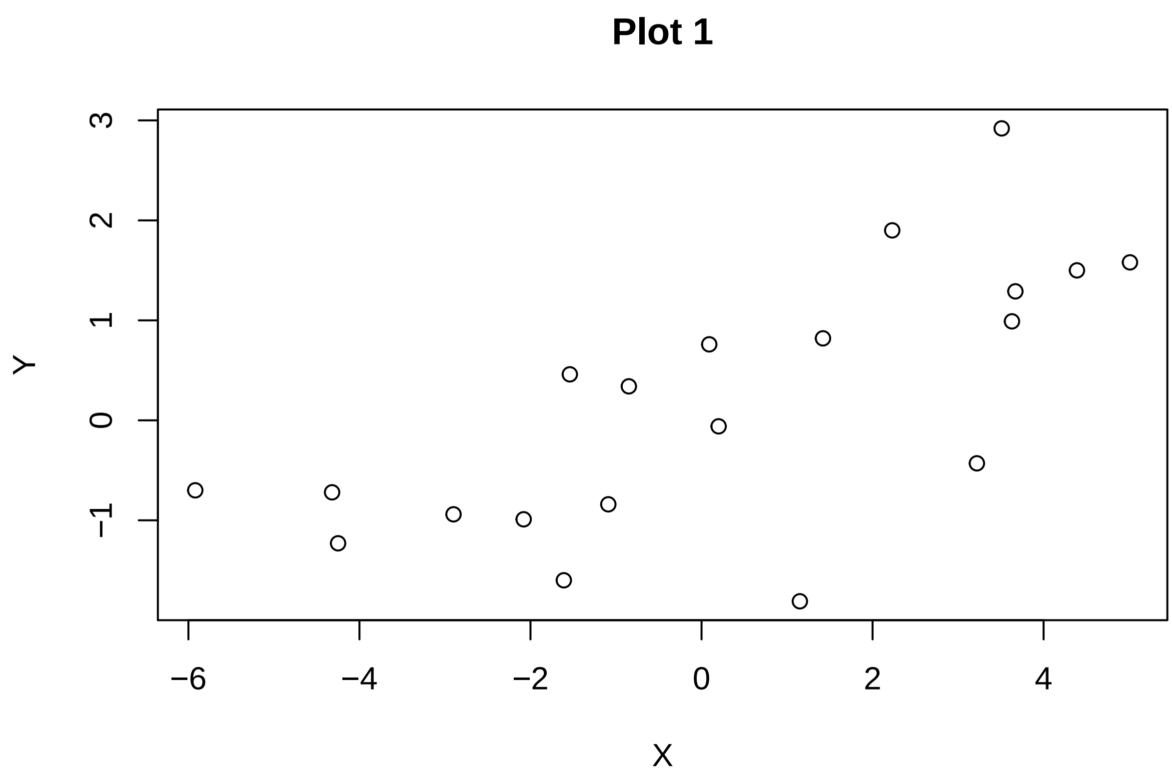
# Homework 1

##Problem 1

#a)

```r
#Fix me, add labels etc.
HW1Q1 <- read_csv("~/Desktop/Multivariate/Homework 1/HW1Q1.csv", show_col_types = FALSE)

p1 <- plot(HW1Q1$X, HW1Q1$Y,
           main = "Plot 1",xlab = "X", ylab= "Y")
```

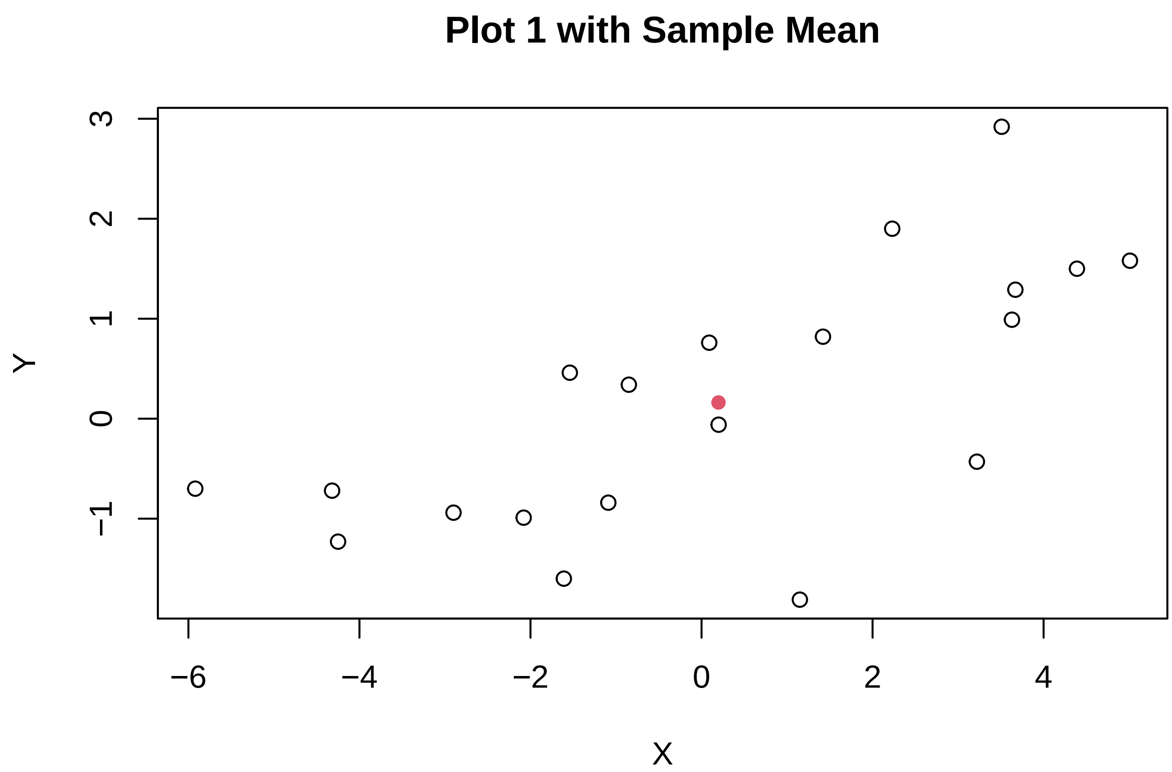



```r
p1
```

```
## NULL
```

#b

```r
mean_q1 <- cbind(mean(HW1Q1$X),  mean(HW1Q1$Y))
p2 <- plot(HW1Q1$X, HW1Q1$Y, main = "Plot 1 with Sample Mean",xlab = "X", ylab= "Y")
points(mean_q1[1],mean_q1[2],pch=16, col=2)
```




#c Find the sample covariance matrix

```r
sampCov_q1 <- cov(HW1Q1[,],use='pairwise.complete.obs')
sampCov_q1
```

```
##           X        Y
## X 10.140227 2.852078
## Y  2.852078 1.668133
```

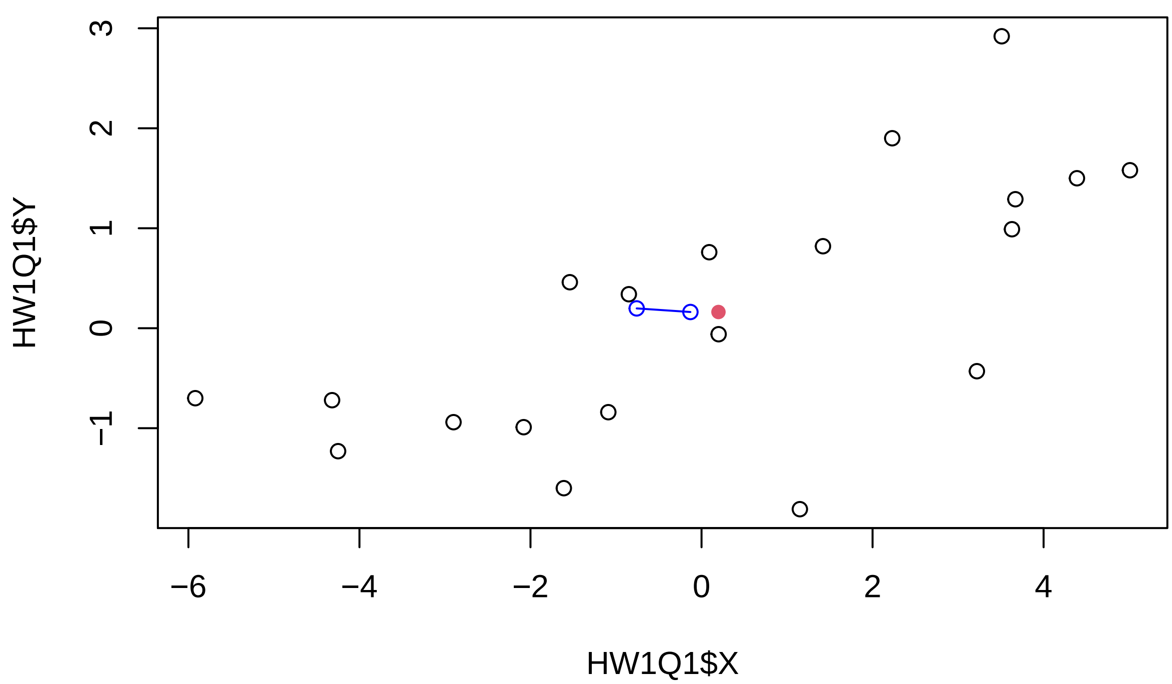
#d

```r
(hw1_eigendeco<- sampCov_q1 %>% eigen())
```

```
## eigen() decomposition
## $values
## [1] 11.0108859  0.7974741
## 
## $vectors
##            [,1]       [,2]
## [1,] -0.9564274  0.2919702
## [2,] -0.2919702 -0.9564274
```

#e

```
point_to <- c(hw1_eigendeco$vectors[1,1]+mean_q1[1], hw1_eigendeco$vectors[2,1]+mean_q1[2])
point_from <- c(mean_q1[1], mean_q1[2])
p3 <- plot(HW1Q1$X, HW1Q1$Y)
points(mean_q1[1],mean_q1[2],pch=16, col=2)
lines(point_to, point_from, type="o", col= "blue")
```




```
#Literally been fighting with this for hours and giving up
```

The eigenvectors should follow the trend of the data in trend and slope. Since we have a positive trend, we should have a positive slope.

##Problem 2 a)

```
CubitData  <- read_csv("~/Desktop/Multivariate/Homework 1/CubitData.csv",show_col_types = FALSE)
(sampmeans_cubit <- cbind(mean(CubitData$height), mean(CubitData$cubit)))
```

```
##           [,1]     [,2]
## [1,] 67.08137 18.07067
```

b)

```r
sampCov_q2 <- cov(CubitData[,],use='pairwise.complete.obs')
sampCov_q2
```

```
##          height     cubit
## height 5.604262 1.4548363
## cubit  1.454836 0.8796708
```

c)

```r
(cubit_eigendeco <- sampCov_q2 %>% eigen())
```

```
## eigen() decomposition
## $values
## [1] 6.0163114 0.4676216
##
## $vectors
##            [,1]       [,2]
## [1,] -0.9621535  0.2725080
## [2,] -0.2725080 -0.9621535
```

d) The eigenvector corresponding to the largest eigenvalue is

$$\begin{bmatrix} -0.9621535 \\ -0.2723080 \end{bmatrix}$$

.

e)

```r
##Plot
```

##Problem 3 Let A be the following matrix:

```r
MatA <- cbind(c(5.125,3.875,2.125,-1.125,0),
              c(3.875,5.125,-1.125,2.125,0),
              c(2.125,-1.125,5.125,3.875,0),
              c(-1.125,2.125,3.875,5.125,0),
              c(0,0,0,0,-3))
MatA
```

```
##        [,1]   [,2]   [,3]   [,4] [,5]
## [1,]  5.125  3.875  2.125 -1.125    0
## [2,]  3.875  5.125 -1.125  2.125    0
## [3,]  2.125 -1.125  5.125  3.875    0
## [4,] -1.125  2.125  3.875  5.125    0
## [5,]  0.000  0.000  0.000  0.000   -3
```

a) Find the eigen decomposition of matrix A:

```r
(eigen_MatA <- MatA %>% eigen())
```

```
## eigen() decomposition
## $values
## [1] 10.0  8.0  4.5 -2.0 -3.0
##
## $vectors
##      [,1] [,2] [,3] [,4] [,5]
## [1,]  0.5 -0.5  0.5  0.5    0
## [2,]  0.5 -0.5 -0.5 -0.5    0
## [3,]  0.5  0.5  0.5 -0.5    0
## [4,]  0.5  0.5 -0.5  0.5    0
## [5,]  0.0  0.0  0.0  0.0    1
```

Thus, we have $\mathbf{A} = \mathbf{V}\lambda\mathbf{V^T}$ where $\mathbf{V} =$

$$\begin{bmatrix} .5 & -.5 & .5 & .5 & 0 \\ 0.5 & -.5 & -.5 & -.5 & 0 \\ 0.5 & 0.5 & 0.5 & -0.5 & 0 \\ 0.5 & 0.5 & -0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and $\lambda =$

$$\begin{bmatrix} 10 & 0 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 & 0 \\ 0 & 0 & 4.5 & 0 & 0 \\ 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 & -3 \end{bmatrix}$$

. b) Is A positive definite?

$\mathbf{A}$ is not positive definite since we do not have positive eigenvalues. If not, give a vector $\mathbf{x}$ for which $\mathbf{x^TAx} < 0$

The simplest $\mathbf{X^TAX}$ would be a 5x1 matrix and we may use either eigenvector associated with with the negative eigenvalues of A to get a $mathbf{x^TAx} <0.

If we take the vector associated with the eigenvalue -2, let

$$\mathbf{x} = \begin{bmatrix} .5 \\ -0.5 \\ -0.5 \\ 0.5 \\ 0 \end{bmatrix}$$

$$\mathbf{x^T} = \begin{bmatrix} 0.5 & -0.5 & -0.5 & 0.5 & 0 \end{bmatrix}$$

```r
x <- cbind(c(.5,-0.5,-0.5,0.5,0))
xT <- rbind(c(.5,-0.5,-0.5,0.5,0))

xTA <- xT %*% MatA
xTA
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]   -1    1    1   -1    0
```

```r
xTA %*% x
```

```
##      [,1]
## [1,]   -2
```

Then,

$$\mathbf{x^T A} = \begin{bmatrix}0.5 & -0.5 & -0.5 & 0.5 & 0\end{bmatrix} \begin{bmatrix}5.125 & 3.875 & 2.125 & -1.125 & 0\\ 3.875 & 5.125 & -1.125 & 2.125 & 0\\ 2.125 & -1.125 & 5.125 & 3.875 & 0\\ 0 & 0 & 0 & 0 & -3\end{bmatrix}$$
$$= \begin{bmatrix}-1 & 1 & 1 & -1 & 0\end{bmatrix}$$

And,

$$\mathbf{x^T A x} = \begin{bmatrix}0.5 & -0.5 & -0.5 & 0.5 & 0\end{bmatrix} \begin{bmatrix}5.125 & 3.875 & 2.125 & -1.125 & 0\\ 3.875 & 5.125 & -1.125 & 2.125 & 0\\ 2.125 & -1.125 & 5.125 & 3.875 & 0\\ 0 & 0 & 0 & 0 & -3\end{bmatrix} \begin{bmatrix}.5\\ -0.5\\ -0.5\\ 0.5\\ 0\end{bmatrix}$$
$$= \begin{bmatrix}-1 & 1 & 1 & -1 & 0\end{bmatrix} \begin{bmatrix}.5\\ -0.5\\ -0.5\\ 0.5\\ 0\end{bmatrix}$$
$$= -2$$

c) Let a vector $\mathbf{x}$ be given by

$$\mathbf{x} = 4\mathbf{v_1} + 2\mathbf{v_5}$$

Where $\mathbf{v_1}$ is the eigenvector corresponding to the largest eigenvalue $\lambda_1$ of $\mathbf{A}$, and $\mathbf{v_5}$ is the eigenvector corresponding to the smallest eigenvalue $\lambda_5$ of $\mathbf{A}$. In terms of $\mathbf{v_1}, \mathbf{v_5}, \lambda_1$, and $\lambda_5$, what is $\mathbf{Ax}$?

```
#v1 is eigenvector for the largest eigenvalue of A
v1 <- eigen_MatA$vectors[,1]

#v5 eigenvector for smallest evalue of A
v5 <- eigen_MatA$vectors[,5]

x <- 4*v1 + 2*v5

MatA %*% x
```

```
##      [,1]
## [1,]   20
## [2,]   20
## [3,]   20
## [4,]   20
## [5,]   -6
```

$$\begin{aligned}\mathbf{Ax} &= \mathbf{A} * \mathbf{x} = \mathbf{A} * (4v_1 + 2v_5)\\ &= \mathbf{A} * 2(2v_1 + v_5)\\ &= 2\mathbf{A} * (2v_1 + v_5)\\ &= 2\begin{bmatrix}\lambda_1\\ \lambda_2\end{bmatrix}(2v_1 + v_5)\\ &= 2(2\lambda_1 v_1 + \lambda_5 v_5)\end{aligned}$$

.

#Problem 4 The Iris Flower Data Set is available as the file 'IrisData.csv' on Canvas. Download the file and read it into R using read.csv().

a)

```
IrisData <- read_csv("~/Desktop/Multivariate/Homework 1/IrisData.csv", show_col_types = FALSE)
Iris_means <- cbind(mean(IrisData$Sepal.Length), mean(IrisData$Sepal.Width), mean(IrisData$Petal.Length)
Iris_means
```

```
##          [,1]     [,2]  [,3]     [,4] [,5]
## [1,] 5.843333 3.057333 3.758 1.199333    2
```

b)

```
means_by_species <- IrisData %>% group_by(Type) %>%
                                summarise(sep_length_mean <- mean(Sepal.Length),
                                sep_width_mean <- mean(Sepal.Width),
                                pet_length_mean <- mean(Petal.Length),
                                pet_width_mean <- mean(Petal.Width)
                                )
means_by_species
```

```
## # A tibble: 3 x 5
##    Type `sep_length_mean ~ `sep_width_mean ~ `pet_length_mean~ `pet_width_mean ~
##   <dbl>              <dbl>             <dbl>             <dbl>             <dbl>
## 1     1               5.01              3.43              1.46             0.246
## 2     2               5.94              2.77              4.26             1.33
## 3     3               6.59              2.97              5.55             2.03
```

c) Find the sample correlation matrix

```
cor_matrix_overall <- IrisData %>%
                        select(-Type) %>%
                        cor()
cor_matrix_overall
```

```
##              Sepal.Length Sepal.Width Petal.Length Petal.Width
## Sepal.Length    1.0000000  -0.1175698    0.8717538   0.8179411
## Sepal.Width    -0.1175698   1.0000000   -0.4284401  -0.3661259
## Petal.Length    0.8717538  -0.4284401    1.0000000   0.9628654
## Petal.Width     0.8179411  -0.3661259    0.9628654   1.0000000
```

The two variables that are most highly correlated are Petal.Length and Petal.Width which have a correlation coefficient of 0.9628654.

d)

```
findCor <- function(i) {
       IrisData%>%
       filter(Type == i) %>%
      select(-Type) %>%
      cor()
}
```

```r
type1_cor <- findCor(1)

type1_cor
```

```
##              Sepal.Length Sepal.Width Petal.Length Petal.Width
## Sepal.Length    1.0000000   0.7425467    0.2671758   0.2780984
## Sepal.Width     0.7425467   1.0000000    0.1777000   0.2327520
## Petal.Length    0.2671758   0.1777000    1.0000000   0.3316300
## Petal.Width     0.2780984   0.2327520    0.3316300   1.0000000
```

The highest correlated variables for type 1 are between Sepal.Width and Sepal.Length with correlation coefficient of 0.7425.

```r
type2_cor <- findCor(2)

type2_cor
```

```
##              Sepal.Length Sepal.Width Petal.Length Petal.Width
## Sepal.Length    1.0000000   0.5259107    0.7540490   0.5464611
## Sepal.Width     0.5259107   1.0000000    0.5605221   0.6639987
## Petal.Length    0.7540490   0.5605221    1.0000000   0.7866681
## Petal.Width     0.5464611   0.6639987    0.7866681   1.0000000
```

The highest correlated variables for type 2 are Petal.Length and Petal.Width with a correlation coefficient of 0.7867.

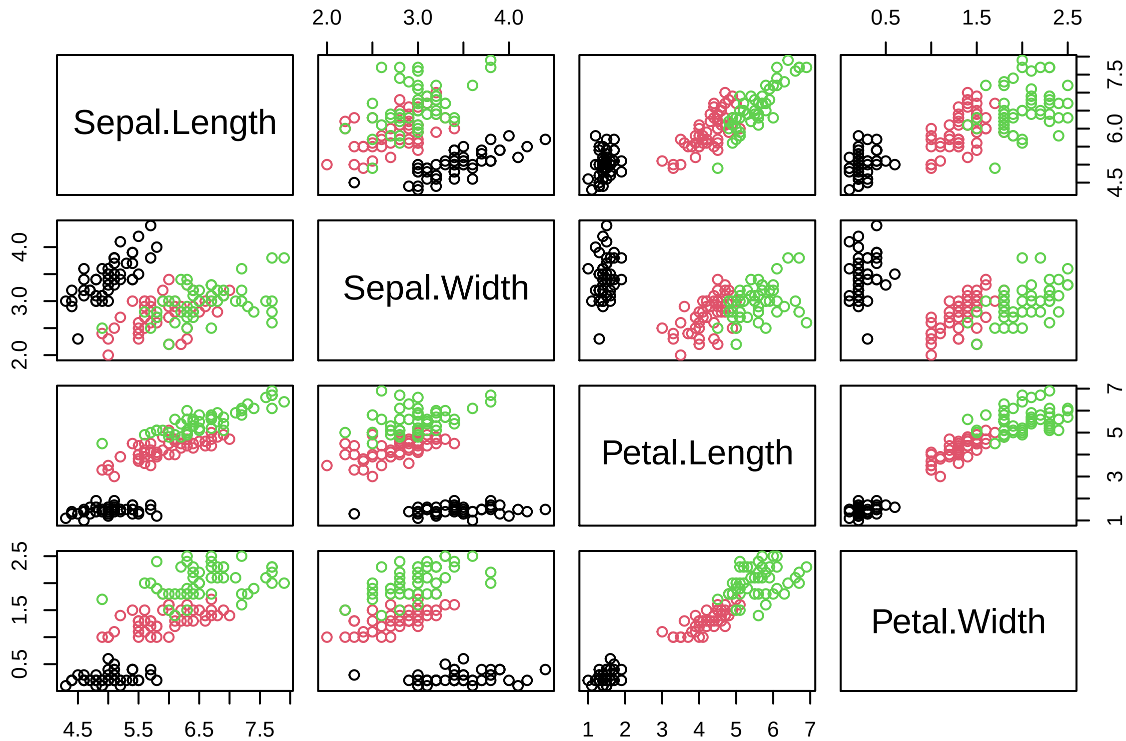
```r
type3_cor <- findCor(3)

type3_cor
```

```
##              Sepal.Length Sepal.Width Petal.Length Petal.Width
## Sepal.Length    1.0000000   0.4572278    0.8642247   0.2811077
## Sepal.Width     0.4572278   1.0000000    0.4010446   0.5377280
## Petal.Length    0.8642247   0.4010446    1.0000000   0.3221082
## Petal.Width     0.2811077   0.5377280    0.3221082   1.0000000
```

Once again the highest correlated variable of individual types do not match the highest correlated variables of the overall dataset. For type 3, the highest correlated variables are Petal.Length and Sepal.Length with correlation coefficient 0.864.

e)

```r
pairs(IrisData[,1:4],col = IrisData$Type)
```




Species 2 and 3 tend to follow a similar pattern, even often having some overlap, whereas Type I is distinct from the other two. Petal Width and Petal Length have distinct groupings where but sepal width and length have less of a clear pattern.

#Problem 5 Let $\mathbf{A}$ be any (n x p) matrix, for arbitrary n and p, and let $\mathbf{B}$ be the product matrix $\mathbf{B} = \mathbf{A^T A}$

a) Show that $\mathbf{B}$ is a symmetric matrix.

Matrix $\mathbf{B}$ is symmetric if $\mathbf{B^T} = \mathbf{B}$. Let $\mathbf{B} = \mathbf{A^T A}$ for some n and p. Then,

$$\begin{aligned}\mathbf{B^T} &= (\mathbf{A^T A})^\mathbf{T} \\ &= \mathbf{A^T}(\mathbf{A^T})^\mathbf{T} \\ &= \mathbf{A^T A} \\ &= \mathbf{B}.\end{aligned}$$

Thus, $mathbf{B}$ is symmetric.

b) Show that $\mathbf{B}$ is a positive semi-definite matrix: $\mathbf{x^T B x} \geq 0$ for all $\mathbf{x} \in R^p$. That is for all p-dimensional vectors $\mathbf{x}$.

Since $\mathbf{B} = \mathbf{A^T A}$,

$$\begin{aligned}\mathbf{x^T B x} &= \mathbf{x^T}(\mathbf{A^T A})\mathbf{x} \\ &= (\mathbf{Ax})^\mathbf{T}\mathbf{Ax} \\ &= ||\mathbf{Ax}||\end{aligned}$$

Since $||\mathbf{Ax}||$ is always positive, $\mathbf{B}$ is a semi-definite matrix.