CA1

ESNCHR001

Continuous Assessment 1

```
ca1 <- read.csv("CA1.csv")
X_total <- as.matrix(ca1)
```

Question 1

Compute and report the sample mean vectors for each of the five time periods.

```
mean_matrix <- matrix(nrow=5, ncol=5)
for(i in 1:5){
  X_group <- X_total[X_total[,5] == i,]
  mean_matrix[i,] <- (1/nrow(X_group))*t(X_group)%*%matrix(1,nrow=nrow(X_group),ncol=1)
}
colnames(mean_matrix) <- c("MaxBreadth", "BaseHeight", "BasLength", "NasHeight",
                           "TimePeriod")
rownames(mean_matrix) <- c("4000 B.C", "3300 B.C.", "1850 B.C.", " 200 B.C.",
                           " 150 A.D.")
(mean_matrix <- t(mean_matrix))
```

```
             4000 B.C 3300 B.C. 1850 B.C.  200 B.C.  150 A.D.
MaxBreadth  131.36667 132.36667 134.46667 135.50000 136.16667
BaseHeight  133.60000 132.70000 133.80000 132.30000 130.33333
BasLength    99.16667  99.06667  96.03333  94.53333  93.50000
NasHeight    50.53333  50.23333  50.56667  51.96667  51.36667
TimePeriod    1.00000   2.00000   3.00000   4.00000   5.00000
```

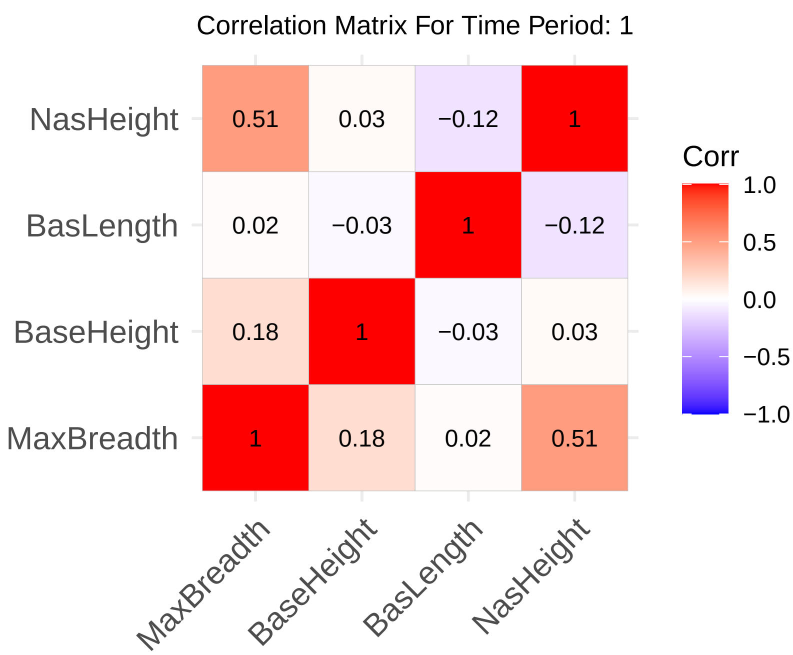
Question 2

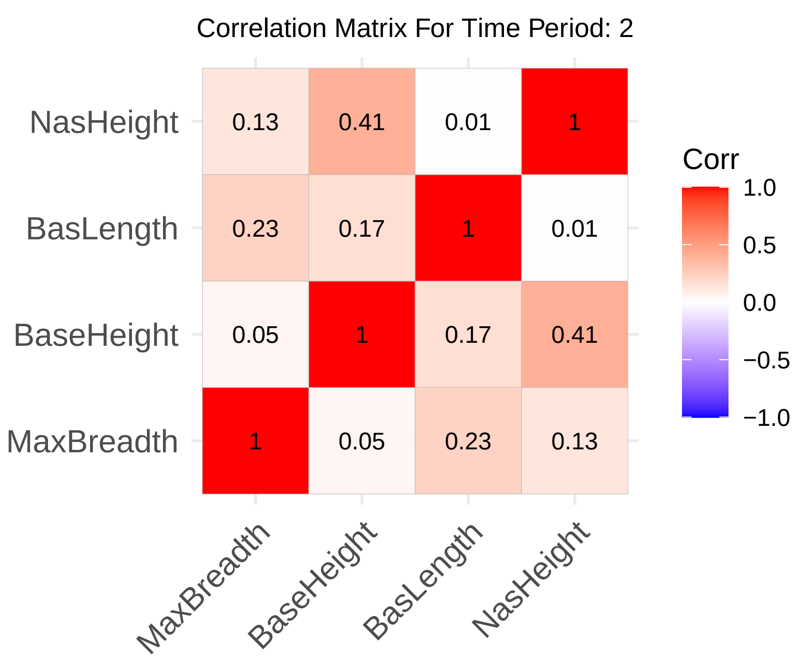
Provide a heat map of the correlation matrix for each time period

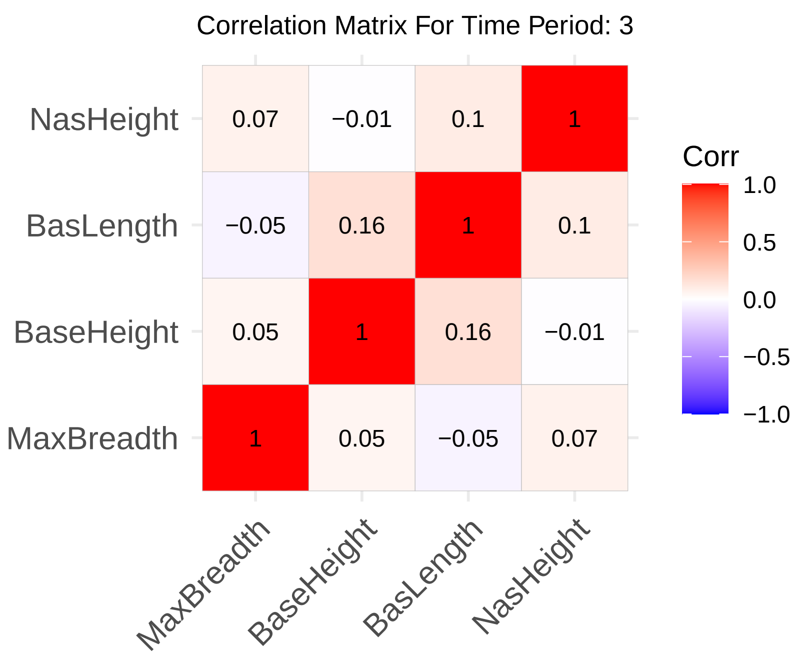
```
library(ggcorrplot)
```

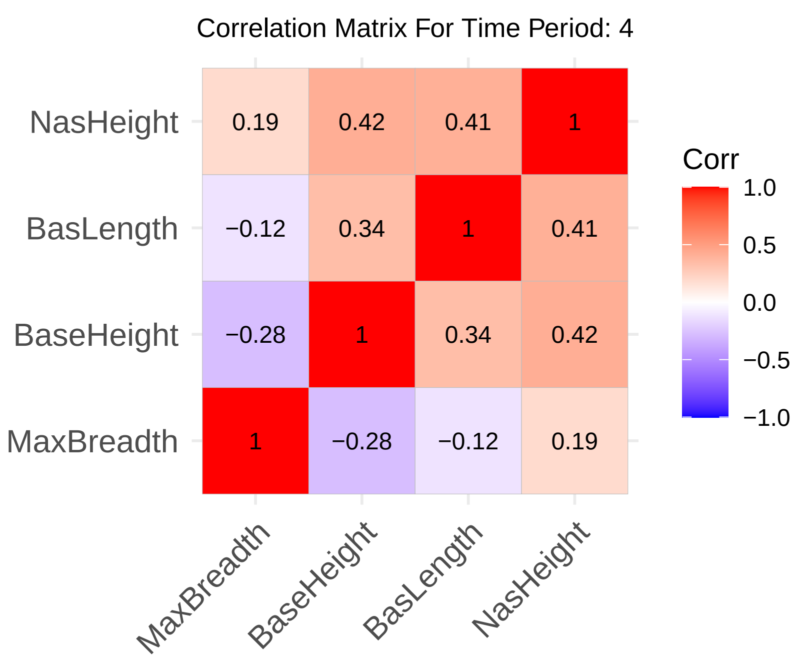
```
Loading required package: ggplot2
```

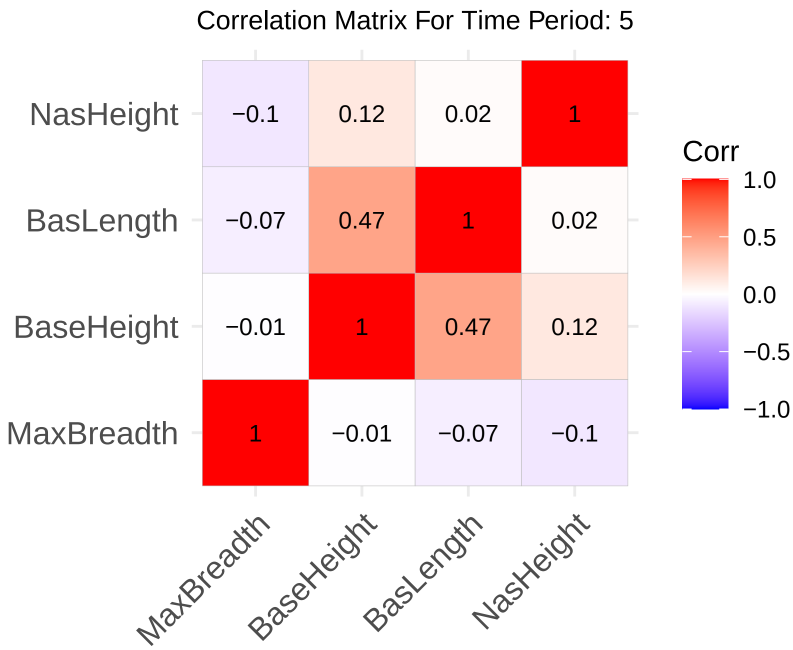
```
library(gridExtra)

for(i in 1:5){
  # X_bar for each time period
  X_bar <- matrix(mean_matrix[1:4,i])

  # X matrix for each time period
  X_group <- X_total[X_total[,5] == i,1:4]

  # Variance-Covariance Matrix
  S <- t(X_group) %*% X_group - nrow(X_group)*X_bar %*% t(X_bar)
  S <- S/(nrow(X_group) - 1)

  # Diagonals Matrix
  D <- diag(diag(S))

  # Correlation Matrix
  R <- solve(D)^(1/2) %*% S %*% solve(D)^(1/2)
  rownames(R) <- c("MaxBreadth", "BaseHeight", "BasLength",
                   "NasHeight")
  colnames(R) <- c("MaxBreadth", "BaseHeight", "BasLength",
                   "NasHeight")

  # Correlation Heat Map Plot
  print(ggcorrplot(R, lab = TRUE, lab_size = 3) +
    ggtitle(paste("Correlation Matrix For Time Period:", i)) +
    theme(plot.title = element_text(hjust = 0.5, size=10),
          axis.text=element_text(size=8)))
}
```


Correlation Matrix For Time Period: 1
NasHeight
BasLength
BaseHeight
MaxBreadth
0.51
0.03
−0.12
1
0.02
−0.03
1
−0.12
0.18
1
−0.03
0.03
1
0.18
0.02
0.51
MaxBreadth
BaseHeight
BasLength
NasHeight
Corr
1.0
0.5
0.0
−0.5
−1.0


Correlation Matrix For Time Period: 2
NasHeight
BasLength
BaseHeight
MaxBreadth
0.13
0.41
0.01
1
0.23
0.17
1
0.01
0.05
1
0.17
0.41
1
0.05
0.23
0.13
MaxBreadth
BaseHeight
BasLength
NasHeight
Corr
1.0
0.5
0.0
−0.5
−1.0


Correlation Matrix For Time Period: 3
NasHeight
BasLength
BaseHeight
MaxBreadth
0.07
−0.01
0.1
1
−0.05
0.16
1
0.1
0.05
1
0.16
−0.01
1
0.05
−0.05
0.07
MaxBreadth
BaseHeight
BasLength
NasHeight
Corr
1.0
0.5
0.0
−0.5
−1.0


Correlation Matrix For Time Period: 4
NasHeight
BasLength
BaseHeight
MaxBreadth
0.19
0.42
0.41
1
−0.12
0.34
1
0.41
−0.28
1
0.34
0.42
1
−0.28
−0.12
0.19
MaxBreadth
BaseHeight
BasLength
NasHeight
Corr
1.0
0.5
0.0
−0.5
−1.0



And briefly interpret the heat map. Are there any noticeable changes over the time periods?

1. The correlation between the MaxBreadth and NasHeight went from being positively correlated to each other in time period 1 to gradually becoming uncorrelated to each other by time period 5}

2. The correlation between BaseHeight and BasLength went from being uncorrelated to each other in time period 1, to slowy become more and more positively correlated to each other by time period 5

3. The correlation between the MaxBreadth and BaseHeight went from being positively correlated in time period 1 to gradually became less and less uncorrelated by time period 5.

Question 3

Calculate the angle between the deviation vectors for X_1 and X_3 in period 1.

```r
# Deviation Vectors
X_center <- scale(X_total[X_total[,5] == 1,], center = T, scale = F)
# same as di = xi - x_bar * 1
d1 <- matrix(X_center[,1])
d2 <- matrix(X_center[,3])

# cos(theta) = angle between deviation vectors X1 and X3
cos_theta <- (t(d1) %*% d2)/((sqrt(t(d1)%*%(d1)))*(sqrt(t(d2)%*%(d2))))
theta <- acos(cos_theta)
print(paste("Correlation:", cos_theta))
```

```
[1] "Correlation: 0.0150425047831862"
```

```r
print(paste("Angle in degrees:", theta*180/pi))
```

```
[1] "Angle in degrees: 89.1380954556786"
```

```r
print(paste("Angle in radians:", theta))
```

```
[1] "Angle in radians: 1.55575325465859"
```

Explain why this value is to be expected by referring to the appropriate value from question 2.

```r
cor(X_total[X_total[,5] == 1,1:4])
```

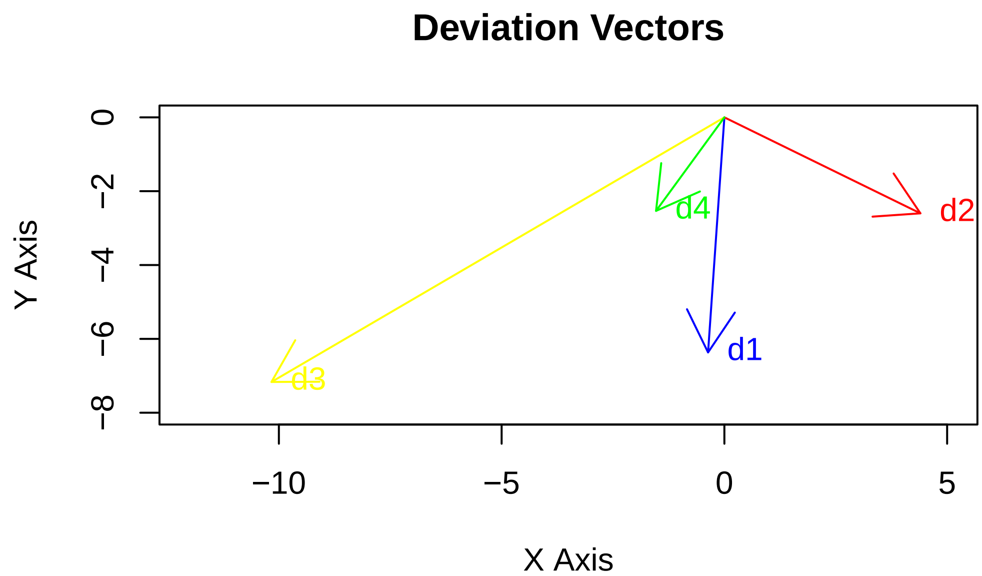
```
           MaxBreadth   BasHeight   BasLength   NasHeight
MaxBreadth  1.0000000  0.18111709  0.01504250  0.51119663
BasHeight   0.1811171  1.00000000 -0.03015856  0.03182998
BasLength   0.0150425 -0.03015856  1.00000000 -0.11804243
NasHeight   0.5111966  0.03182998 -0.11804243  1.00000000
```

```r
# The correlation between X1 (MaxBreadth) and X3 (BasLength) is 0.01504250
# which is the same as finding the cos(angle) between the two deviation vectors
cos(theta)
```

```
          [,1]
[1,] 0.0150425
```

For a bonus mark, plot all the deviation vectors for period 1 across the first two observations.

```
X_center <- scale(X_total[X_total[,5] == 1,1:4], center = T, scale = F)

# Plot deviation Vectors
d1 <- matrix(X_center[1:2,1])
d2 <- matrix(X_center[1:2,2])
d3 <- matrix(X_center[1:2,3])
d4 <- matrix(X_center[1:2,4])

library(ggplot2)

# Base Plot Setup
plot(0,0, type = "n", xlim=c(-12,5), ylim=c(-8,0), xlab ="X Axis",
     ylab = "Y Axis", main = "Deviation Vectors")

# Add Arrow
arrows(0, 0, d1[1], d1[2], col = "blue")
arrows(0, 0, d2[1], d2[2], col = "red")
arrows(0, 0, d3[1], d3[2], col = "yellow")
arrows(0, 0, d4[1], d4[2], col = "green")

# Add labels near arrow tips
text(d1[1], d1[2], labels = "d1", pos = 4, col = "blue")
text(d2[1], d2[2], labels = "d2", pos = 4, col = "red")
text(d3[1], d3[2], labels = "d3", pos = 4, col = "yellow")
text(d4[1], d4[2], labels = "d4", pos = 4, col = "green")
```



Question 4

Suppose researchers are interested in the quantity $Y_i = 3X_4 - X_1$ for time periods i = 1, . . . ,5. Use your answers from question 1 and an appropriate vector b to determine the sample means $\bar{y_1}, ..., \bar{y_5}$. Also give the covariance matrix of Y = [Y1 Y2 Y3 Y4 Y5]'

```
Y <- matrix(0, nrow=nrow(X_total[X_total[,5]==1,1:4]), ncol=5)
Y_Expectation <- matrix(0, nrow = 5,ncol=1)
b <- matrix(data = c(-1,0,0,3,0), nrow = 5, ncol = 1)

for(i in 1:5){
  # Yi = 3*X4-X1 so X*b_vector for each time period
  Y[,i] <- (X_total[X_total[,5]==i,]) %*% b

  # E(b'X) = E(b1*x1 + b2*x2 + b3*x3 + b4*x4 + b5*x5)
  # = b1*E(x1) + b2*E(x2) + b3*E(x3) + b4*E(x4) + b5*E(x5)
  X_bar <- matrix(mean_matrix[,i])
  Y_Expectation[i] <- t(b) %*% X_bar
}
```

```
rownames(Y_Expectation) <- c("1", "2", "3", "4", "5")
colnames(Y) <- c("Y1", "Y2", "Y3", "Y4", "Y5")

print("Sample means yi for time periods i = 1,2,3,4,5:")
```

```
[1] "Sample means yi for time periods i = 1,2,3,4,5:"
```

```
Y_Expectation
```

```
      [,1]
1 20.23333
2 18.33333
3 17.23333
4 20.40000
5 17.93333
```

```
# Finding the covariance for Y
Y_bar <- (1/nrow(Y)) * t(Y) %*% matrix(1, nrow=30, ncol=1)
# Same as Y_Expectation (checking)
S <- t(Y) %*% Y- nrow(Y)*Y_bar%*%t(Y_bar)
S <- S/(nrow(Y)-1)
print("Covariances Yi for time periods i = 1,2,3,4,5:")
```

```
[1] "Covariances Yi for time periods i = 1,2,3,4,5:"
```

```
S
```

```
           Y1        Y2        Y3         Y4          Y5
Y1 51.5643678  1.402299  23.32299  -7.924138   0.2229885
Y2  1.4022989 90.712644  -6.08046  29.862069   2.8850575
Y3 23.3229885 -6.080460 120.11609 -10.131034 -33.4321839
Y4 -7.9241379 29.862069 -10.13103  74.731034  14.6482759
Y5  0.2229885  2.885057 -33.43218  14.648276 165.0298851
```