week6 exercise

Xiang Li

2023/11/1

Exercise 1

1

```
f1 = function (df){
  num_df = df[, sapply(df, is.numeric)]
  summary_df = data.frame(colname = colnames(num_df), mean = unname(sapply(num_df, mean)))
  result = list(dimensions = dim(df), names = colnames(df), summary = summary_df)
  return(result)
}
```

2

```
data(iris)
iris_df = as.data.frame(iris)
f1_result = f1(iris_df)
f1_result
```

```
## $dimensions
## [1] 150   5
##
## $names
## [1] "Sepal.Length" "Sepal.Width"  "Petal.Length" "Petal.Width"  "Species"
##
## $summary
##        colname     mean
## 1 Sepal.Length 5.843333
## 2  Sepal.Width 3.057333
## 3 Petal.Length 3.758000
## 4  Petal.Width 1.199333
```

3

```
print(f1_result[[3]])
```

```
##        colname     mean
## 1 Sepal.Length 5.843333
## 2  Sepal.Width 3.057333
## 3 Petal.Length 3.758000
## 4  Petal.Width 1.199333
```

4

```
print(f1_result[[2]][1])
```

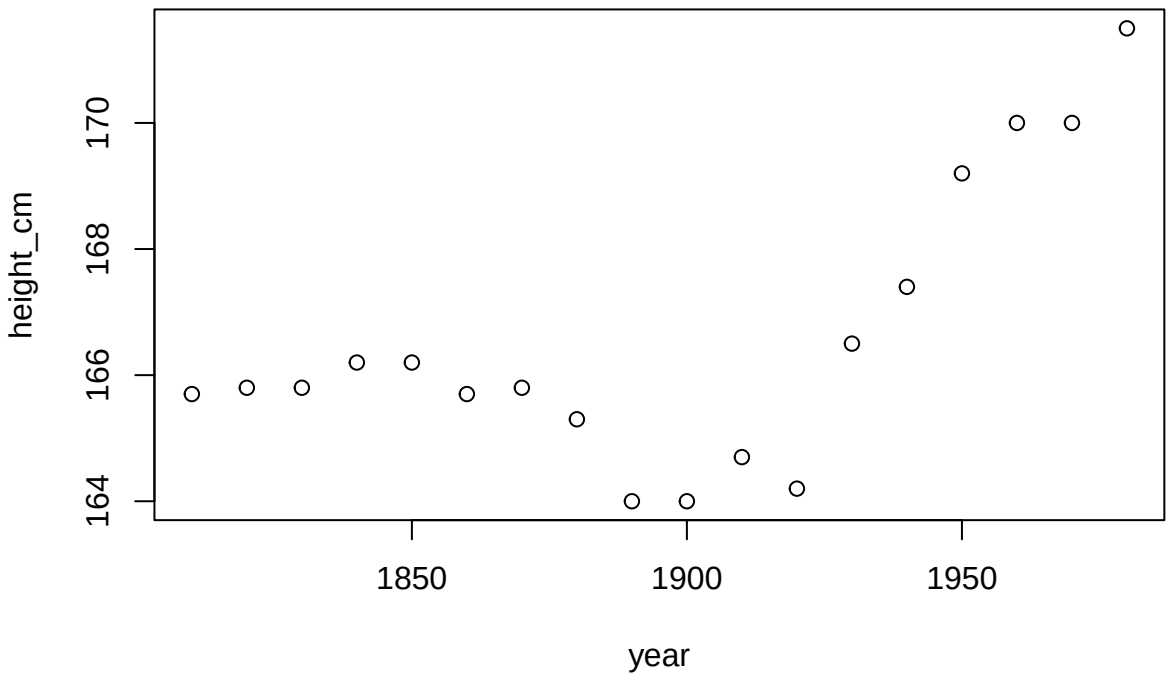
```
## [1] "Sepal.Length"
```

Exercise 2

```
library(brolgar)
data(heights)
heights_df = as.data.frame(heights)
```

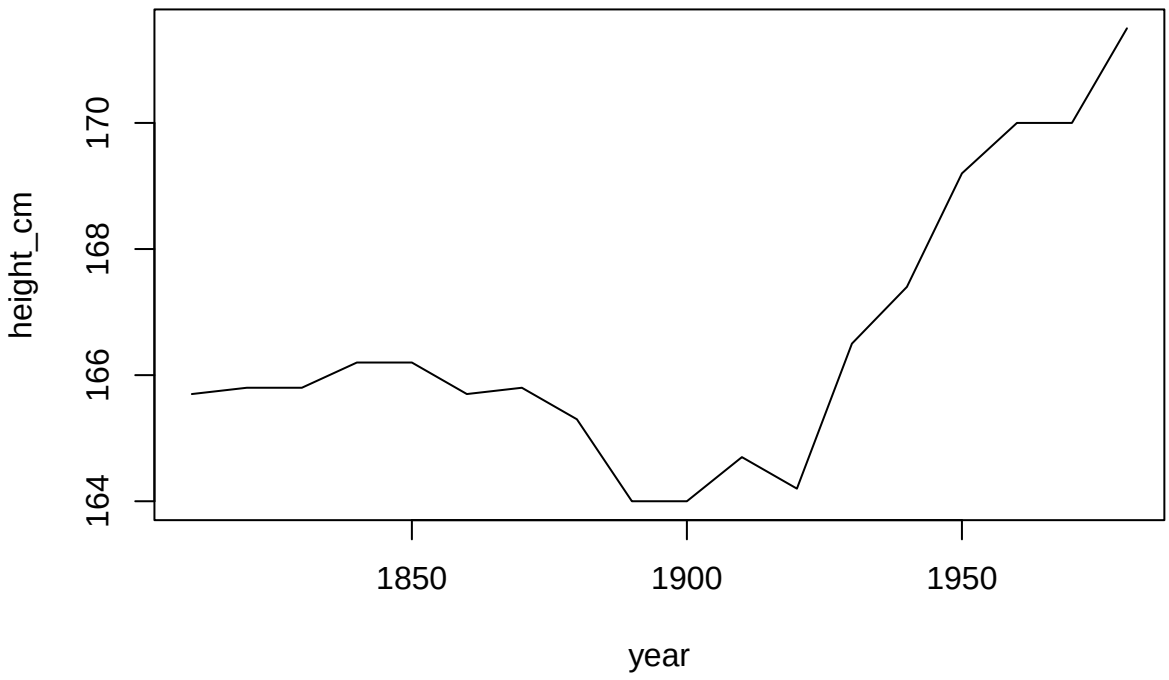
1

```
height_scatter = function(country_name){
  coun_heights_df = heights_df[heights_df$country == country_name, ]
  plot(coun_heights_df$year, coun_heights_df$height_cm, type = 'p', xlab = 'year', ylab = 'height_cm')
}
height_scatter('China')
```



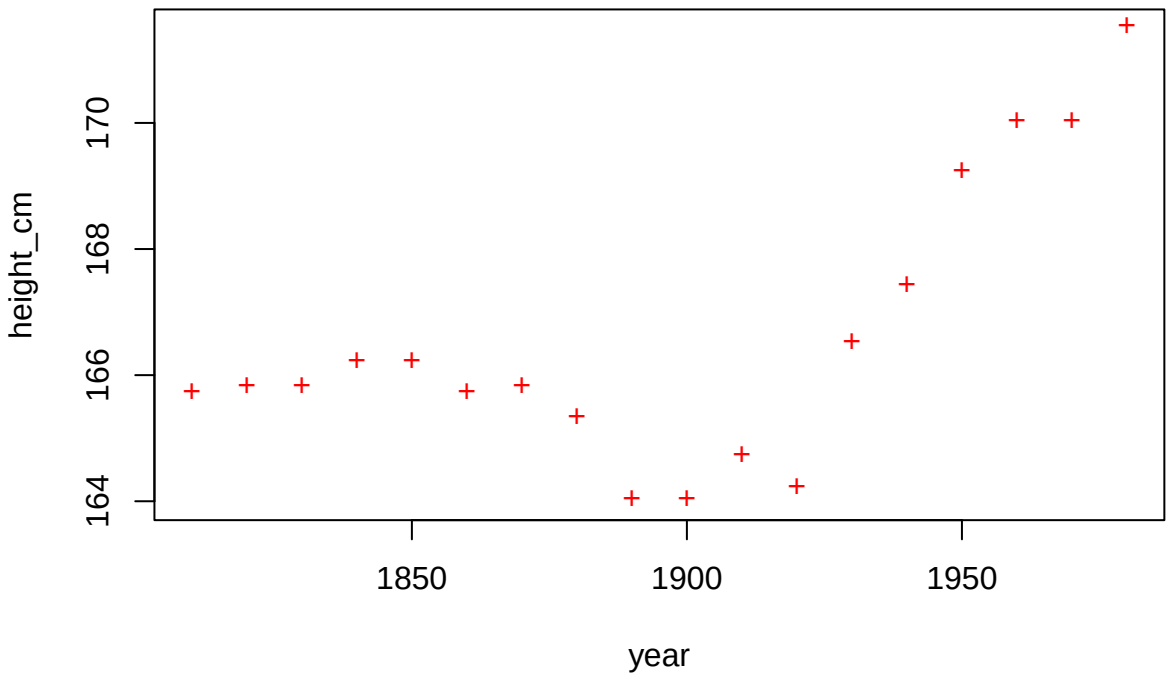
2

```
height_line = function(country_name){
  coun_heights_df = heights_df[heights_df$country == country_name, ]
  plot(coun_heights_df$year, coun_heights_df$height_cm, type = 'l', xlab = 'year', ylab = 'height_cm')
}
height_line('China')
```



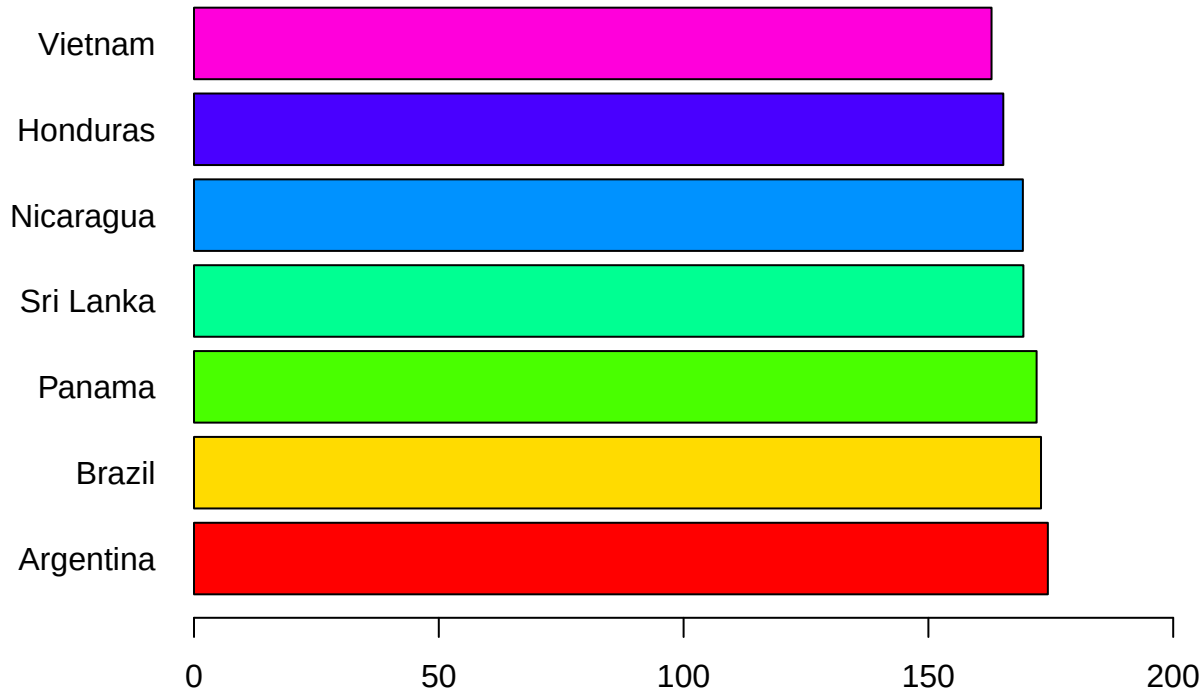
3

```
height_scatter = function(country_name, shape = '.', col = 'blue'){
  coun_heights_df = heights_df[heights_df$country == country_name, ]
  plot(coun_heights_df$year, coun_heights_df$height_cm, type = 'p', xlab = 'year', ylab = 'height_cm',
}
height_scatter('China', shape = '+', col = 'red')
```



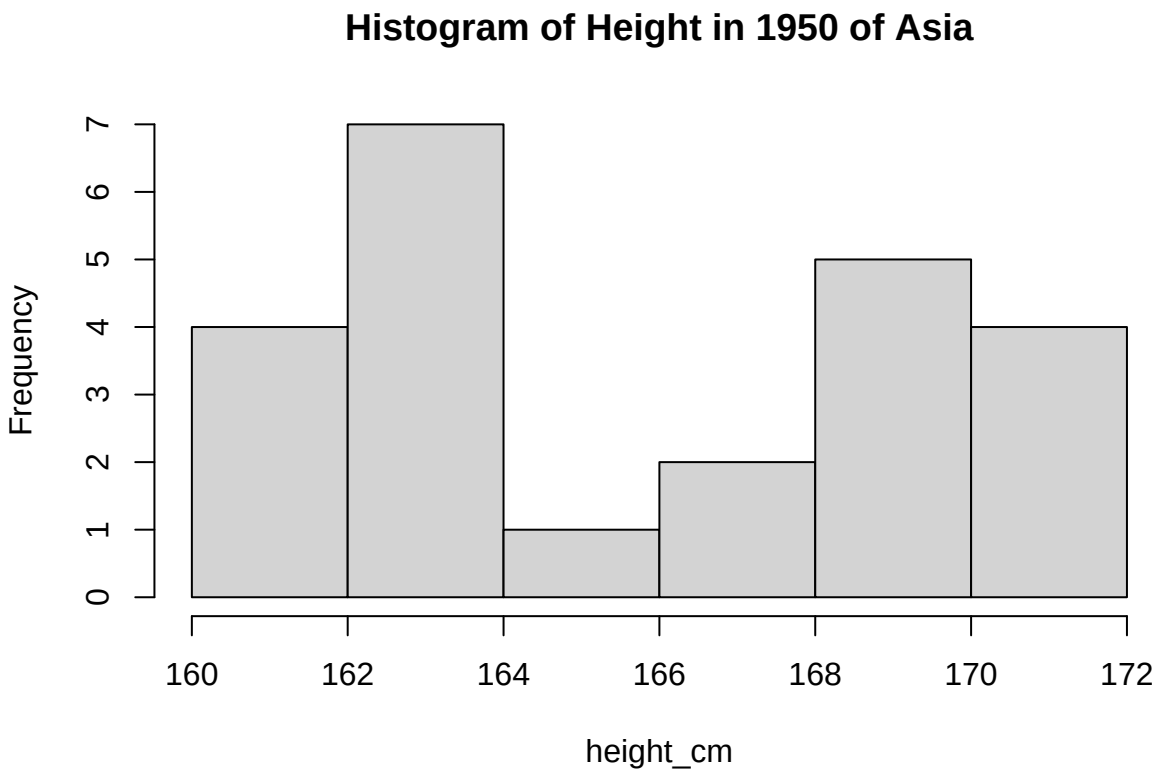
4

```
coun_v = c('Argentina', 'Brazil', 'Honduras', 'Nicaragua', 'Panama', 'Sri Lanka', 'Vietnam')
n = length(coun_v)
coun_heights_df = heights_df[(heights_df$country %in% coun_v) & (heights_df$year == 1990), ]
coun_heights_df = coun_heights_df[order(coun_heights_df$height_cm, decreasing = TRUE), ]
par(mar=c(3,5,3,2))
barplot(coun_heights_df$height_cm, names.arg = coun_heights_df$country, horiz = TRUE, xlab = 'height_cm
```

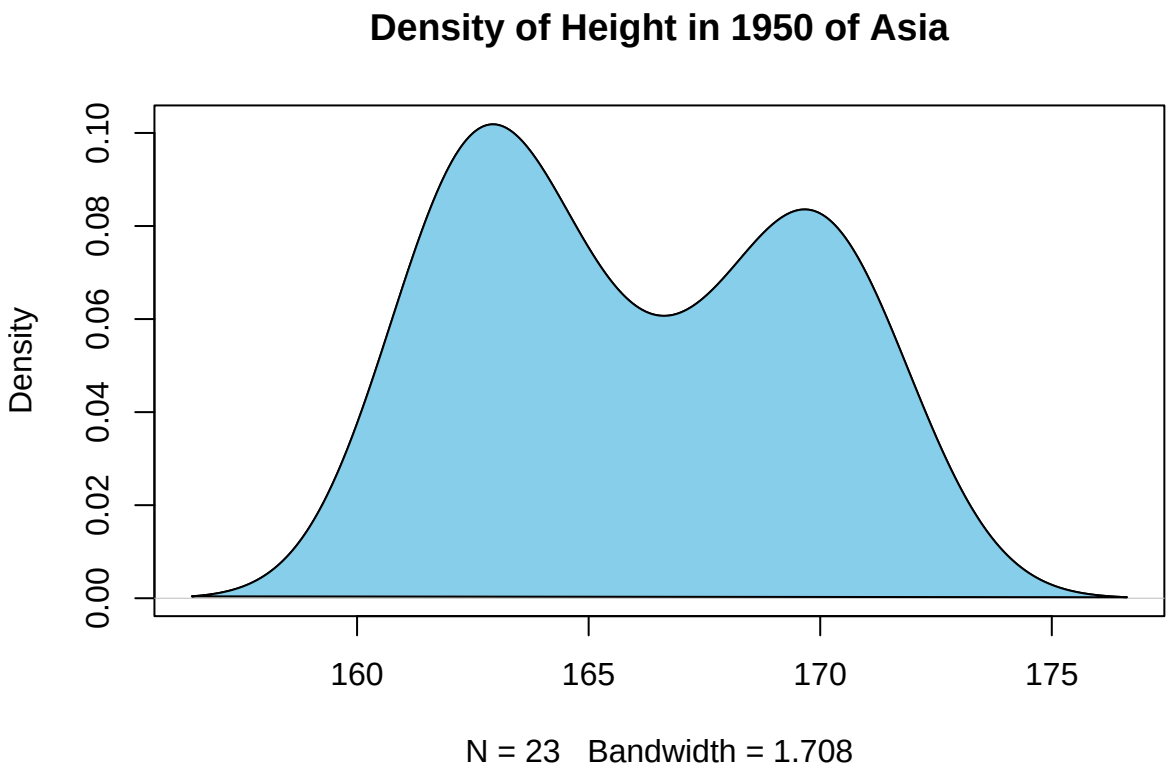


5

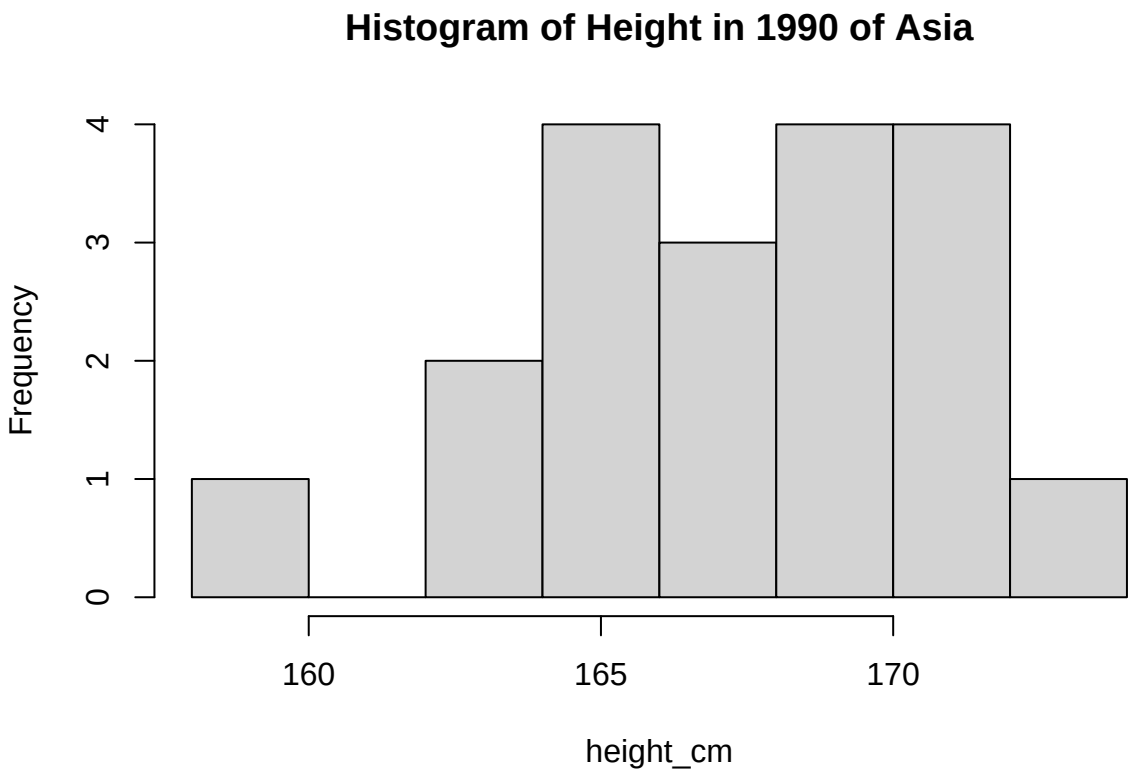
```
heights1950_df = heights_df[(heights_df$continent == 'Asia') & (heights_df$year == 1950), ]
heights1990_df = heights_df[(heights_df$continent == 'Asia') & (heights_df$year == 1990), ]
hist(heights1950_df$height_cm, xlab = 'height_cm', main = 'Histogram of Height in 1950 of Asia')
```



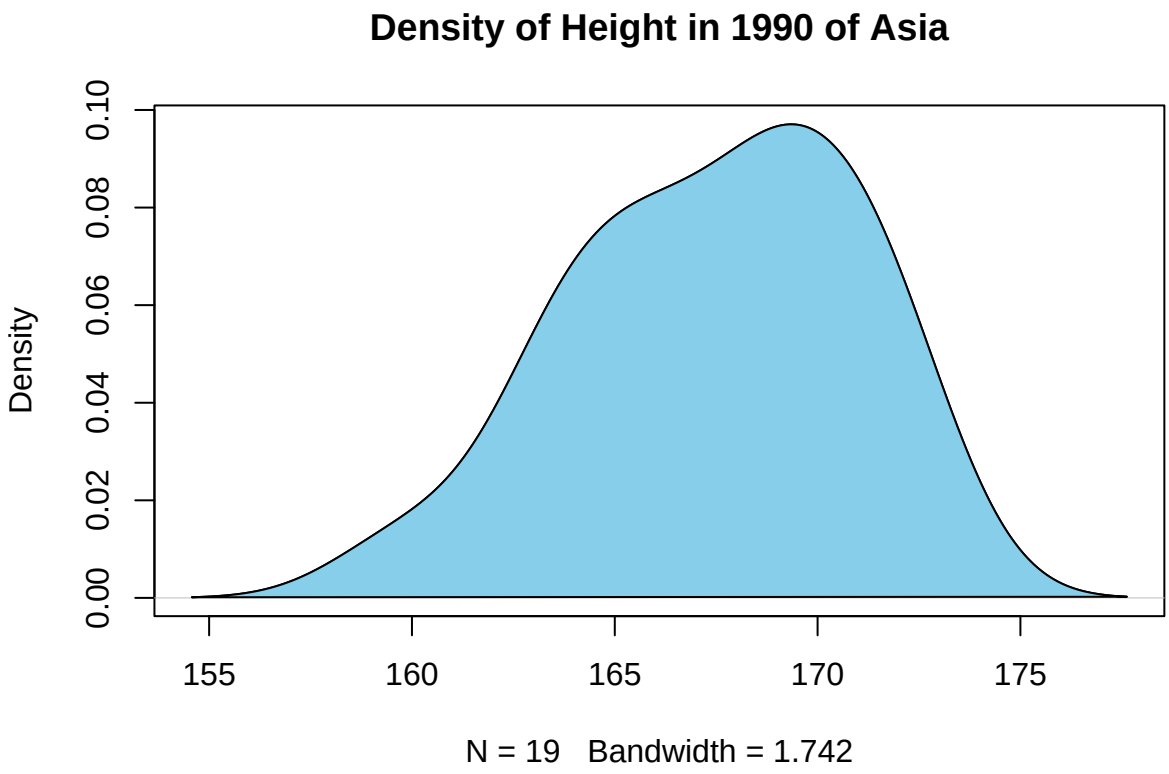

```
plot(density(heights1950_df$height_cm), main = 'Density of Height in 1950 of Asia')
polygon(density(heights1950_df$height_cm), col = 'skyblue')
```




```
hist(heights1990_df$height_cm, xlab = 'height_cm', main = 'Histogram of Height in 1990 of Asia')
```




```
plot(density(heights1990_df$height_cm), main = 'Density of Height in 1990 of Asia')
polygon(density(heights1990_df$height_cm), col = 'skyblue')
```



Exercise 3

```
data_df = read.csv('data/data.csv', skip = 4)
```

1

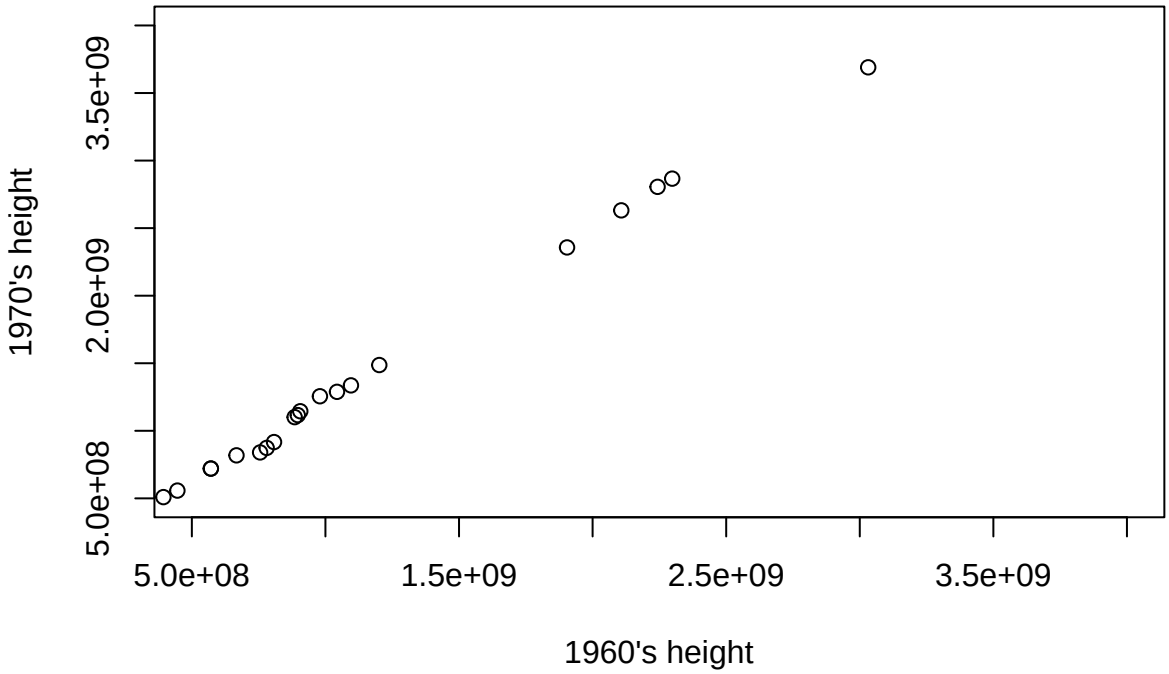
```
data1990_df = data_df[order(data_df$X1990, decreasing = TRUE), ]
top20 = head(data1990_df$Country.Name, n = 20)
top20
```

```
##  [1] "World"
##  [2] "IDA & IBRD total"
##  [3] "Low & middle income"
##  [4] "Middle income"
##  [5] "IBRD only"
##  [6] "Upper middle income"
##  [7] "Early-demographic dividend"
##  [8] "Late-demographic dividend"
##  [9] "Lower middle income"
## [10] "East Asia & Pacific"
```

```
## [11] "East Asia & Pacific (excluding high income)"
## [12] "East Asia & Pacific (IDA & IBRD countries)"
## [13] "South Asia"
## [14] "South Asia (IDA & IBRD)"
## [15] "China"
## [16] "OECD members"
## [17] "High income"
## [18] "Post-demographic dividend"
## [19] "India"
## [20] "IDA total"
```

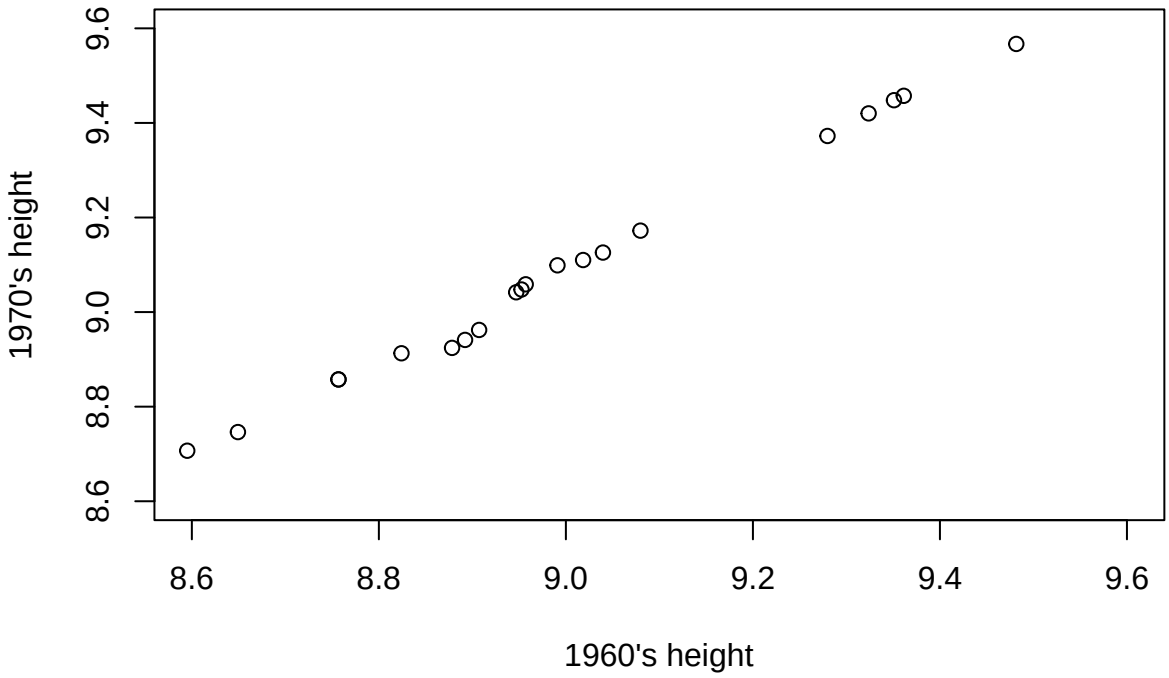
2

```
top20_df = data_df[data_df$Country.Name %in% top20, ]
plot(top20_df$X1960, top20_df$X1970, type ='p', xlab = "1960's height", ylab = "1970's height", xlim = c
```



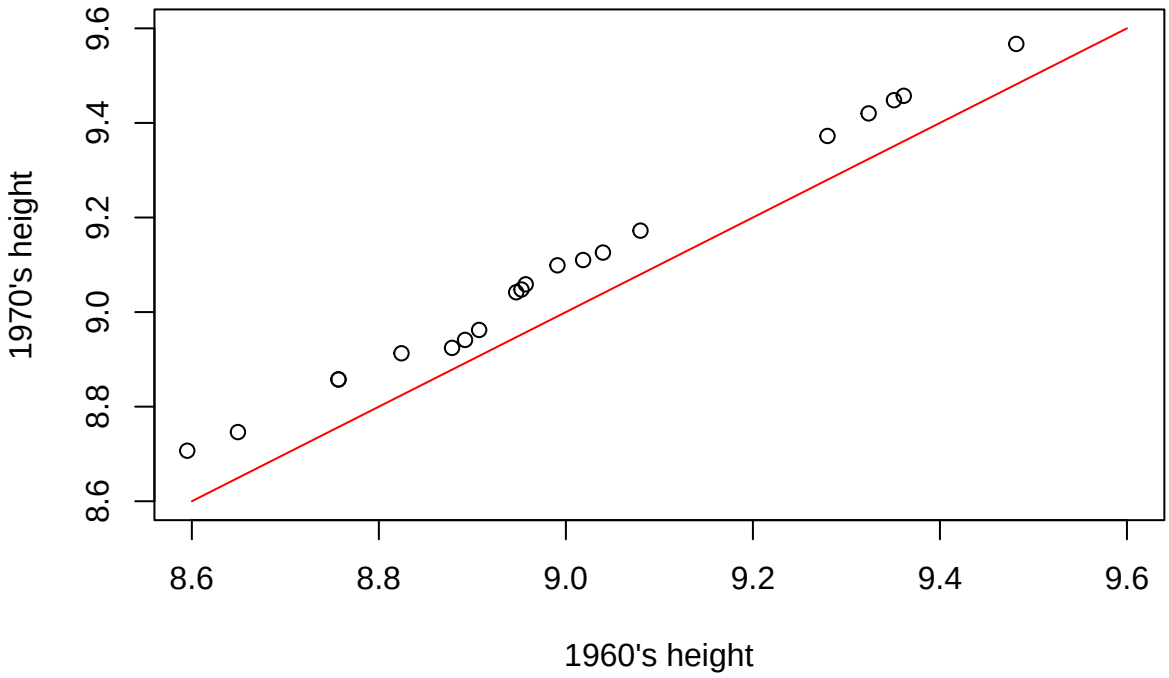
3

```
plot(log(top20_df$X1960, base = 10), log(top20_df$X1970, base = 10), type ='p', xlab = "1960's height",
```



4

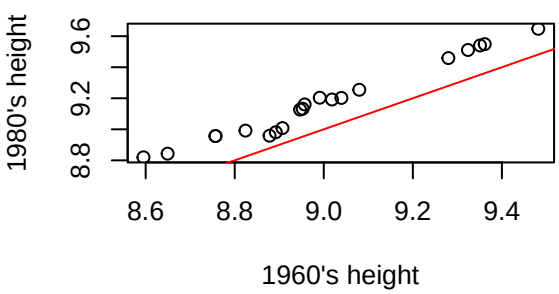
```
plot(log(top20_df$X1960, base = 10), log(top20_df$X1970, base = 10), type ='p', xlab = "1960's height",
lines(x = seq(8.6, 9.6, 0.1), y = seq(8.6, 9.6, 0.1), col = 'red', type = 'l')
```

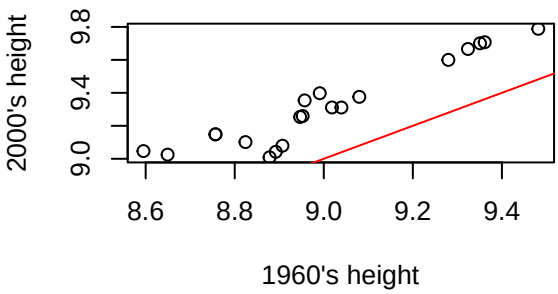


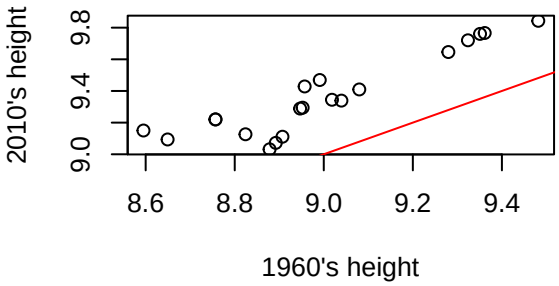
All of top20 countries experienced a population growth.

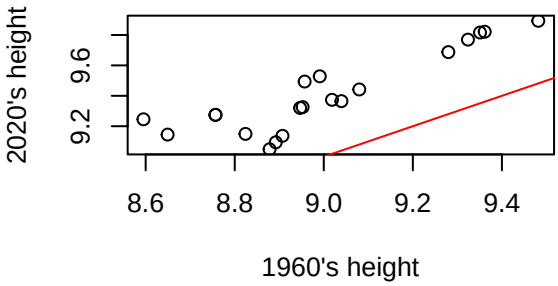
5

```
par(mfrow = c(2, 2))
plot(log(top20_df$X1960, base = 10), log(top20_df$X1980, base = 10), type ='p', xlab = "1960's height",
lines(x = seq(8, 10, 0.1), y = seq(8, 10, 0.1), col = 'red', type = 'l')
plot(log(top20_df$X1960, base = 10), log(top20_df$X2000, base = 10), type ='p', xlab = "1960's height",
lines(x = seq(8, 10, 0.1), y = seq(8, 10, 0.1), col = 'red', type = 'l')
plot(log(top20_df$X1960, base = 10), log(top20_df$X2010, base = 10), type ='p', xlab = "1960's height",
lines(x = seq(8, 10, 0.1), y = seq(8, 10, 0.1), col = 'red', type = 'l')
plot(log(top20_df$X1960, base = 10), log(top20_df$X2020, base = 10), type ='p', xlab = "1960's height",
lines(x = seq(8, 10, 0.1), y = seq(8, 10, 0.1), col = 'red', type = 'l')
```







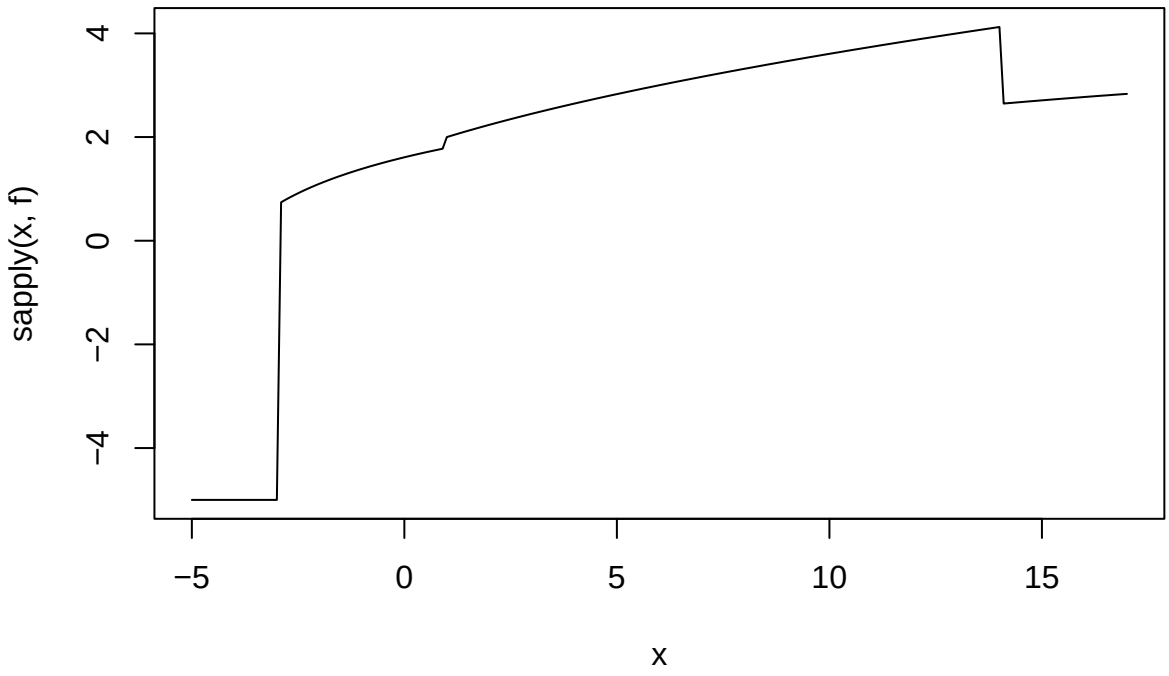


Exercise 4

1

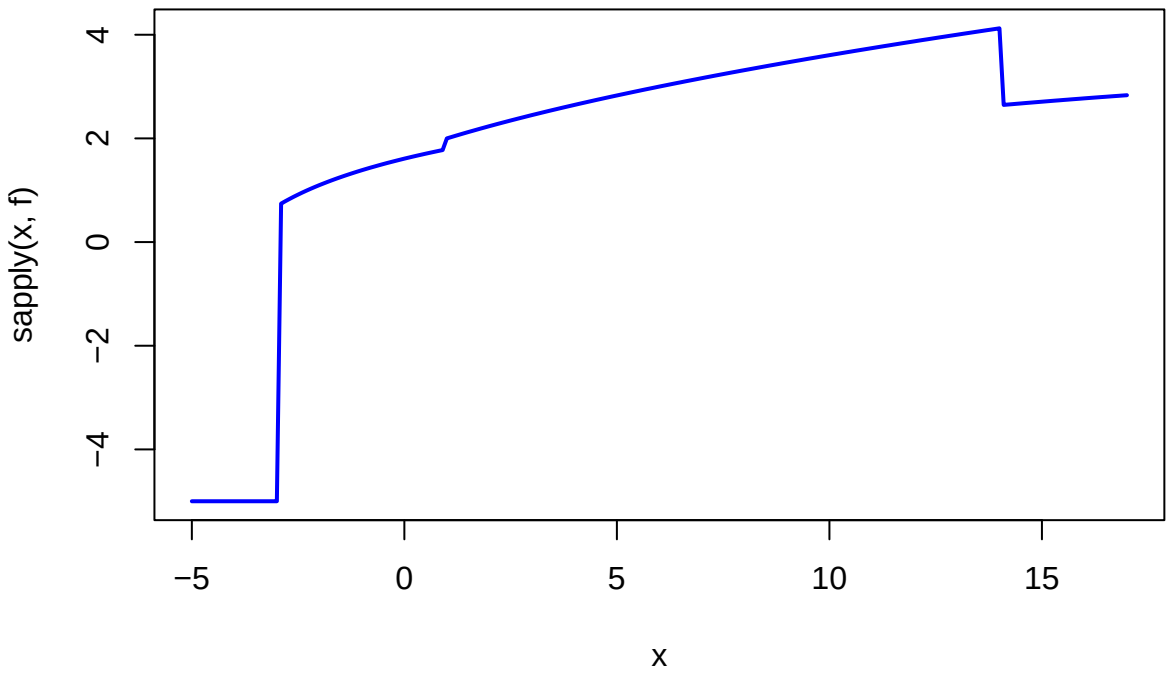
```
f = function (x){
  if (x <= -3){
    result = -5
  }
  else if((x > -3) & (x < 1)){
    result = log(x+5)
  }
  else if (x == 1){
    result = 2
  }
  else if ((x > 1) & (x <= 14)){
    result = (x+3)**0.5
  }
  else {
    result = log(x)
  }
  return(result)
}
```

```r
x = seq(-5, 17, 0.1)
plot(x, sapply(x, f), type = 'l')
```



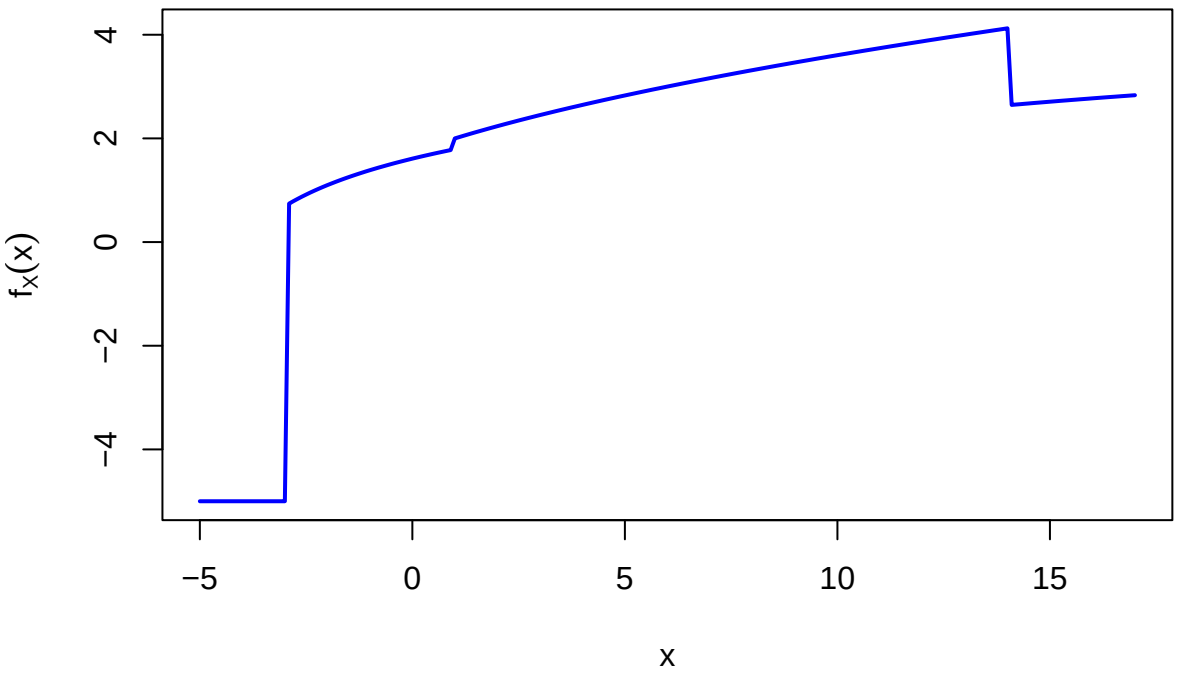
2

```r
plot(x, sapply(x, f), type = 'l', col = 'blue', lwd = 2)
```



3

```
plot(x, sapply(x, f), type = 'l', col = 'blue', lwd = 2, ylab = expression(f[X](x)))
```



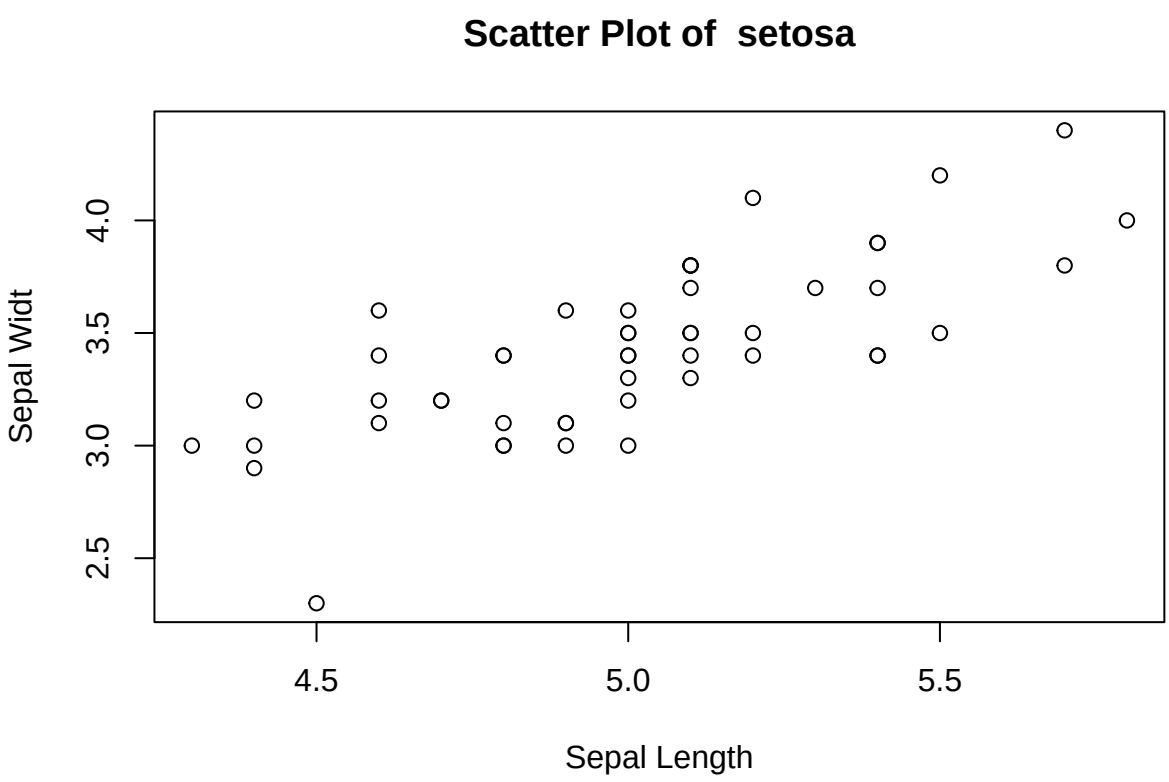
Exercise 5

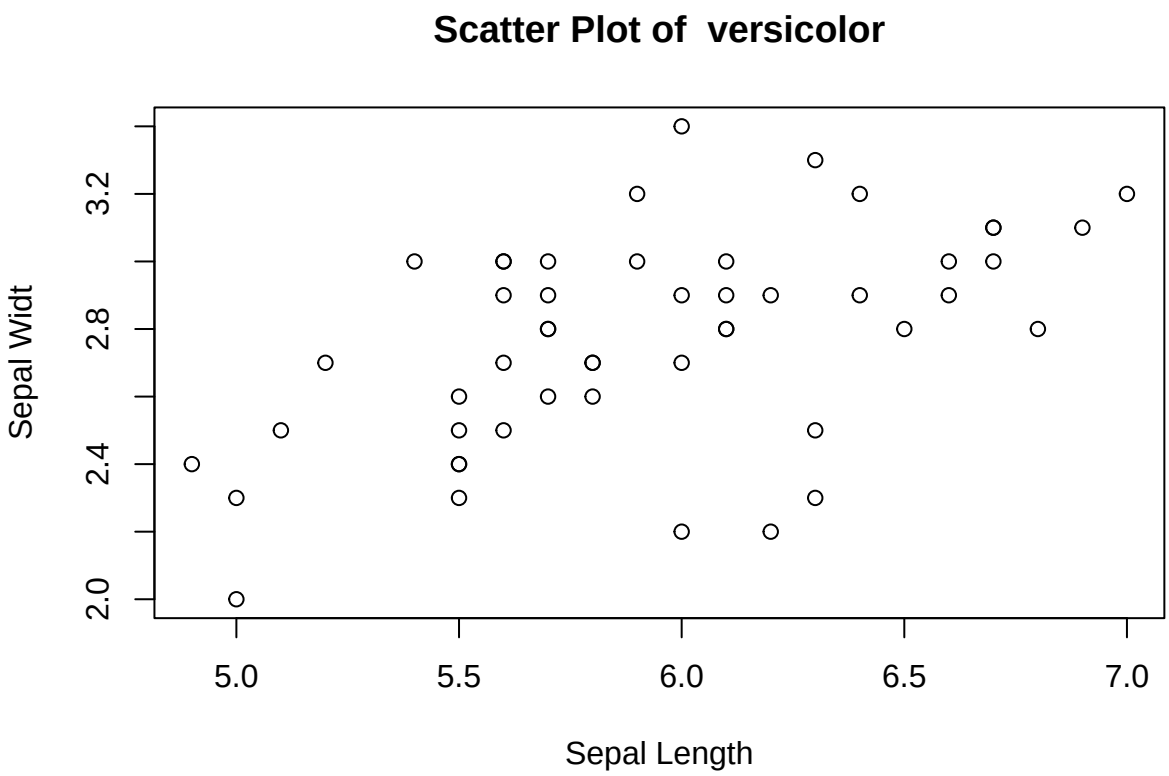
1

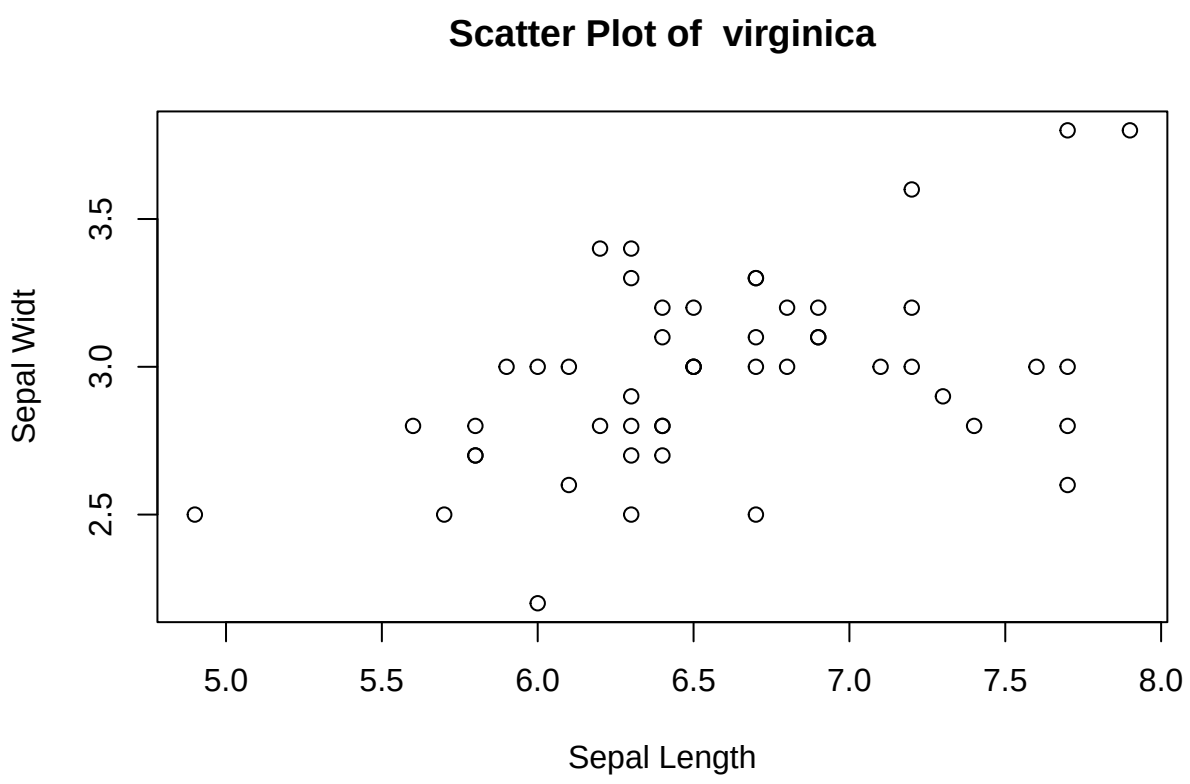
```
splitiris = split(iris, iris$Species)
```

2

```
n = length(splitiris)
for (i in 1:n){
  i_df = splitiris[[i]]
  plot(x = i_df$Sepal.Length, y = i_df$Sepal.Width, xlab = 'Sepal Length', ylab = 'Sepal Widt', main = p
}
```


Scatter Plot of setosa
Sepal Widt
2.5
3.0
3.5
4.0
4.5
5.0
5.5
Sepal Length

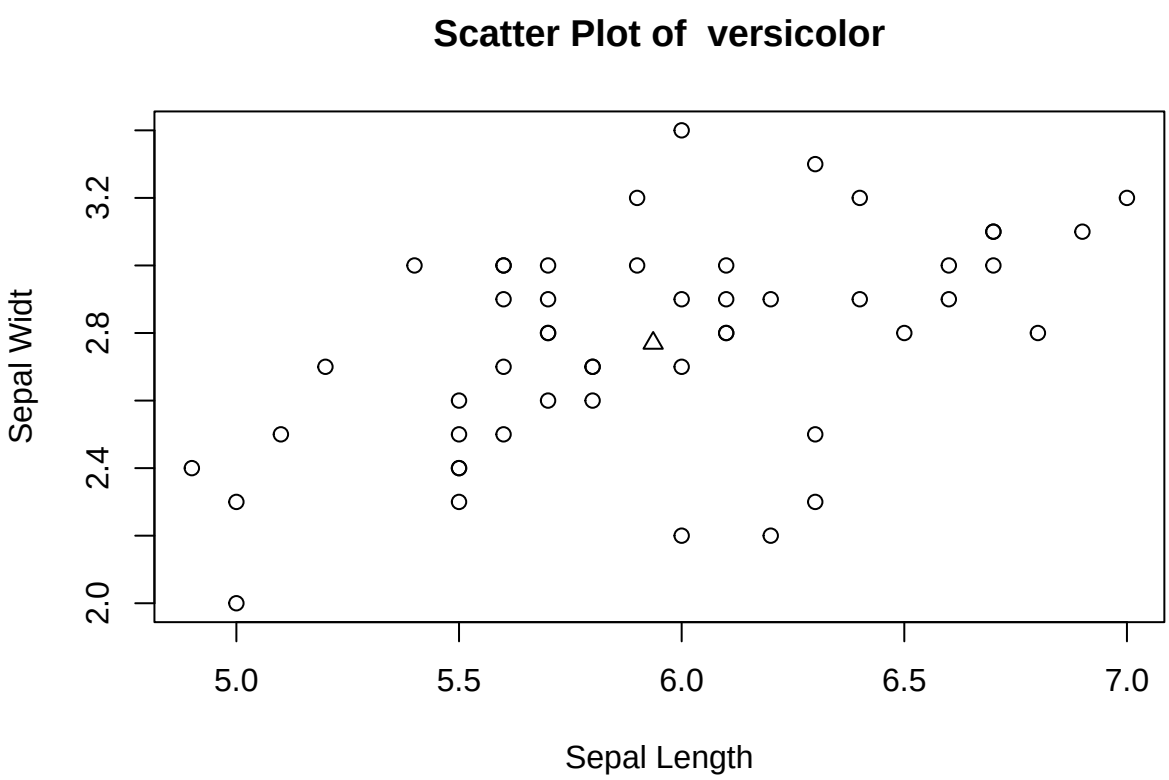

Scatter Plot of versicolor
Sepal Widt
2.0
2.4
2.8
3.2
5.0
5.5
6.0
6.5
7.0
Sepal Length

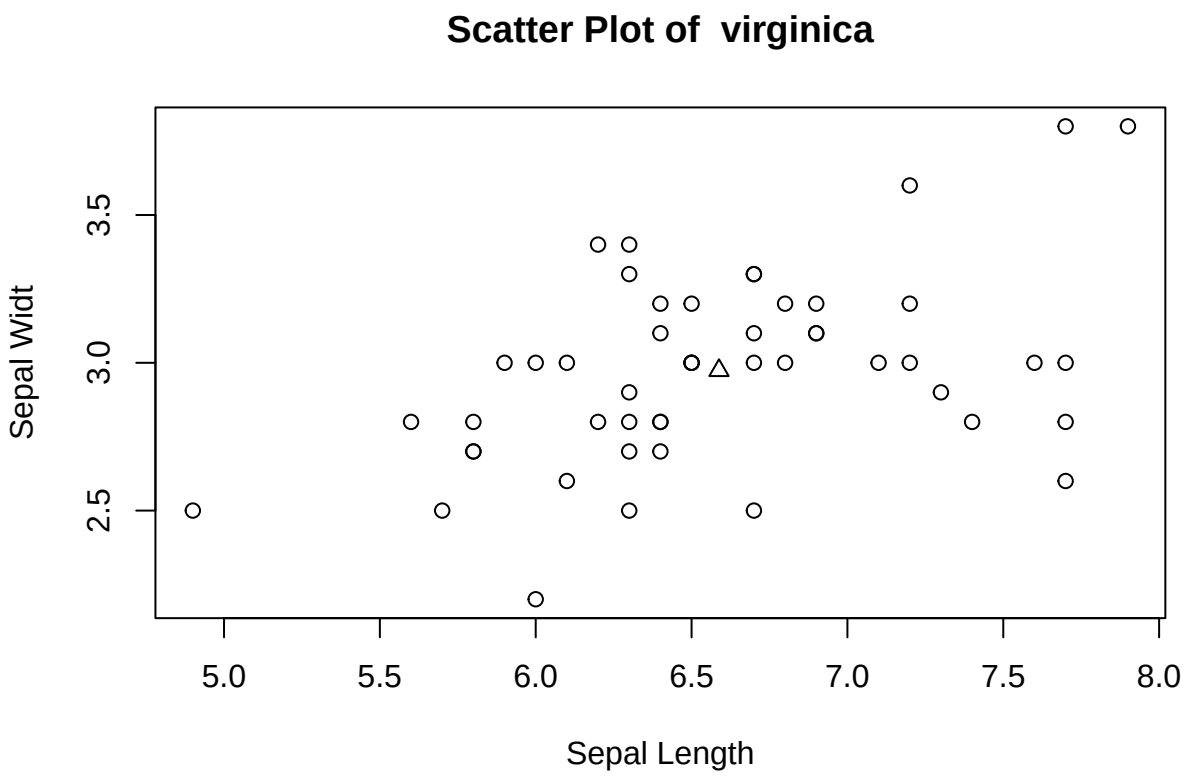


3

```
n = length(splitiris)
for (i in 1:n){
  i_df = splitiris[[i]]
  plot(x = i_df$Sepal.Length, y = i_df$Sepal.Width, xlab = 'Sepal Length', ylab = 'Sepal Widt', main = 
  points(x = mean(i_df$Sepal.Length), y = mean(i_df$Sepal.Width), pch = 24,col="black")
}
```


Scatter Plot of setosa
Sepal Widt
2.5
3.0
3.5
4.0
4.5
5.0
5.5
Sepal Length


Scatter Plot of versicolor
Sepal Widt
2.0
2.4
2.8
3.2
5.0
5.5
6.0
6.5
7.0
Sepal Length



4

```
splitiris1 = splitiris[-which(names(splitiris) == 'setosa')]
```