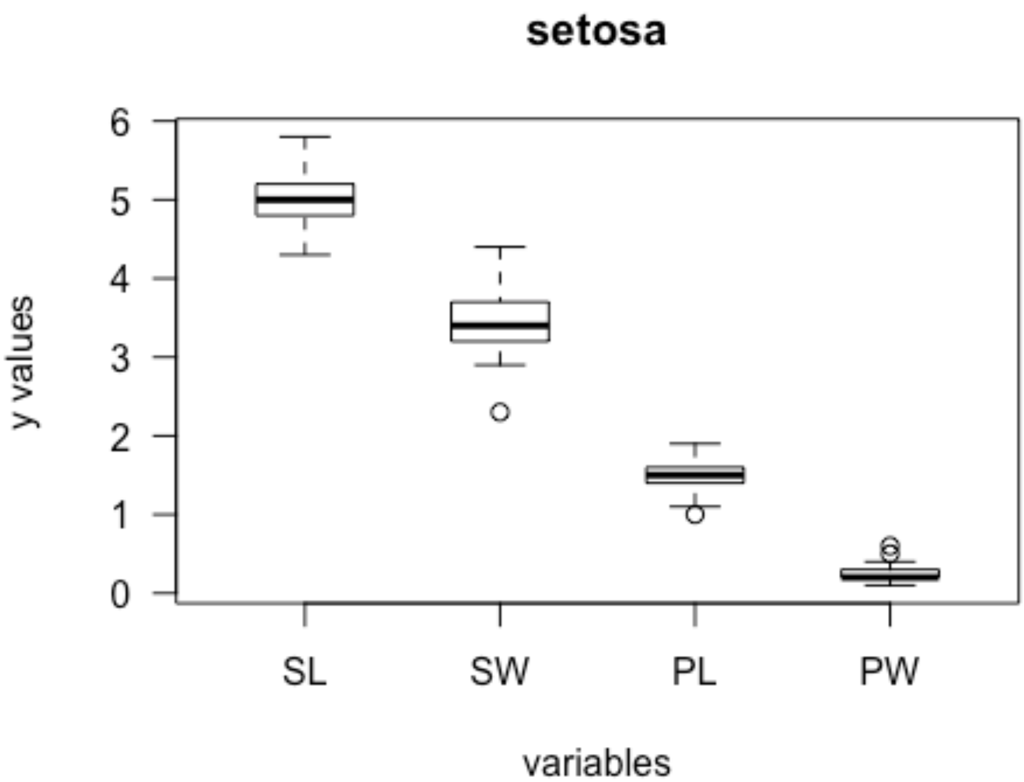
Question 1d

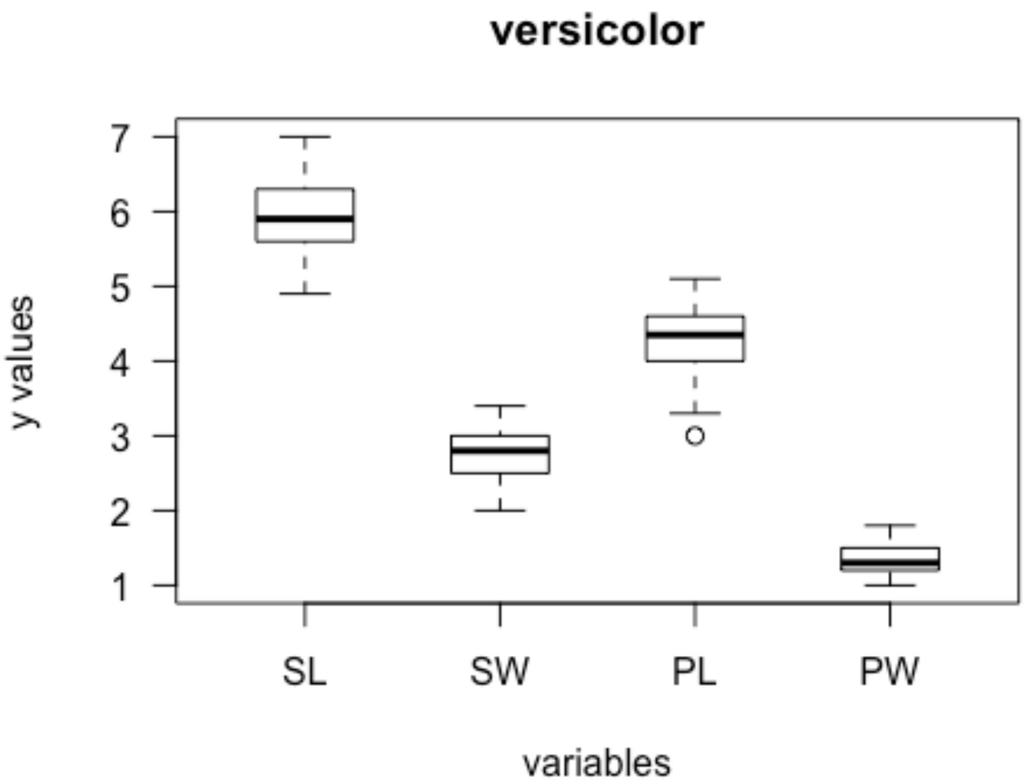
```r
#Find the length of x;
sqrt(sum(5*5+1*1+3*3))

## [1] 5.91608

#Find the length of y;
sqrt(sum((-1)*(-1)+ 3*3 +1*1))

## [1] 3.316625

#Find xTy;
c1 <- c(5,1,3)
x <- matrix(c1,3,1)
c2 <- c(-1,3,1)
y <- matrix(c2,3,1)
t(x) %*% y

##      [,1]
## [1,]    1
```

Question 3a

```r
#Extract the setosa species from iris data;
attach(iris)
SE <- iris[Species == "setosa", 1:4]
#Build the boxplot of setosa;
boxplot(SE, boxwex = 0.5, las=1, names=c("SL","SW","PL","PW"), main="setosa",

        xlab="variables",ylab="y values")
```



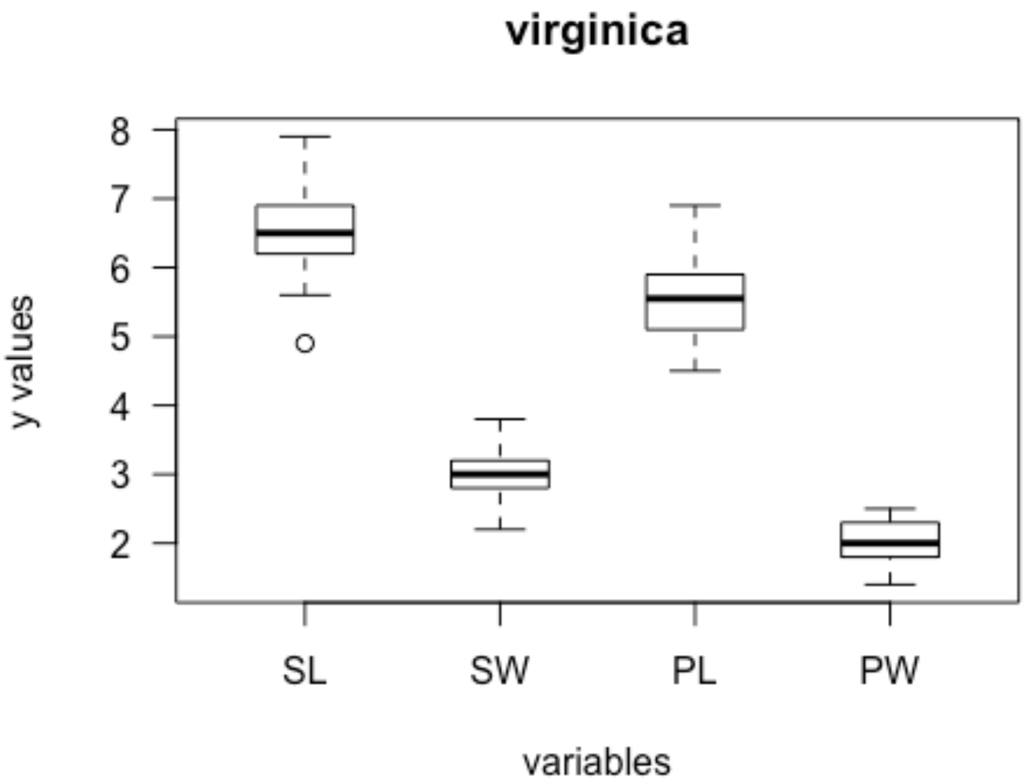

```r
#Extract the versicolor species from iris data;
VER <- iris[Species == "versicolor", 1:4]
#Build the boxplot of versicolor;
boxplot(VER, boxwex = 0.5, las=1, names=c("SL","SW","PL","PW"), main="versicolor",
        xlab="variables",ylab="y values")
```



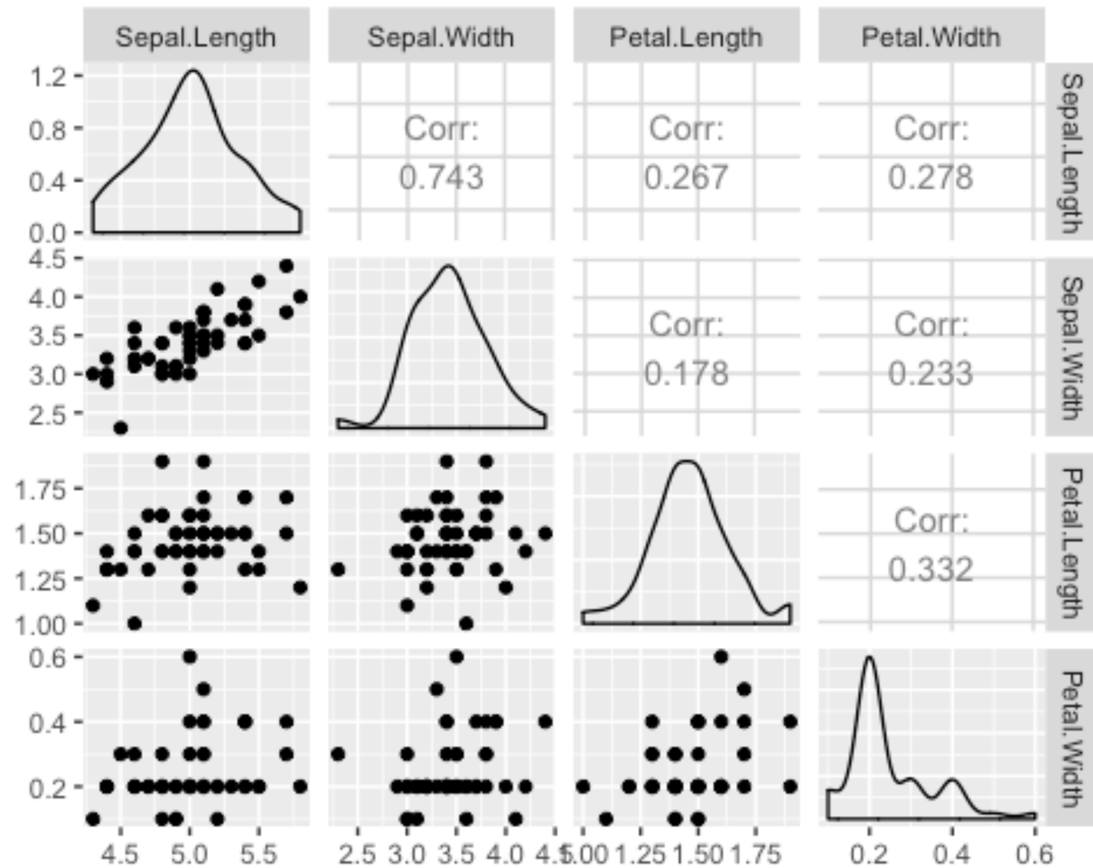

```
#Extract the virginica species from iris data;
VIR <- iris[Species == "virginica", 1:4]
#Build the boxplot of virginica;
boxplot(VIR, boxwex = 0.5, las=1, names=c("SL","SW","PL","PW"), main="virgini
ca",
        xlab="variables",ylab="y values")
```

Question 3b

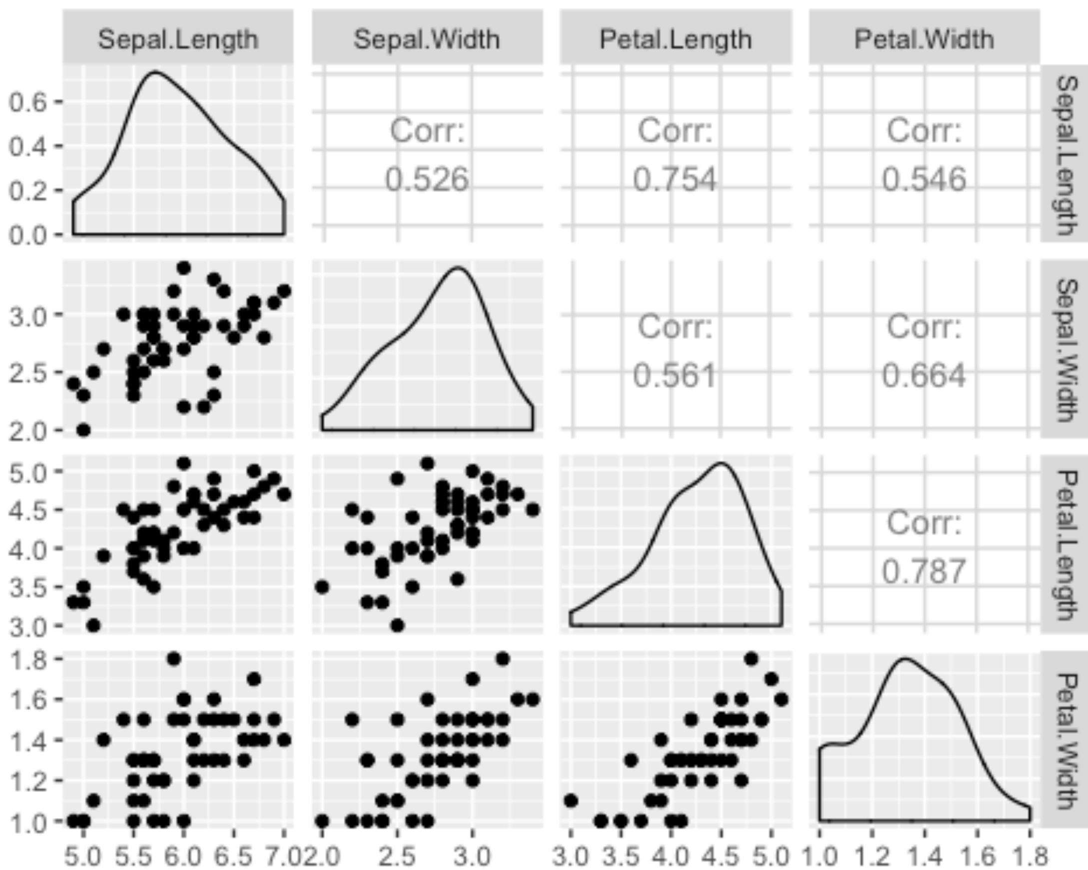
```
#load the library of ggplot2 and GGally;
library(ggplot2)
```



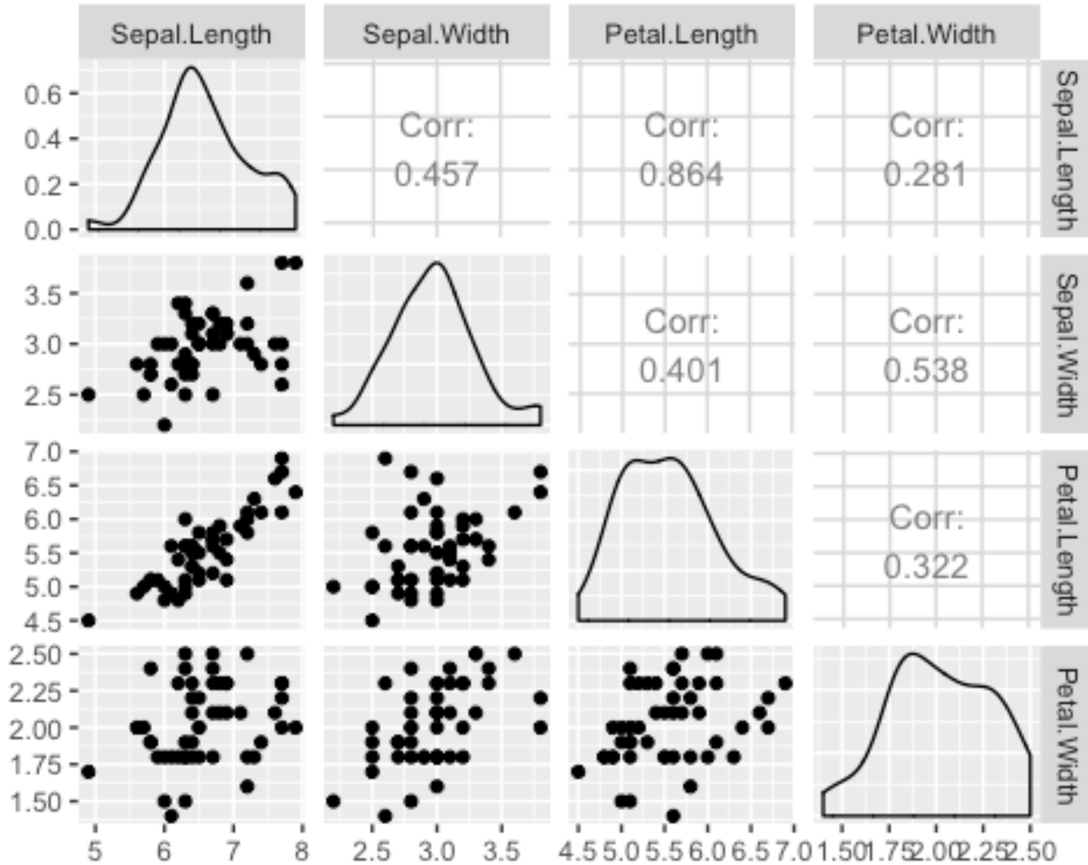

```r
library(GGally)
#Make the pairs-plot of setosa;
ggpairs(SE)
```




```
#Make the pairs-plot of versicolor;
ggpairs(VER)
```




```r
#Make the pairs-plot of virginica;
ggpairs(VIR)
```



Question 3c

```
#Estimate the μ of setosa;
SE_means <- colMeans(SE)
SE_means

## Sepal.Length  Sepal.Width Petal.Length  Petal.Width 
##        5.006        3.428        1.462        0.246

#Compute the sample covariance matrix;
SE_S <- cov(SE)
round(SE_S, 4)

##              Sepal.Length Sepal.Width Petal.Length Petal.Width
## Sepal.Length       0.1242      0.0992       0.0164      0.0103
## Sepal.Width        0.0992      0.1437       0.0117      0.0093
## Petal.Length       0.0164      0.0117       0.0302      0.0061
## Petal.Width        0.0103      0.0093       0.0061      0.0111

#Compute the sample correlation matrix;
cor(SE)

##              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

#Sample size of setosa flowers;
SE_n <- nrow(SE)
#Estimate covariance of setosa;
round(SE_S/SE_n, 5)

##                Sepal.Length Sepal.Width Petal.Length Petal.Width
## Sepal.Length        0.00248     0.00198      0.00033     0.00021
## Sepal.Width         0.00198     0.00287      0.00023     0.00019
## Petal.Length        0.00033     0.00023      0.00060     0.00012
## Petal.Width         0.00021     0.00019      0.00012     0.00022

#Estimate the μ of versicolor;
VER_means <- colMeans(VER)
VER_means

## Sepal.Length  Sepal.Width Petal.Length  Petal.Width
##        5.936        2.770        4.260        1.326

#Compute the sample covariance matrix;
VER_S <- cov(VER)
round(VER_S, 4)

##                Sepal.Length Sepal.Width Petal.Length Petal.Width
## Sepal.Length         0.2664      0.0852       0.1829      0.0558
## Sepal.Width          0.0852      0.0985       0.0827      0.0412
## Petal.Length         0.1829      0.0827       0.2208      0.0731
## Petal.Width          0.0558      0.0412       0.0731      0.0391

#Compute the sample correlation matrix;
cor(VER)

##                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

#Sample size of versicolor flowers;
VER_n <- nrow(VER)
#Estimate the covariance of versicolor;
round(VER_S/VER_n, 5)

##                Sepal.Length Sepal.Width Petal.Length Petal.Width
## Sepal.Length        0.00533     0.00170      0.00366     0.00112
## Sepal.Width         0.00170     0.00197      0.00165     0.00082
## Petal.Length        0.00366     0.00165      0.00442     0.00146
## Petal.Width         0.00112     0.00082      0.00146     0.00078
```

```r
#Estimate the μ of virginica;
VIR_means <- colMeans(VIR)
VIR_means
```

```
## Sepal.Length  Sepal.Width Petal.Length  Petal.Width 
##        6.588        2.974        5.552        2.026
```

```r
#Compute the sample covariance matrix;
VIR_S <- cov(VIR)
round(VIR_S,4)
```

```
##              Sepal.Length Sepal.Width Petal.Length Petal.Width
## Sepal.Length       0.4043      0.0938       0.3033      0.0491
## Sepal.Width        0.0938      0.1040       0.0714      0.0476
## Petal.Length       0.3033      0.0714       0.3046      0.0488
## Petal.Width        0.0491      0.0476       0.0488      0.0754
```

```r
#Compute the sample correlation matrix;
cor(VIR)
```

```
##              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
```

```r
#Sample size of virginica flowers;
VIR_n <- nrow(VIR)
#Estimate the covariance of virginica;
round(VIR_S/VIR_n, 5)
```

```
##              Sepal.Length Sepal.Width Petal.Length Petal.Width
## Sepal.Length      0.00809     0.00188      0.00607     0.00098
## Sepal.Width       0.00188     0.00208      0.00143     0.00095
## Petal.Length      0.00607     0.00143      0.00609     0.00098
## Petal.Width       0.00098     0.00095      0.00098     0.00151
```

Question 3d

Yes, for all three species, there are positive correlation between any variables (length, width of Sepal or Petal). But when we compare the pattern of them, the virginica has more corelated between variables, the setosa has the least correlation between variables.

Question 4a

Population: population consist of all skulls with epoch c4000BC;
Parameter: true mean of mb, bh, bl, nh of all skulls with epoch c4000BC;
Sample: sample consists of 30 epoch c4000BC for which data collected;
Statistic: estimate the mean of mb, bh, bl, nh of all skulls with epoch c4000BC by sample mean.

Question 4b

```r
#Extract all data of c4000BC;
library(HSAUR3)

## Loading required package: tools

attach(skulls)
epoch_c4000 <- skulls[epoch == "c4000BC", 2:5]
#Estimate of the true mean of all skulls with epoch c4000BC;
epoch_c4000_mean <- colMeans(epoch_c4000)
epoch_c4000_mean

##        mb        bh        bl        nh 
## 131.36667 133.60000  99.16667  50.53333
```

Question 4c

```r
#Sample variance-covariance was used to estimate
S_eopch_c4000 <- cov(epoch_c4000)
S_eopch_c4000

##           mb         bh         bl         nh
## mb 26.309195  4.1517241  0.4540230  7.2459770
## bh  4.151724 19.9724138 -0.7931034  0.3931034
## bl  0.454023 -0.7931034 34.6264368 -1.9195402
## nh  7.245977  0.3931034 -1.9195402  7.6367816
```

Question 4d

```r
#Sample size of c4000BC;
n_c4000 <- nrow(epoch_c4000)
#Estimate the population variance covariance of the population mean;
E <- S_eopch_c4000/n_c4000
E

##           mb          bh          bl          nh
## mb 0.8769732  0.13839080  0.01513410  0.24153257
## bh 0.1383908  0.66574713 -0.02643678  0.01310345
## bl 0.0151341 -0.02643678  1.15421456 -0.06398467
## nh 0.2415326  0.01310345 -0.06398467  0.25455939
```

Question 4e

```r
#According to the calculation, the matrix of A can be written as;
A <- matrix(c(1,0,0,-1,0,1,0,-1), nrow=2, ncol=4, byrow=TRUE)
A

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

#The covariacne of the estimator is:
A %*% E %*% t(A)
```

```
##            [,1]      [,2]
## [1,] 0.6484674 0.1383142
## [2,] 0.1383142 0.8940996
```

Question 5a

The basic difference between multivariate and longitudinal data is that the order of the repeated measurements is essential in the analysis of longitudinal data, whereas permuting the order of the variables in multivariate analysis and yields same results.

Question 5b

Because we want to know how variable and quantify how variablily between two variables.

Question 5c

No, that only proves X and Y doesn't have linear realtionship, they may have non-liner relationship.

Question 5d

No, as we know, the longitude data should be meausred as the same unit at different time spots, however in problem 4 they did at the different spots, but the testing targets are different from each time tests.