189 Section B

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

```
##                  model  mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## 1            Mazda RX4 21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
## 2        Mazda RX4 Wag 21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
## 3           Datsun 710 22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
## 4       Hornet 4 Drive 21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
## 5    Hornet Sportabout 18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
## 6              Valiant 18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
## 7           Duster 360 14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
## 8            Merc 240D 24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
## 9             Merc 230 22.8   4 140.8  95 3.92 3.150 22.90  1  0    4    2
## 10            Merc 280 19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
## 11           Merc 280C 17.8   6 167.6 123 3.92 3.440 18.90  1  0    4    4
## 12          Merc 450SE 16.4   8 275.8 180 3.07 4.070 17.40  0  0    3    3
## 13          Merc 450SL 17.3   8 275.8 180 3.07 3.730 17.60  0  0    3    3
## 14         Merc 450SLC 15.2   8 275.8 180 3.07 3.780 18.00  0  0    3    3
## 15  Cadillac Fleetwood 10.4   8 472.0 205 2.93 5.250 17.98  0  0    3    4
## 16 Lincoln Continental 10.4   8 460.0 215 3.00 5.424 17.82  0  0    3    4
## 17   Chrysler Imperial 14.7   8 440.0 230 3.23 5.345 17.42  0  0    3    4
## 18            Fiat 128 32.4   4  78.7  66 4.08 2.200 19.47  1  1    4    1
## 19         Honda Civic 30.4   4  75.7  52 4.93 1.615 18.52  1  1    4    2
## 20      Toyota Corolla 33.9   4  71.1  65 4.22 1.835 19.90  1  1    4    1
## 21       Toyota Corona 21.5   4 120.1  97 3.70 2.465 20.01  1  0    3    1
## 22    Dodge Challenger 15.5   8 318.0 150 2.76 3.520 16.87  0  0    3    2
## 23         AMC Javelin 15.2   8 304.0 150 3.15 3.435 17.30  0  0    3    2
## 24          Camaro Z28 13.3   8 350.0 245 3.73 3.840 15.41  0  0    3    4
## 25    Pontiac Firebird 19.2   8 400.0 175 3.08 3.845 17.05  0  0    3    2
## 26           Fiat X1-9 27.3   4  79.0  66 4.08 1.935 18.90  1  1    4    1
## 27       Porsche 914-2 26.0   4 120.3  91 4.43 2.140 16.70  0  1    5    2
## 28        Lotus Europa 30.4   4  95.1 113 3.77 1.513 16.90  1  1    5    2
## 29      Ford Pantera L 15.8   8 351.0 264 4.22 3.170 14.50  0  1    5    4
## 30        Ferrari Dino 19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
## 31       Maserati Bora 15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
## 32          Volvo 142E 21.4   4 121.0 109 4.11 2.780 18.60  1  1    4    2
```

1. Calculate Sample mean and Variance

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

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

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

```
##          mpg          cyl         disp           hp         drat           wt
## 3.632410e+01 3.189516e+00 1.536080e+04 4.700867e+03 2.858814e-01 9.573790e-01
##         qsec           vs           am         gear         carb
## 3.193166e+00 2.540323e-01 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.

An example where examining the corrleation matrix can reveal relationships/trends in our data is the birth weight example from class. Given a woman's vitamin A intake (or protein or Vitamin C, etc.), see how that correlates with birth weight and create a correlation matrix from all these variables.

```
cov(cars)
```

```
##              mpg         cyl        disp          hp         drat          wt
## mpg    36.324103  -9.1723790  -633.09721 -320.732056   2.19506351  -5.1166847
## cyl    -9.172379   3.1895161   199.66028  101.931452  -0.66836694   1.3673710
## disp -633.097208 199.6602823 15360.79983 6721.158669 -47.06401915 107.6842040
## hp   -320.732056 101.9314516  6721.15867 4700.866935 -16.45110887  44.1926613
## drat    2.195064  -0.6683669   -47.06402  -16.451109   0.28588135  -0.3727207
## wt     -5.116685   1.3673710   107.68420   44.192661  -0.37272073   0.9573790
## qsec    4.509149  -1.8868548   -96.05168  -86.770081   0.08714073  -0.3054816
## vs      2.017137  -0.7298387   -44.37762  -24.987903   0.11864919  -0.2736613
## am      1.803931  -0.4657258   -36.56401   -8.320565   0.19015121  -0.3381048
## gear    2.135685  -0.6491935   -50.80262   -6.358871   0.27598790  -0.4210806
## carb   -5.363105   1.5201613    79.06875   83.036290  -0.07840726   0.6757903
##             qsec          vs          am       gear        carb
## mpg    4.50914919  2.01713710  1.80393145  2.1356855 -5.36310484
## cyl   -1.88685484 -0.72983871 -0.46572581 -0.6491935  1.52016129
```

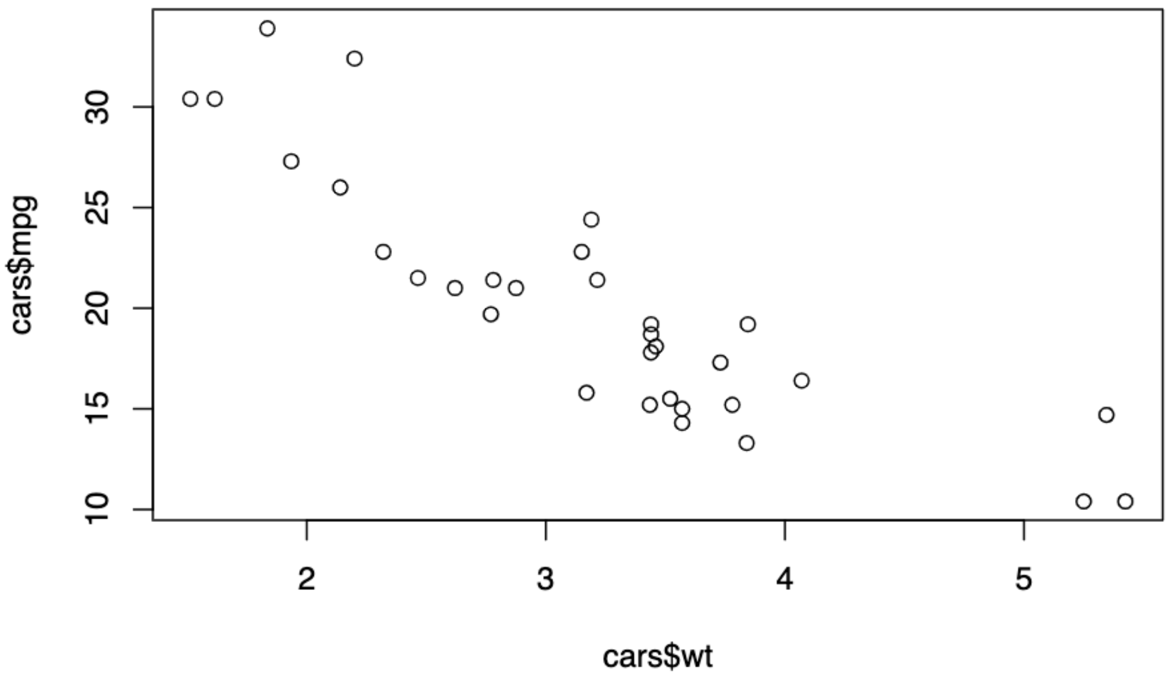
```
## disp -96.05168145 -44.37762097 -36.56401210 -50.8026210 79.06875000
## hp   -86.77008065 -24.98790323  -8.32056452  -6.3588710 83.03629032
## drat   0.08714073   0.11864919   0.19015121   0.2759879 -0.07840726
## wt    -0.30548161  -0.27366129  -0.33810484  -0.4210806  0.67579032
## qsec   3.19316613   0.67056452  -0.20495968  -0.2804032 -1.89411290
## vs     0.67056452   0.25403226   0.04233871   0.0766129 -0.46370968
## am    -0.20495968   0.04233871   0.24899194   0.2923387  0.04637097
## gear  -0.28040323   0.07661290   0.29233871   0.5443548  0.32661290
## carb  -1.89411290  -0.46370968   0.04637097   0.3266129  2.60887097
```

```
cor(cars)
```

```
##             mpg        cyl       disp         hp        drat         wt
## mpg   1.0000000 -0.8521620 -0.8475514 -0.7761684  0.68117191 -0.8676594
## cyl  -0.8521620  1.0000000  0.9020329  0.8324475 -0.69993811  0.7824958
## disp -0.8475514  0.9020329  1.0000000  0.7909486 -0.71021393  0.8879799
## hp   -0.7761684  0.8324475  0.7909486  1.0000000 -0.44875912  0.6587479
## drat  0.6811719 -0.6999381 -0.7102139 -0.4487591  1.00000000 -0.7124406
## wt   -0.8676594  0.7824958  0.8879799  0.6587479 -0.71244065  1.0000000
## qsec  0.4186840 -0.5912421 -0.4336979 -0.7082234  0.09120476 -0.1747159
## vs    0.6640389 -0.8108118 -0.7104159 -0.7230967  0.44027846 -0.5549157
## am    0.5998324 -0.5226070 -0.5912270 -0.2432043  0.71271113 -0.6924953
## gear  0.4802848 -0.4926866 -0.5555692 -0.1257043  0.69961013 -0.5832870
## carb -0.5509251  0.5269883  0.3949769  0.7498125 -0.09078980  0.4276059
##             qsec         vs          am       gear        carb
## mpg   0.41868403  0.6640389  0.59983243  0.4802848 -0.55092507
## cyl  -0.59124207 -0.8108118 -0.52260705 -0.4926866  0.52698829
## disp -0.43369788 -0.7104159 -0.59122704 -0.5555692  0.39497686
## hp   -0.70822339 -0.7230967 -0.24320426 -0.1257043  0.74981247
## drat  0.09120476  0.4402785  0.71271113  0.6996101 -0.09078980
## wt   -0.17471588 -0.5549157 -0.69249526 -0.5832870  0.42760594
## qsec  1.00000000  0.7445354 -0.22986086 -0.2126822 -0.65624923
## vs    0.74453544  1.0000000  0.16834512  0.2060233 -0.56960714
## am   -0.22986086  0.1683451  1.00000000  0.7940588  0.05753435
## gear -0.21268223  0.2060233  0.79405876  1.0000000  0.27407284
## carb -0.65624923 -0.5696071  0.05753435  0.2740728  1.00000000
```

3. Scatterplot between Wt and Mpg

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

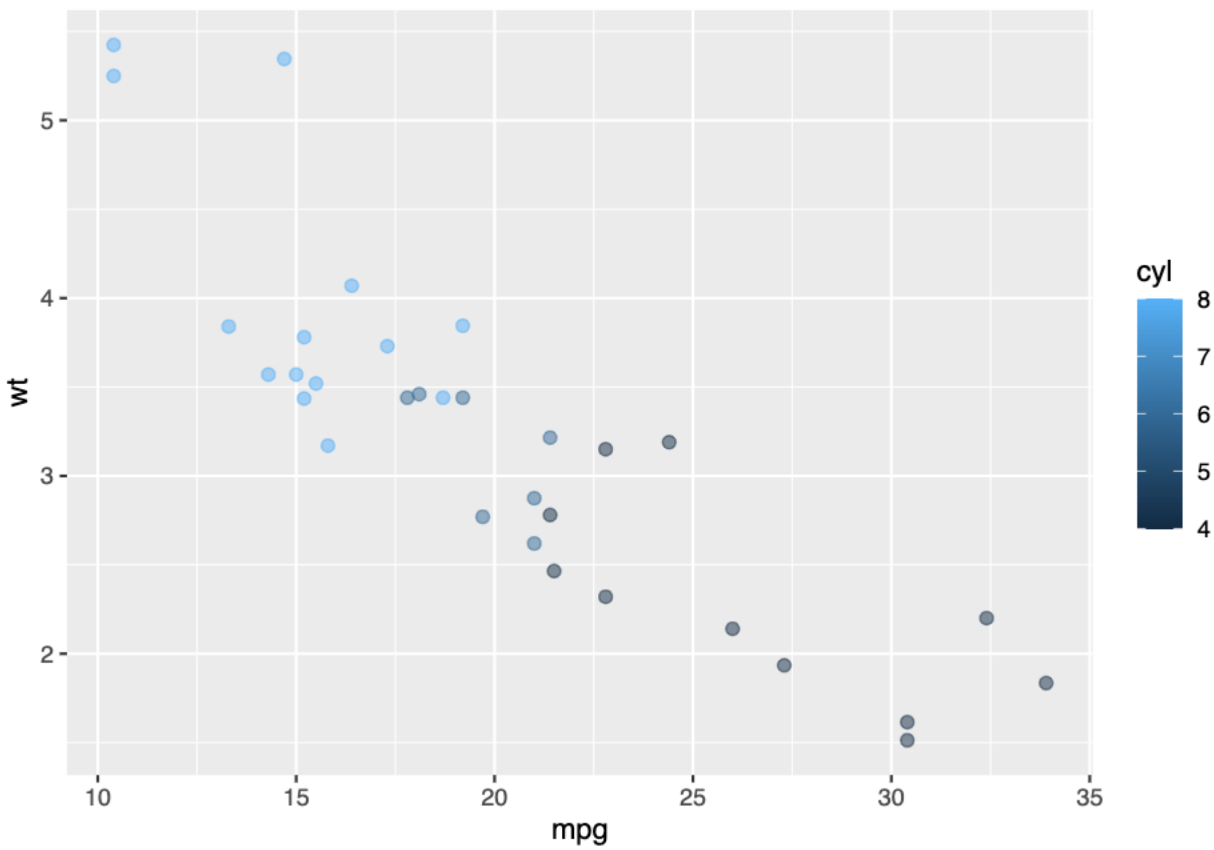


4. Drawing 3D scatterplot using columns of mtcars

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

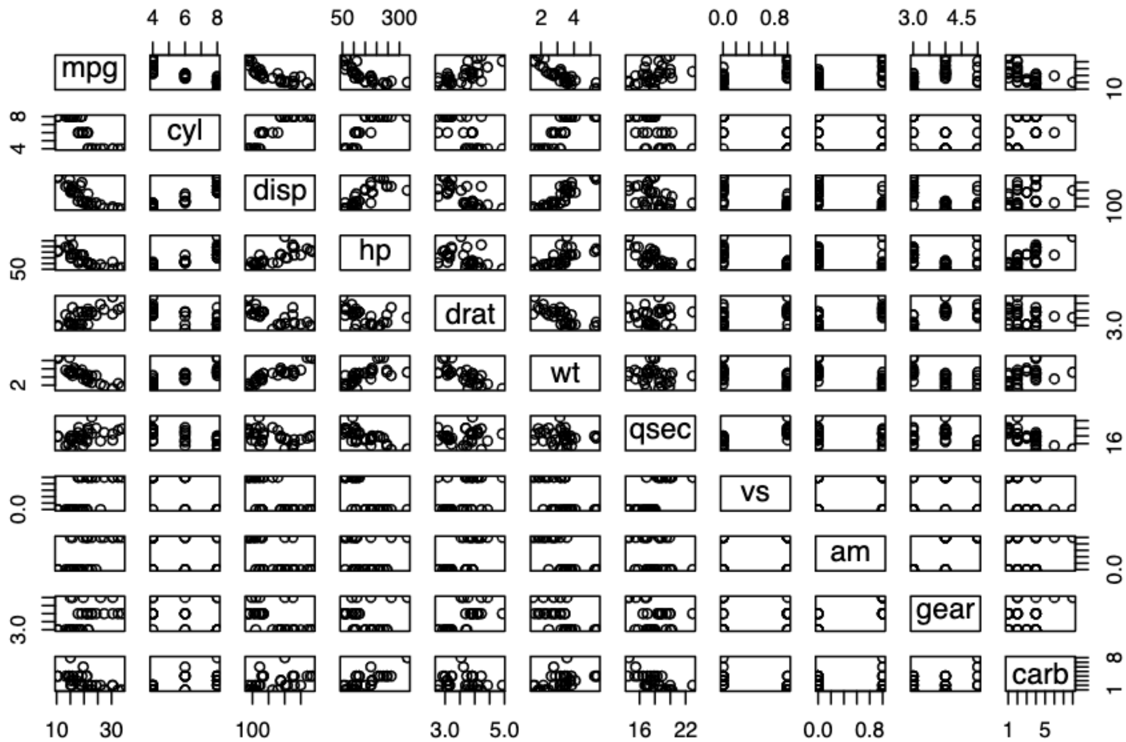
```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --v ggplot2 3.3.5     v
## v tibble  3.1.8     v dplyr   1.0.7
## v tidyr   1.2.1     v stringr 1.4.0
## v readr   2.1.3     v forcats 0.5.1-- Conflicts ------------------------------------------ tidyverse_
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
#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

```
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).