STOR 565 Spring 2018 Homework 1

Shilpa Kancharla

Exercise 1. *(5 pt)* Using the `c`, `rep` or `seq` commands, create the following 6 vectors:

$x1 = (2, .5, 4, 2)$;

$x2 = (2, .5, 4, 2, 1, 1, 1, 1)$;

$x3 = (1, 0, -1, -2)$;

$x4 = ("Hello", " ", "World, "!", "Hello World!")$;

Hint. Try `paste` function.

$x5 = (TRUE, TRUE, NA, FALSE)$;

Hint. Check `?NA` and `class(NA)` to learn more about the missing value object `NA`.

$x6 = (1, 2, 1, 2, 1, 1, 2, 2)$.

```
x1 <- c(2, .5, 4, 2);
x2 <- c(2, .5, 4, 2, rep(1, 4));
x3 <- seq(1, -2, -1);
x4 <- paste("Hello"," ", "World", "!", "Hello World!", sep=" ", collapse = NULL);
x5 <- c(rep(TRUE,2), NA, TRUE);
x6 <- c(1, 2, 1, 2, rep(1, 2), rep(2, 2));
x1
```

```
## [1] 2.0 0.5 4.0 2.0
```

```
x2
```

```
## [1] 2.0 0.5 4.0 2.0 1.0 1.0 1.0 1.0
```

```
x3
```

```
## [1]  1  0 -1 -2
```

```
x4
```

```
## [1] "Hello   World ! Hello World!"
```

```
x5
```

```
## [1] TRUE TRUE   NA TRUE
```

```
x6
```

```
## [1] 1 2 1 2 1 1 2 2
```

Exercise 2. *(5 pt)* Using `matrix`, `rbind` and `cbind`, create

$$\mathbf{X} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 0 & -1 & -2 \\ 2 & .5 & 4 & 2 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

```
A = matrix(c(1,1,2,1), 4, 1)
B = matrix(c(2,0,.5,1), 4, 1)
C = matrix(c(3, -1, 4, 1), 4, 1)
D = matrix(c(4, -2, 2, 1), 4, 1)
```

```
X <- cbind(A, B, C, D)
X
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    1  2.0    3    4
## [2,]    1  0.0   -1   -2
## [3,]    2  0.5    4    2
## [4,]    1  1.0    1    1
```

```
students <- data.frame( id      = c("001", "002", "003"), # ids are characters
                        score_A = c(95, 97, 90),           # scores are numericss
                        score_B = c(80, 75, 84))
students
```

Exercise 3. *(5 pt)* Applying the conditional selection technique (without using `subset`), extract the record of 003 in `students`.

```
students3 <- students[3,]
students3
```

```
##    id score_A score_B
## 3 003      90      84
```

One can also create a matrix or a legitimate list first and then convert it into a data.frame as follows.

```
scores <- matrix(c(95, 97, 90, 80, 75, 84), 3, 2)
scores <- data.frame(scores)
colnames(scores) <- c("score_A", "score_B")
id <- c("001", "002", "003")
students1 <- cbind(id, scores)
students2 <- data.frame( list( id      = c("001", "002", "003"),
                               score_A = c(95, 97, 90),
                               score_B = c(80, 75, 84))
                       )
```

Exercise 4. *(10 pt)* Create a data.frame object to display the calendar for Jan 2018 as follows.

```
## Sun Mon Tue Wed Thu Fri Sat
##       NY   2   3   4   5   6
##   7   8   9  10  11  12  13
## MLK  15  16  17  18  19  20
##  21  22  23  24  25  26  27
##  28  29  30  31
calendar <- matrix(c(" ", 7, "MLK", 21, 28, "NY", 8,
                     15, 22, 29, 2, 9, 16, 23, 30, 3,
                     10, 17, 24, 31, 4, 11, 18, 25, " ",
                     5, 12, 19, 26, " ", 6, 13, 20, 27, " "), 5, 7)
calendar <- data.frame(calendar)
colnames(calendar) <- c("Sun", "Mon", "Tue", "Wed", "Thu", "Fri", "Sat")
print(calendar, row.names=FALSE)
```

```
##  Sun Mon Tue Wed Thu Fri Sat
##        NY   2   3   4   5   6
##    7   8   9  10  11  12  13
##  MLK  15  16  17  18  19  20
##   21  22  23  24  25  26  27
##   28  29  30  31
```

Factors

Factor is a special data structure in **R** in representing categorical variables and facilitating the data labels and subgroups. It's basically a character vector that keeps track of its distinct values called levels. Consider the longitudinal layout of the previous score table.

```
id        <- rep(c("001","002","003"), 2)
subj      <- rep(c("A","B"), each = 3)
score     <- c(95, 97, 90, 80, 75, 84)
students3 <- data.frame(id, subj, score)  # try cbind(id, subj, score) to see the difference

# students3$id and students3$subj are automatically formatted as factors
class(students3$id)
levels(students3$id)

class(students3$subj)
levels(students3$subj)

# combind student 003 with 002 via level rename
students4 <- students3    # work on a copy in case of direct modification of students3
levels(students4$id)[3] <- "002"
levels(students4$id)
students4
```

Exercise 5. *(5 pt)* Create a factor variable `grade` in `students3`, where the `score` variable is divided into $[90, 100]$, $[80, 89)$ and $[0, 79)$ corresponding to A, B and C in `grade` respectively.

```
grades <- cut(students3$score, breaks = c(100, 89, 79, 0), labels=c("C","B","A"))
grades
```

```
## [1] A A A B C B
## Levels: C B A
```

```
students3 <- data.frame(id, subj, score, grades)
students3
```

```
##     id subj score grades
## 1 001    A    95      A
## 2 002    A    97      A
## 3 003    A    90      A
## 4 001    B    80      B
## 5 002    B    75      C
## 6 003    B    84      B
```

Operations and Functions

Exercise 6. *(5 pt)* Without using the `var` and `scale` functions, compute the sample covariance `X.var` of the data matrix `X` as in **Exercise 2**.

```
k <- ncol(X) #number of variables
n <- nrow(X) #number of subjects
X_mean <- matrix(data=1,nrow=n) %*% cbind(mean(A),mean(B),mean(C),mean(D)) #creates means for each colu
H <- X - X_mean #creates a difference matrix
X.var <- (n-1)^-1 * t(H) %*% H #creates covariance matrix
X.var
```

```
##          [,1]      [,2]      [,3]  [,4]
```

```
## [1,]   0.250 -0.1250000 0.7500000 0.250
## [2,] -0.125   0.7291667 0.9583333 1.875
## [3,]   0.750   0.9583333 4.9166667 4.750
## [4,]   0.250   1.8750000 4.7500000 6.250
```

```r
cov(X)
```

```
##          [,1]      [,2]      [,3]  [,4]
## [1,]   0.250 -0.1250000 0.7500000 0.250
## [2,] -0.125   0.7291667 0.9583333 1.875
## [3,]   0.750   0.9583333 4.9166667 4.750
## [4,]   0.250   1.8750000 4.7500000 6.250
```

Exercise 7. *(10 pt)* Create a variable `score.mean` in `students3`, taking value as the mean score among students who have the same `subj` value.

```r
A_mean <- mean(score[students3$subj == "A"])
B_mean <- mean(score[students3$subj == "B"])
score.mean <- c(A_mean,B_mean)
score.mean
```

```
## [1] 94.00000 79.66667
```

Writing your own functions

Exercise 8. *(15 pt)* Write a function `bisect(f, lower, upper, tol = 1e-6)` to find the root of the univariate function `f` on the interval $[\text{lower}, \text{upper}]$ with precision tolerance $\leqslant$ `tol` (defaulted to be 10^{-6}) via bisection, which returns a list consisting of `root`, `f.root` (`f` evaluated at `root`), `iter` (number of iterations) and `estim.prec` (estimated precision). Apply it to the function

$$f(x) = x^3 - x - 1$$

on $[1, 2]$ with precision tolerance 10^{-6}. Compare it with the built-in function `uniroot`.

```r
bisection <- function(f, a, b, n = 1000, tol = 1e-6) {
  if (!(f(a) < 0) && (f(b) > 0)) {
    stop('signs of f(a) and f(b) differ')
  } else if ((f(a) > 0) && (f(b) < 0)) {
    stop('signs of f(a) and f(b) differ')
  }

  for (i in 1:n) {
    c <- (a + b) / 2 # Calculate midpoint
    if ((f(c) == 0) || ((b - a) / 2) < tol) {
      return(c)
    }
    ifelse(sign(f(c)) == sign(f(a)),
           a <- c,
           b <- c)
  }
  print('Too many iterations')
}
f <- function(x) {x^3 - x -1} #function
a <- 1 #lower
b <- 2 #upper
tol <- 1e-6 #tolerance
```

	V1	V2	V3	V4
V1		-0.29	0.68	0.2
V2			0.51	0.88
V3				0.86
V4				

Table 1

```r
n <- 10000 #iterations
bisection(f, a, b, n, tol)
```

```
## [1] 1.324718
```

Exercise 9. *(15 pt)* Without using `cor`, compute the sample correlation matrix `X.cor` from `X.var` in **Exercise 6.** Output `X.cor` to a text file "X_cor.txt" which displays as in Table 1. Then input "X_cor.txt" in **R** and reproduce the correlation matrix `X.cor1`.

Hint. 1) Functions `round` and `lower.tri`; 2) the `NA` trick; 3) options `sep = "\t"`, `col.names = NA`.

```
##        V1    V2   V3   V4
## V1   1.00 -0.29 0.68 0.20
## V2  -0.29  1.00 0.51 0.88
## V3   0.68  0.20 1.00 0.86
## V4   0.51  0.88 0.86 1.00
```

```r
value <- sqrt(diag(X.var))
value1 <- solve(diag(value))
X.cor <- value1 %*% X.var %*% value1
X.cor
```

```
##              [,1]       [,2]      [,3]      [,4]
## [1,]   1.0000000 -0.2927700 0.6764814 0.2000000
## [2,]  -0.2927700  1.0000000 0.5061367 0.8783101
## [3,]   0.6764814  0.5061367 1.0000000 0.8568765
## [4,]   0.2000000  0.8783101 0.8568765 1.0000000
```

```r
#Output as text file
write.table(X.cor, file="X_cor.txt", sep="\t", col.names=NA, quote=FALSE)
X.cor1 <- read.table("X_cor.txt", header=FALSE, sep = "\t")
X.cor1
```

```
##    V1                  V2                 V3                 V4
## 1 NA                  V1                 V2                 V3
## 2  1                   1 -0.29277002188456 0.676481425202546
## 3  2 -0.29277002188456                  1 0.506136675356011
## 4  3 0.676481425202546 0.506136675356011                  1
## 5  4               0.2  0.87831006565368 0.856876471923225
##                 V5
## 1               V4
## 2              0.2
## 3  0.87831006565368
## 4 0.856876471923225
## 5                1
```

```r
cov2cor(X.var)
```

```
##              [,1]       [,2]      [,3]      [,4]
## [1,]   1.0000000 -0.2927700 0.6764814 0.2000000
```

```
## [2,] -0.2927700  1.0000000 0.5061367 0.8783101
## [3,]  0.6764814  0.5061367 1.0000000 0.8568765
## [4,]  0.2000000  0.8783101 0.8568765 1.0000000
```

Exercise 10. *(10 pt)* `iris` is a built-in dataset in **R**. Check `?iris` for more information. Randomly divide `iris` into five subsets `iris1` to `iris5` stratified to `iris$Species` (namely, the proportion of `iris$Species` among different levels remains identical across all subsets).

```
set.seed(5)
irisnew <- split(iris, sample(1:5))
iris1 <- irisnew[1]
iris2 <- irisnew[2]
iris3 <- irisnew[3]
iris4 <- irisnew[4]
iris5 <- irisnew[5]
iris.5fold <- list(iris1, iris2, iris3, iris4, iris5)
iris.5fold
```

```
## [[1]]
## [[1]]$`1`
##     Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
## 4            4.6         3.1          1.5         0.2     setosa
## 9            4.4         2.9          1.4         0.2     setosa
## 14           4.3         3.0          1.1         0.1     setosa
## 19           5.7         3.8          1.7         0.3     setosa
## 24           5.1         3.3          1.7         0.5     setosa
## 29           5.2         3.4          1.4         0.2     setosa
## 34           5.5         4.2          1.4         0.2     setosa
## 39           4.4         3.0          1.3         0.2     setosa
## 44           5.0         3.5          1.6         0.6     setosa
## 49           5.3         3.7          1.5         0.2     setosa
## 54           5.5         2.3          4.0         1.3 versicolor
## 59           6.6         2.9          4.6         1.3 versicolor
## 64           6.1         2.9          4.7         1.4 versicolor
## 69           6.2         2.2          4.5         1.5 versicolor
## 74           6.1         2.8          4.7         1.2 versicolor
## 79           6.0         2.9          4.5         1.5 versicolor
## 84           6.0         2.7          5.1         1.6 versicolor
## 89           5.6         3.0          4.1         1.3 versicolor
## 94           5.0         2.3          3.3         1.0 versicolor
## 99           5.1         2.5          3.0         1.1 versicolor
## 104          6.3         2.9          5.6         1.8  virginica
## 109          6.7         2.5          5.8         1.8  virginica
## 114          5.7         2.5          5.0         2.0  virginica
## 119          7.7         2.6          6.9         2.3  virginica
## 124          6.3         2.7          4.9         1.8  virginica
## 129          6.4         2.8          5.6         2.1  virginica
## 134          6.3         2.8          5.1         1.5  virginica
## 139          6.0         3.0          4.8         1.8  virginica
## 144          6.8         3.2          5.9         2.3  virginica
## 149          6.2         3.4          5.4         2.3  virginica
##
##
## [[2]]
## [[2]]$`2`
```

```
##     Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
## 1            5.1         3.5          1.4         0.2     setosa
## 6            5.4         3.9          1.7         0.4     setosa
## 11           5.4         3.7          1.5         0.2     setosa
## 16           5.7         4.4          1.5         0.4     setosa
## 21           5.4         3.4          1.7         0.2     setosa
## 26           5.0         3.0          1.6         0.2     setosa
## 31           4.8         3.1          1.6         0.2     setosa
## 36           5.0         3.2          1.2         0.2     setosa
## 41           5.0         3.5          1.3         0.3     setosa
## 46           4.8         3.0          1.4         0.3     setosa
## 51           7.0         3.2          4.7         1.4 versicolor
## 56           5.7         2.8          4.5         1.3 versicolor
## 61           5.0         2.0          3.5         1.0 versicolor
## 66           6.7         3.1          4.4         1.4 versicolor
## 71           5.9         3.2          4.8         1.8 versicolor
## 76           6.6         3.0          4.4         1.4 versicolor
## 81           5.5         2.4          3.8         1.1 versicolor
## 86           6.0         3.4          4.5         1.6 versicolor
## 91           5.5         2.6          4.4         1.2 versicolor
## 96           5.7         3.0          4.2         1.2 versicolor
## 101          6.3         3.3          6.0         2.5  virginica
## 106          7.6         3.0          6.6         2.1  virginica
## 111          6.5         3.2          5.1         2.0  virginica
## 116          6.4         3.2          5.3         2.3  virginica
## 121          6.9         3.2          5.7         2.3  virginica
## 126          7.2         3.2          6.0         1.8  virginica
## 131          7.4         2.8          6.1         1.9  virginica
## 136          7.7         3.0          6.1         2.3  virginica
## 141          6.7         3.1          5.6         2.4  virginica
## 146          6.7         3.0          5.2         2.3  virginica
##
##
## [[3]]
## [[3]]$`3`
##     Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
## 2            4.9         3.0          1.4         0.2     setosa
## 7            4.6         3.4          1.4         0.3     setosa
## 12           4.8         3.4          1.6         0.2     setosa
## 17           5.4         3.9          1.3         0.4     setosa
## 22           5.1         3.7          1.5         0.4     setosa
## 27           5.0         3.4          1.6         0.4     setosa
## 32           5.4         3.4          1.5         0.4     setosa
## 37           5.5         3.5          1.3         0.2     setosa
## 42           4.5         2.3          1.3         0.3     setosa
## 47           5.1         3.8          1.6         0.2     setosa
## 52           6.4         3.2          4.5         1.5 versicolor
## 57           6.3         3.3          4.7         1.6 versicolor
## 62           5.9         3.0          4.2         1.5 versicolor
## 67           5.6         3.0          4.5         1.5 versicolor
## 72           6.1         2.8          4.0         1.3 versicolor
## 77           6.8         2.8          4.8         1.4 versicolor
## 82           5.5         2.4          3.7         1.0 versicolor
## 87           6.7         3.1          4.7         1.5 versicolor
```

```
## 92            6.1            3.0            4.6            1.4 versicolor
## 97            5.7            2.9            4.2            1.3 versicolor
## 102           5.8            2.7            5.1            1.9  virginica
## 107           4.9            2.5            4.5            1.7  virginica
## 112           6.4            2.7            5.3            1.9  virginica
## 117           6.5            3.0            5.5            1.8  virginica
## 122           5.6            2.8            4.9            2.0  virginica
## 127           6.2            2.8            4.8            1.8  virginica
## 132           7.9            3.8            6.4            2.0  virginica
## 137           6.3            3.4            5.6            2.4  virginica
## 142           6.9            3.1            5.1            2.3  virginica
## 147           6.3            2.5            5.0            1.9  virginica
##
##
## [[4]]
## [[4]]$`4`
##     Sepal.Length Sepal.Width Petal.Length Petal.Width   Species
## 3            4.7            3.2            1.3            0.2   setosa
## 8            5.0            3.4            1.5            0.2   setosa
## 13           4.8            3.0            1.4            0.1   setosa
## 18           5.1            3.5            1.4            0.3   setosa
## 23           4.6            3.6            1.0            0.2   setosa
## 28           5.2            3.5            1.5            0.2   setosa
## 33           5.2            4.1            1.5            0.1   setosa
## 38           4.9            3.6            1.4            0.1   setosa
## 43           4.4            3.2            1.3            0.2   setosa
## 48           4.6            3.2            1.4            0.2   setosa
## 53           6.9            3.1            4.9            1.5 versicolor
## 58           4.9            2.4            3.3            1.0 versicolor
## 63           6.0            2.2            4.0            1.0 versicolor
## 68           5.8            2.7            4.1            1.0 versicolor
## 73           6.3            2.5            4.9            1.5 versicolor
## 78           6.7            3.0            5.0            1.7 versicolor
## 83           5.8            2.7            3.9            1.2 versicolor
## 88           6.3            2.3            4.4            1.3 versicolor
## 93           5.8            2.6            4.0            1.2 versicolor
## 98           6.2            2.9            4.3            1.3 versicolor
## 103          7.1            3.0            5.9            2.1  virginica
## 108          7.3            2.9            6.3            1.8  virginica
## 113          6.8            3.0            5.5            2.1  virginica
## 118          7.7            3.8            6.7            2.2  virginica
## 123          7.7            2.8            6.7            2.0  virginica
## 128          6.1            3.0            4.9            1.8  virginica
## 133          6.4            2.8            5.6            2.2  virginica
## 138          6.4            3.1            5.5            1.8  virginica
## 143          5.8            2.7            5.1            1.9  virginica
## 148          6.5            3.0            5.2            2.0  virginica
##
##
## [[5]]
## [[5]]$`5`
##     Sepal.Length Sepal.Width Petal.Length Petal.Width   Species
## 5            5.0            3.6            1.4            0.2   setosa
## 10           4.9            3.1            1.5            0.1   setosa
```

```
## 15              5.8         4.0         1.2         0.2       setosa
## 20              5.1         3.8         1.5         0.3       setosa
## 25              4.8         3.4         1.9         0.2       setosa
## 30              4.7         3.2         1.6         0.2       setosa
## 35              4.9         3.1         1.5         0.2       setosa
## 40              5.1         3.4         1.5         0.2       setosa
## 45              5.1         3.8         1.9         0.4       setosa
## 50              5.0         3.3         1.4         0.2       setosa
## 55              6.5         2.8         4.6         1.5 versicolor
## 60              5.2         2.7         3.9         1.4 versicolor
## 65              5.6         2.9         3.6         1.3 versicolor
## 70              5.6         2.5         3.9         1.1 versicolor
## 75              6.4         2.9         4.3         1.3 versicolor
## 80              5.7         2.6         3.5         1.0 versicolor
## 85              5.4         3.0         4.5         1.5 versicolor
## 90              5.5         2.5         4.0         1.3 versicolor
## 95              5.6         2.7         4.2         1.3 versicolor
## 100             5.7         2.8         4.1         1.3 versicolor
## 105             6.5         3.0         5.8         2.2  virginica
## 110             7.2         3.6         6.1         2.5  virginica
## 115             5.8         2.8         5.1         2.4  virginica
## 120             6.0         2.2         5.0         1.5  virginica
## 125             6.7         3.3         5.7         2.1  virginica
## 130             7.2         3.0         5.8         1.6  virginica
## 135             6.1         2.6         5.6         1.4  virginica
## 140             6.9         3.1         5.4         2.1  virginica
## 145             6.7         3.3         5.7         2.5  virginica
## 150             5.9         3.0         5.1         1.8  virginica
```
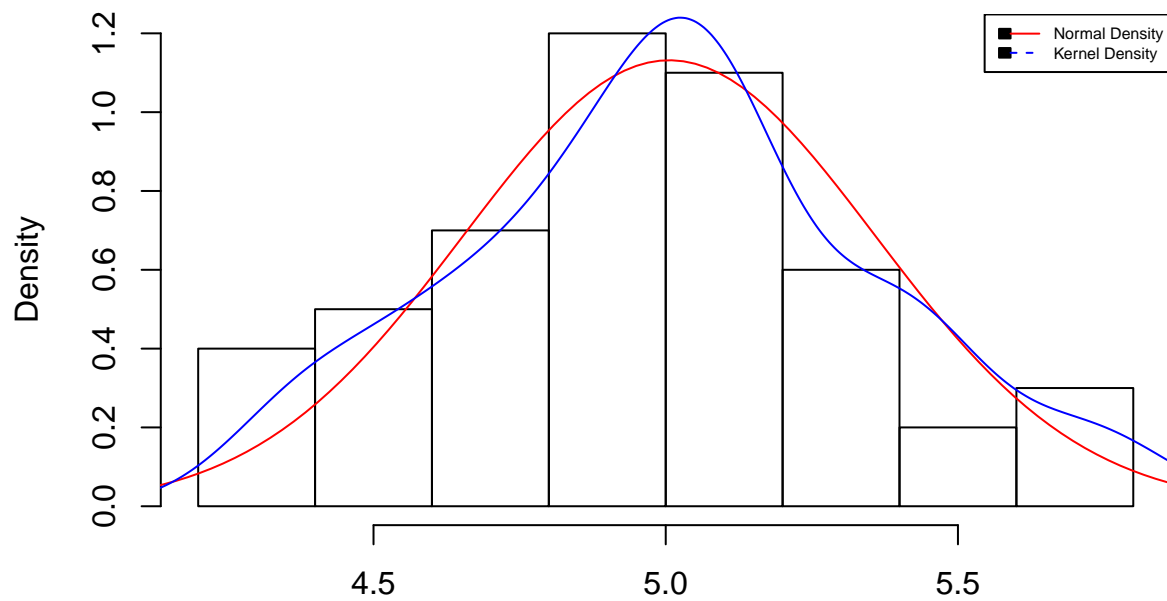
Exercise 11. *(15 pt)* Reproduce the code that generates the following plot about `Sepal.Length` in `iris`.

```
knitr::include_graphics("Iris_Sepal_Length.pdf")
```

Hint. 1) Most decorations are based on defaults in `hist` with `ylim = c(0, 1.3)`; 2) let `h` be the object resulted by `hist` and set `xlim = range(h$breaks)`; 3) set `cex = 0.5` in `legend`; 4) use `curve` function in plotting the normal density with specified parameters.
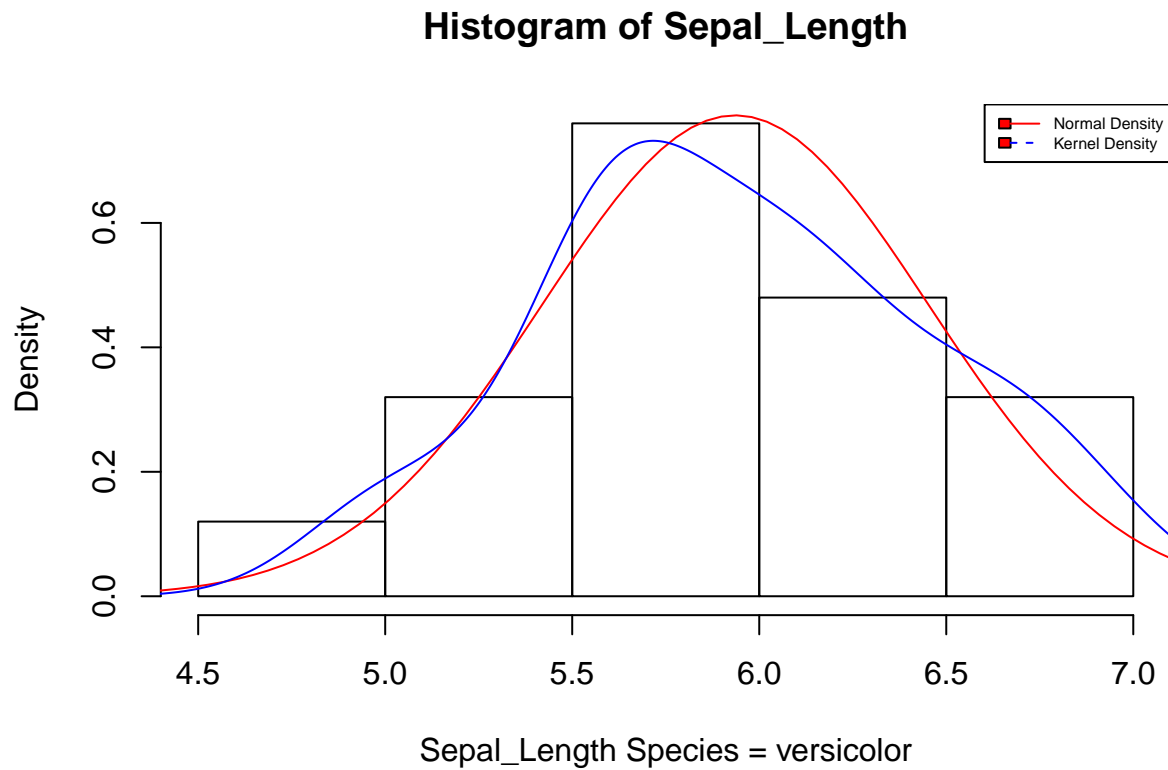
```r
setosa_length <- iris$Sepal.Length[iris$Species == "setosa"]
hist(setosa_length, main = "Histogram of Sepal_Length",
     xlab = "Sepal_Length Species = setosa", freq=FALSE)
x <- seq(4, 6, length.out=100)
y <- with(iris, dnorm(x, mean(setosa_length), sd(setosa_length)))
lines(x, y, col="red")
lines(density(setosa_length), col="blue")
legend("topright", 5, 1.3, legend=c("Normal Density", "Kernel Density"),
       col=c("red", "blue"), lty=1:2, cex=0.5)
```

Histogram of Sepal_Length

Sepal_Length Species = setosa

```
versicolor_length <- iris$Sepal.Length[iris$Species == "versicolor"]
hist(versicolor_length, main = "Histogram of Sepal_Length",
     xlab = "Sepal_Length Species = versicolor", freq=FALSE)
x <- seq(4, 8, length.out=100)
y <- with(iris, dnorm(x, mean(versicolor_length), sd(versicolor_length)))
lines(x, y, col="red")
lines(density(versicolor_length), col="blue")
legend("topright", 5, 2, legend=c("Normal Density", "Kernel Density"),
       col=c("red", "blue"), lty=1:2, cex=0.5)
```

Histogram of Sepal_Length

Sepal_Length Species = versicolor

```r
virginica_length <- iris$Sepal.Length[iris$Species == "virginica"]
hist(virginica_length, main = "Histogram of Sepal_Length",
     xlab = "Sepal_Length Species = virginica", freq=FALSE)
x <- seq(4, 8, length.out=100)
y <- with(iris, dnorm(x, mean(virginica_length), sd(virginica_length)))
lines(x, y, col="red")
lines(density(virginica_length), col="blue")
legend("topright", 5, 2, legend=c("Normal Density", "Kernel Density"),
       col=c("red", "blue"), lty=1:2, cex=0.5)
```

Histogram of Sepal_Length

Sepal_Length Species = virginica