

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])
```

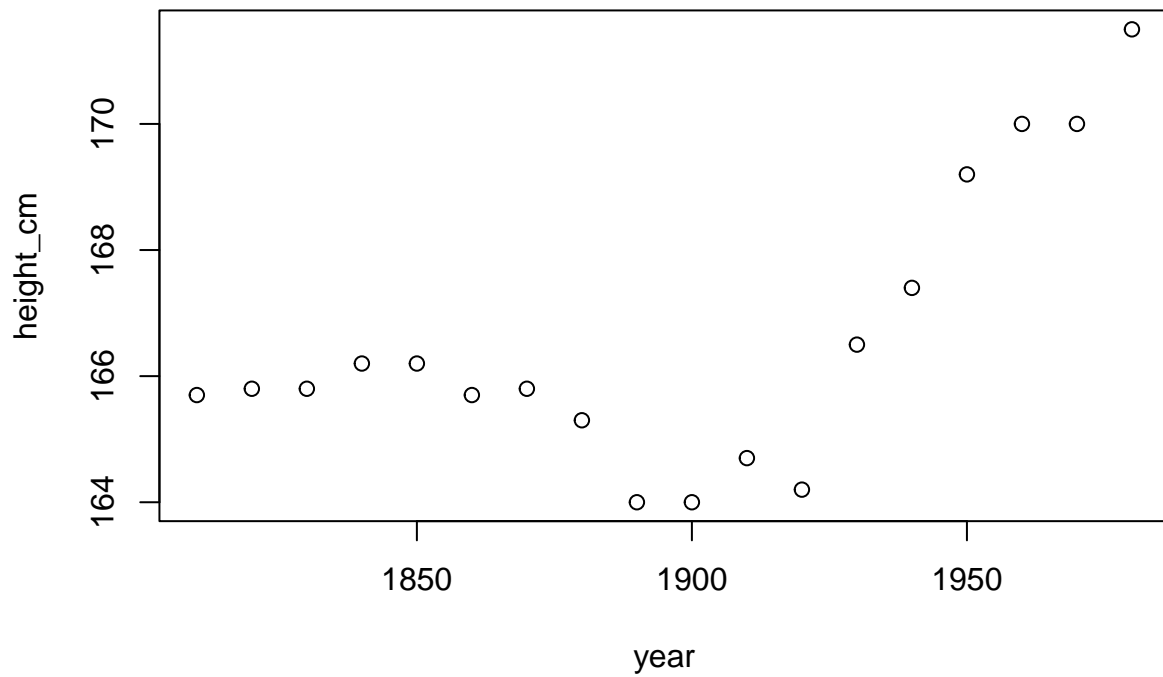
```
## [1] "Sepal.Length"
```

Exercise 2

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

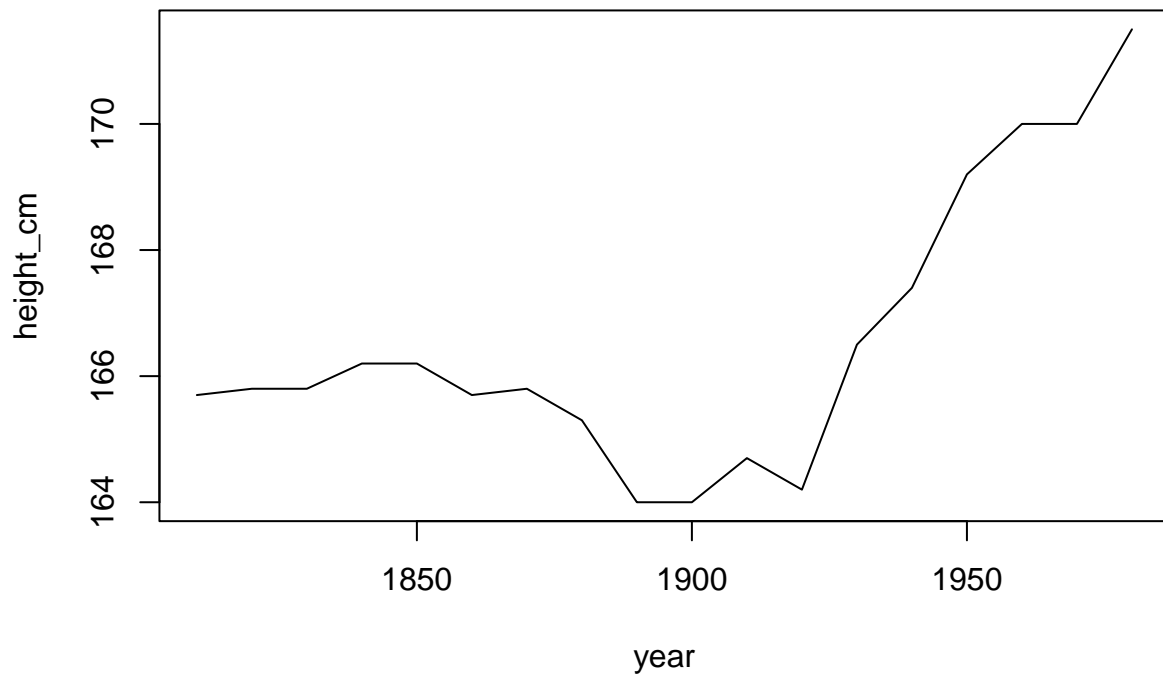
1

```
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')
```



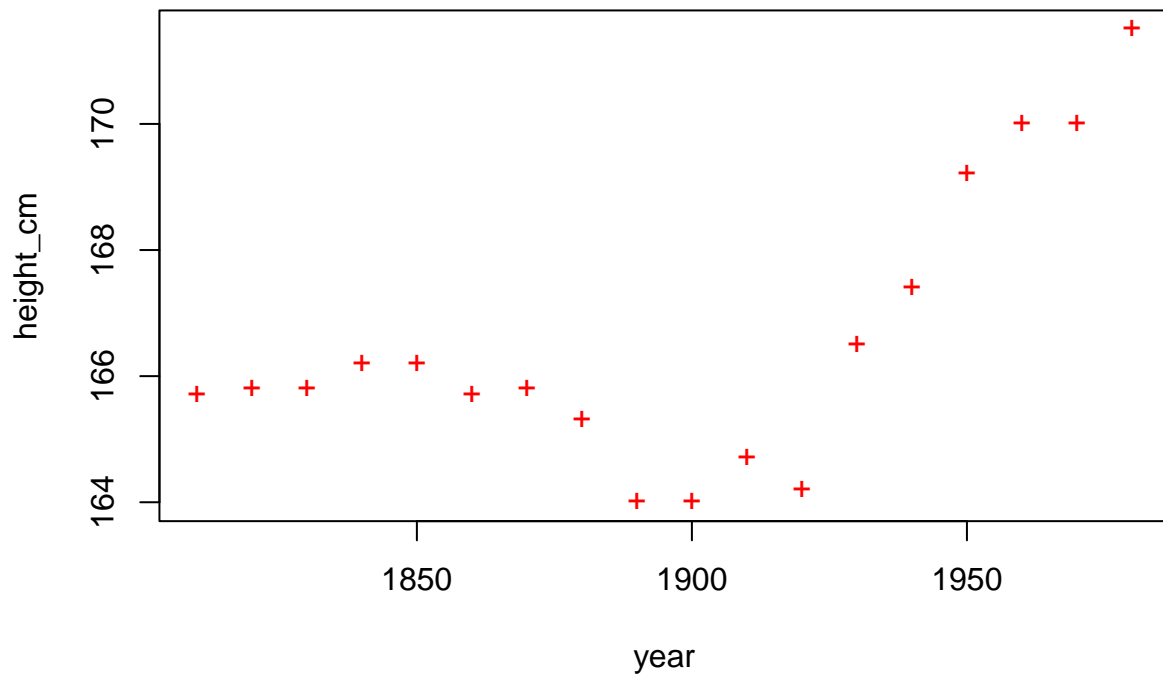
2

```
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')
```



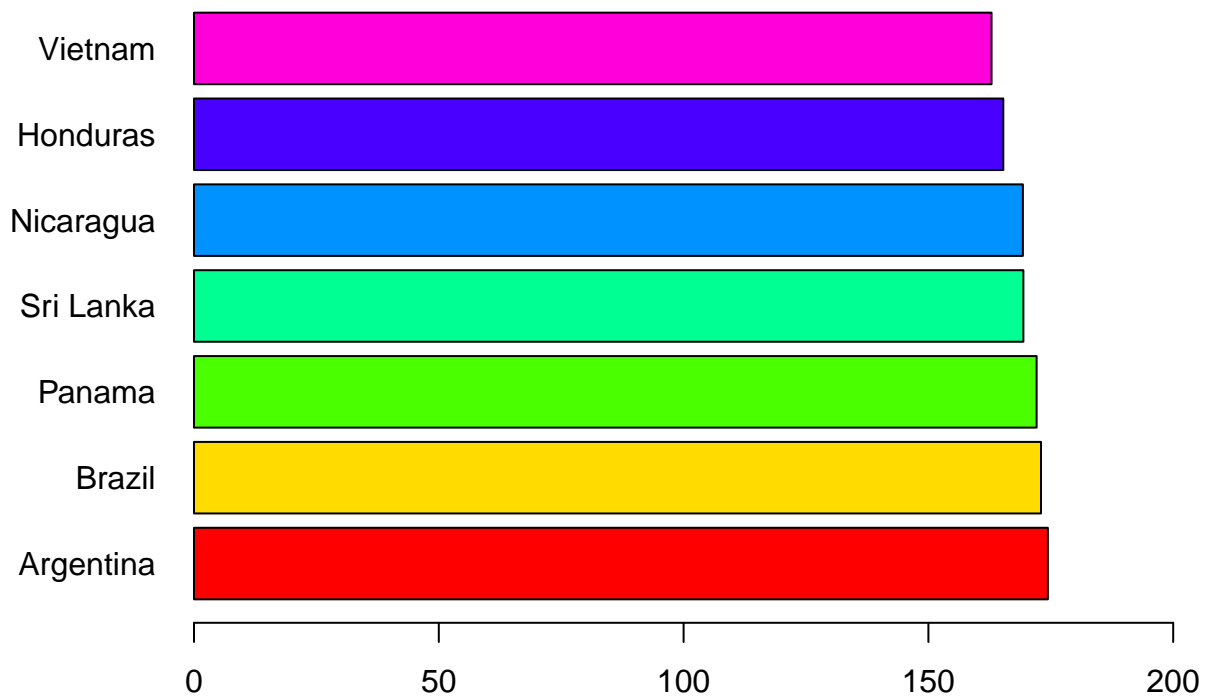
3

```
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', col = col, shape = shape)  
}  
height_scatter('China', shape = '+', col = 'red')
```



4

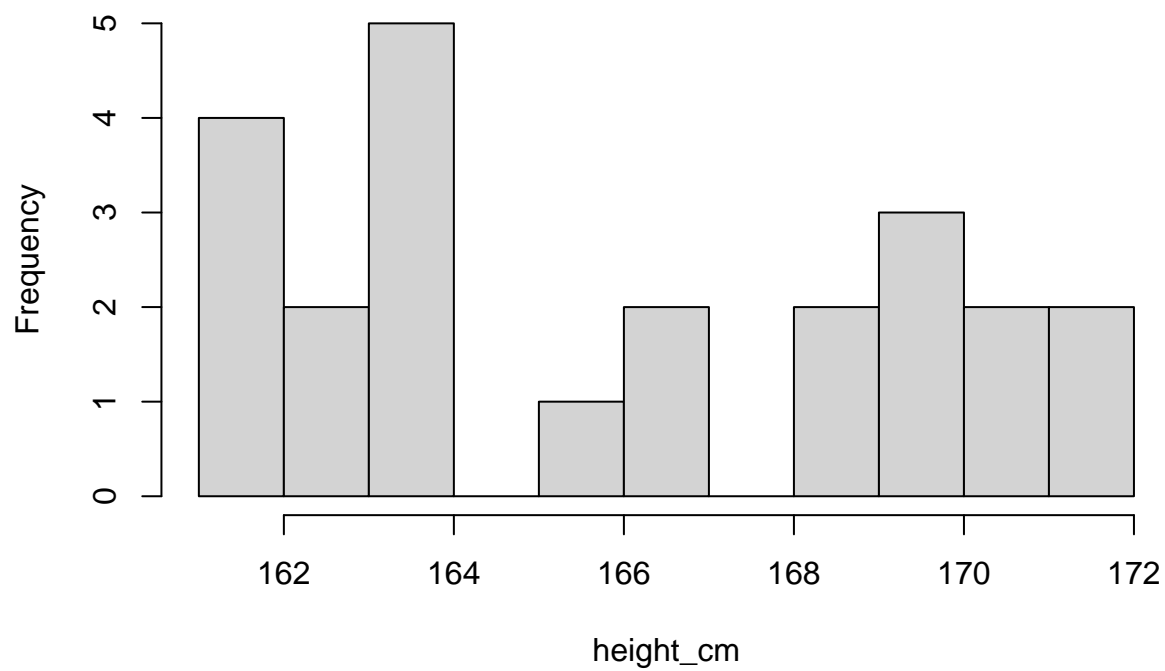
```
coun_v = c('Argentina', 'Brazil', 'Honduras', 'Nicaragua', 'Panama', 'Sri Lanka', 'Vietnam')
n = length(coun_v)
coun_heights_df = heights_df[(heights_df$country %in% coun_v) & (heights_df$year == 1990), ]
coun_heights_df = coun_heights_df[order(coun_heights_df$height_cm, decreasing = TRUE), ]
par(mar=c(3,5,3,2))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



5

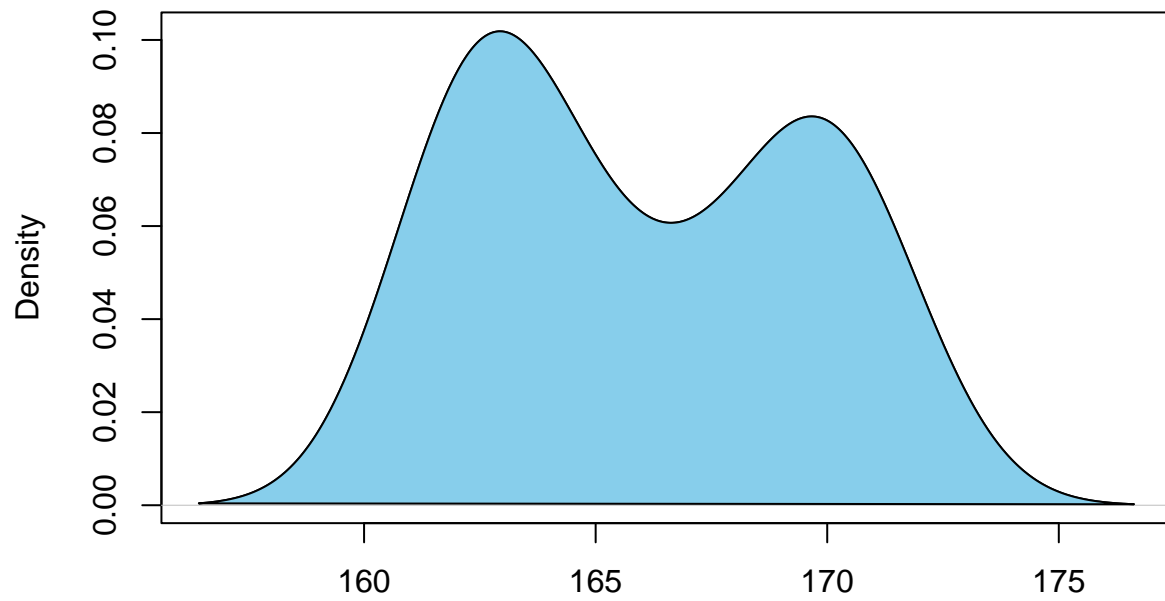
```
heights1950_df = heights_df[(heights_df$continent == 'Asia') & (heights_df$year == 1950), ]  
heights1990_df = heights_df[(heights_df$continent == 'Asia') & (heights_df$year == 1990), ]  
hist(heights1950_df$height_cm, breaks = 10, xlab = 'height_cm', main = 'Histogram of Height in 1950 of A
```

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')
```

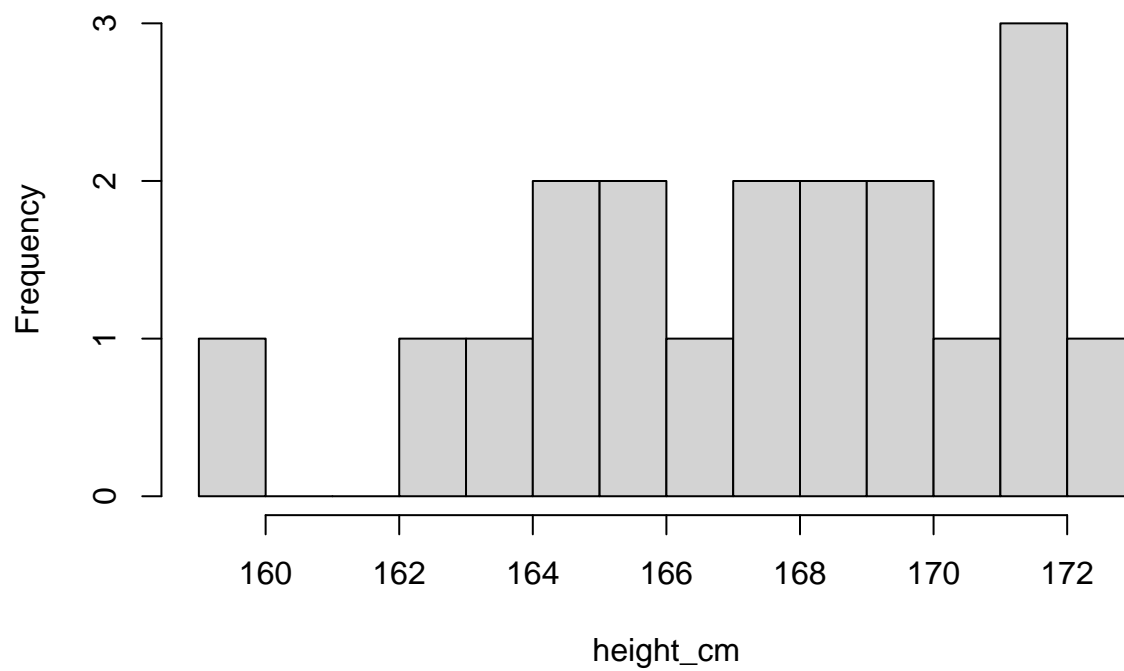
Density of Height in 1950 of Asia



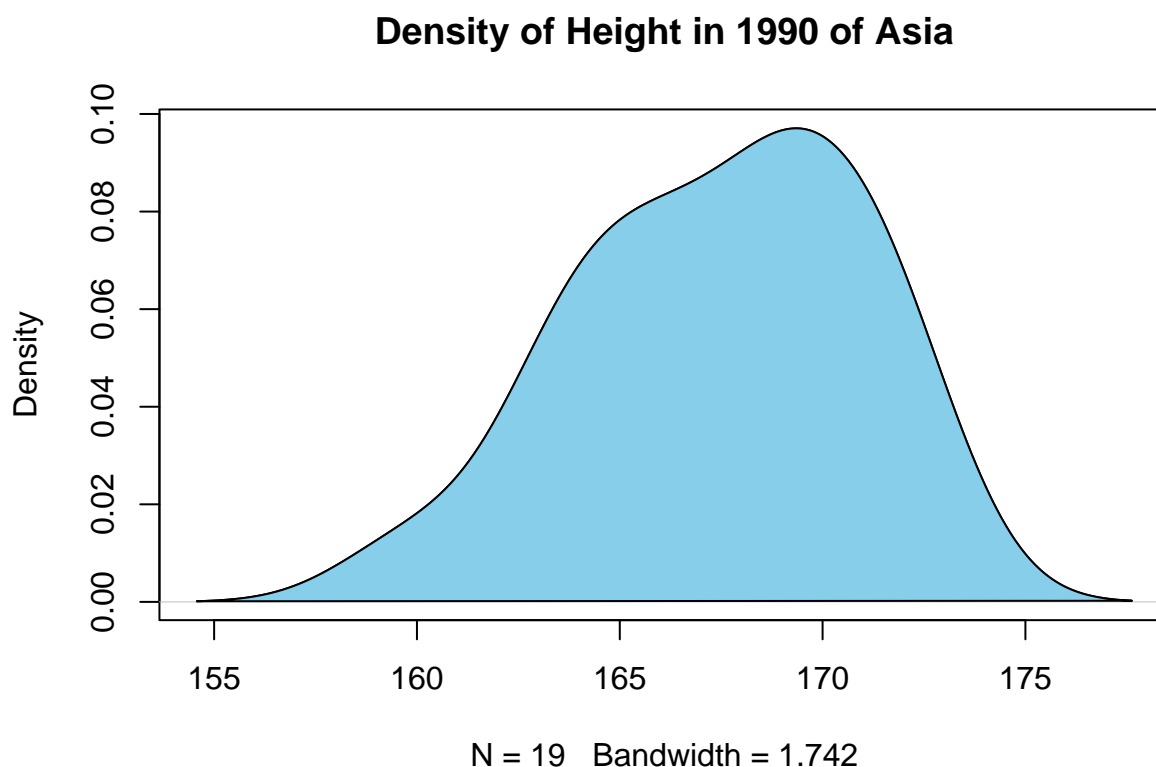
N = 23 Bandwidth = 1.708

```
hist(heights1990_df$height_cm, breaks = 10, xlab = 'height_cm', main = 'Histogram of Height in 1990 of Asia')
```


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)
notCountry = c("World", "IDA & IBRD total", "Low & middle income", "Middle income", "IBRD only",
               "Early-demographic dividend", "Lower middle income", "Upper middle income", "Late-demogr",
               "East Asia & Pacific (excluding high income)",
               "East Asia & Pacific (IDA & IBRD countries)", "South Asia",
               "South Asia (IDA & IBRD)", "OECD members", "High income",
               "Post-demographic dividend", "Europe & Central Asia", "IDA total", "IDA only",
               "Sub-Saharan Africa", "Sub-Saharan Africa (IDA & IBRD countries)",
               "Sub-Saharan Africa (excluding high income)",
               "Least developed countries: UN classification", "Latin America & Caribbean",
               "Fragile and conflict affected situations",
               "Europe & Central Asia (IDA & IBRD countries)",
               "Latin America & the Caribbean (IDA & IBRD countries)", "European Union",
               "Pre-demographic dividend", "Latin America & Caribbean (excluding high income)",
               "Europe & Central Asia (excluding high income)",
               "Heavily indebted poor countries (HIPC)", "Euro area", "Africa Eastern and Southern",
               "Low income", "IDA blend", "North America", "Middle East & North Africa",
               "Middle East & North Africa (excluding high income)",
               "Middle East & North Africa (IDA & IBRD countries)", "Arab World",
               "Africa Western and Central", "Central Europe and the Baltics")
data_df1 = data_df[! data_df$Country.Name %in% notCountry, ]
```

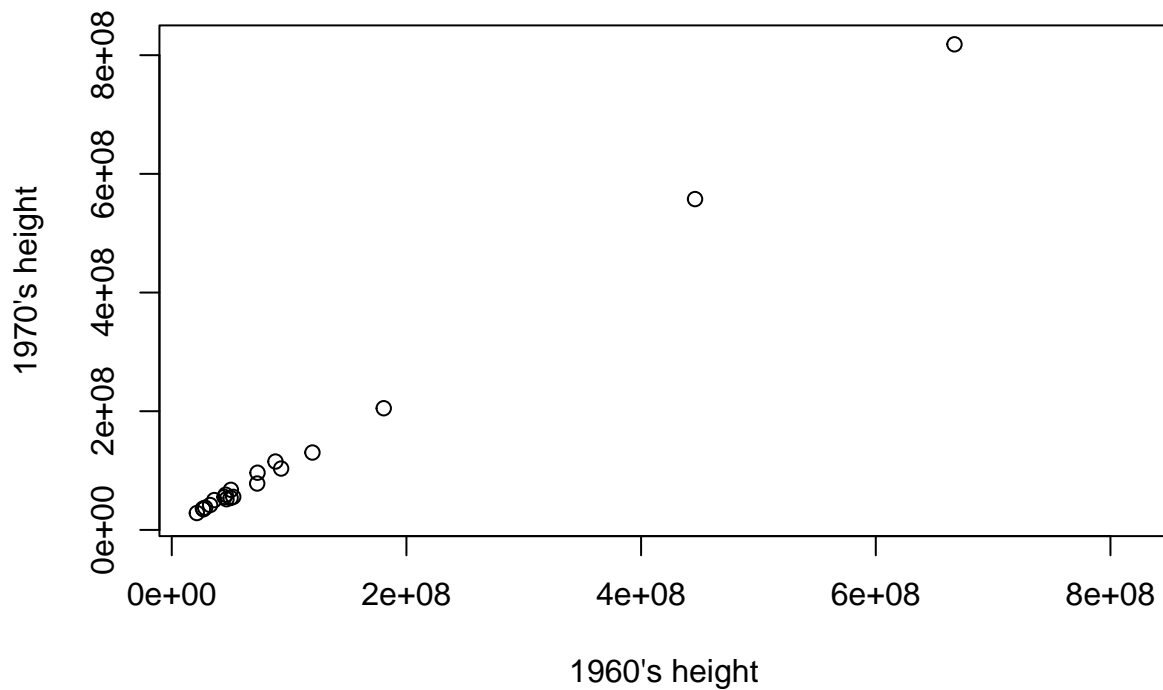
1

```
data1990_df = data_df1[order(data_df1$X1990, decreasing = TRUE), ]
top20 = head(data1990_df$Country.Name, n = 20)
top20
```

```
## [1] "China"          "India"          "United States"
## [4] "Indonesia"      "Brazil"         "Russian Federation"
## [7] "Japan"          "Pakistan"       "Bangladesh"
## [10] "Nigeria"       "Mexico"        "Germany"
## [13] "Vietnam"       "Philippines"   "France"
## [16] "United Kingdom" "Egypt, Arab Rep." "Italy"
## [19] "Iran, Islamic Rep." "Thailand"
```

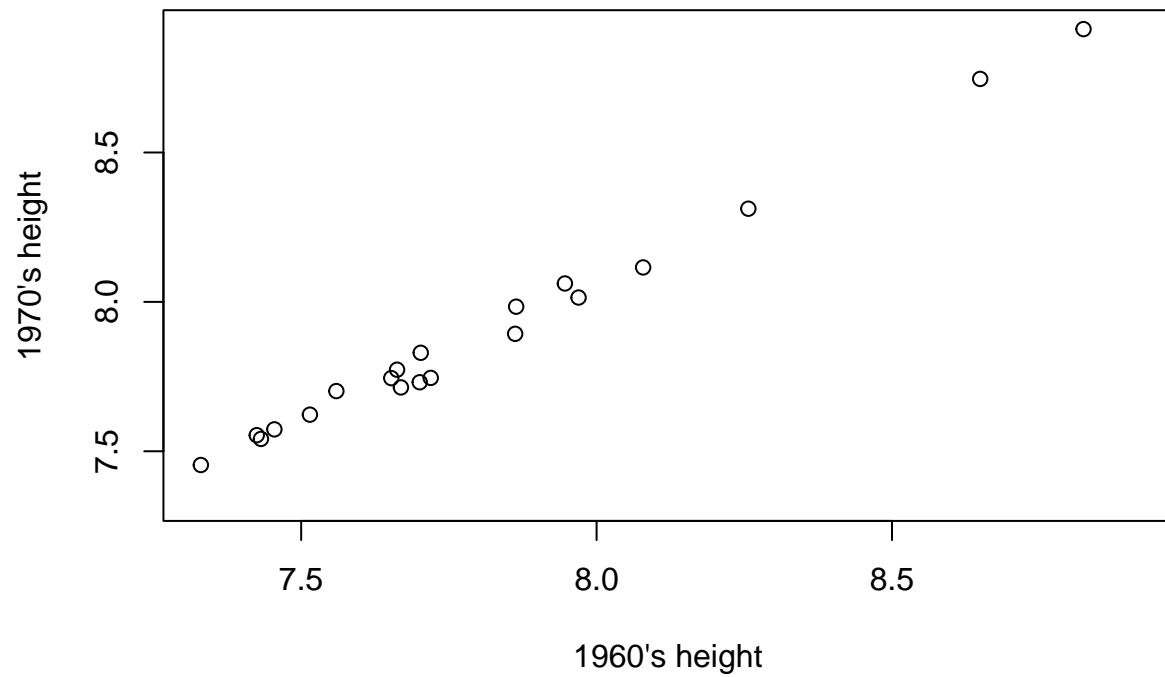
2

```
top20_df = data_df1[data_df1$Country.Name %in% top20, ]
min_ = min(top20_df[, c('X1960', 'X1970')])
max_ = max(top20_df[, c('X1960', 'X1970')])
plot(top20_df$X1960, top20_df$X1970, type = 'p', xlab = "1960's height", ylab = "1970's height", xlim = c(0, max_), ylim = c(0, max_))
```



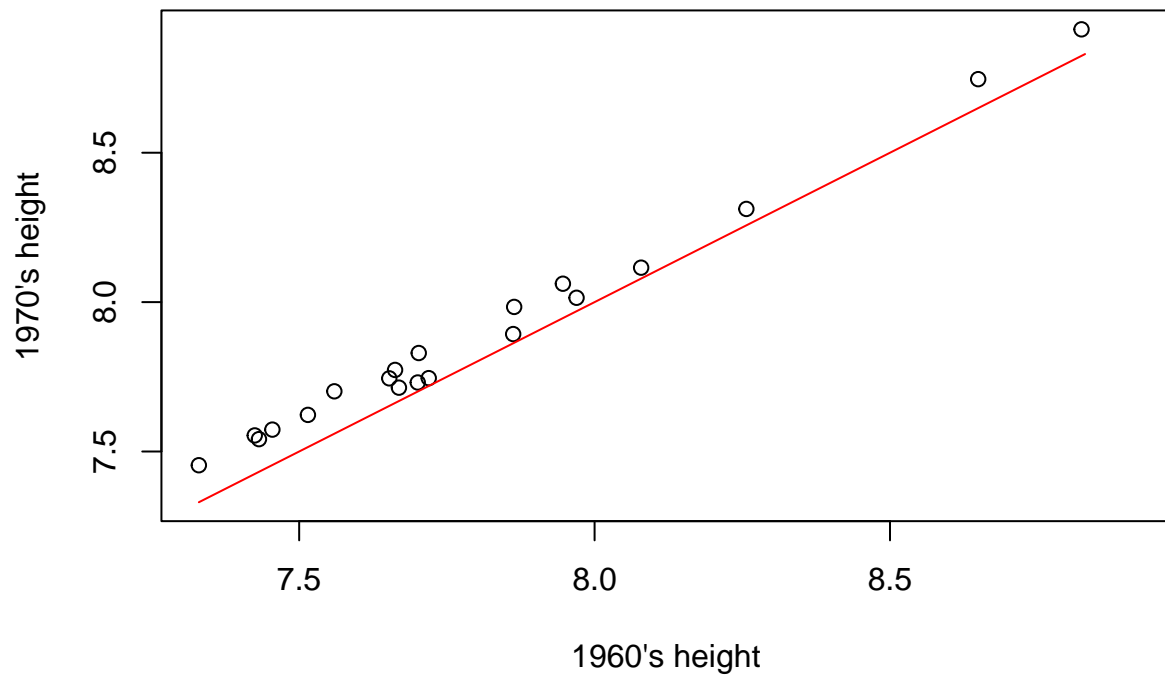
3

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



4

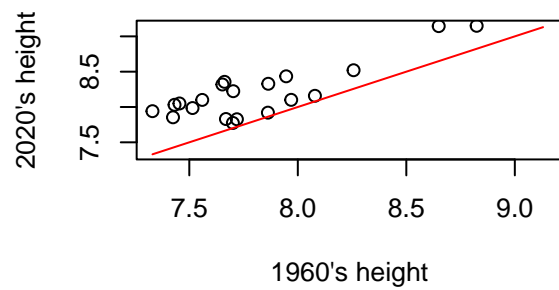
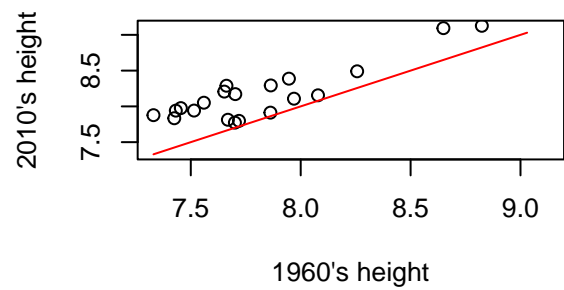
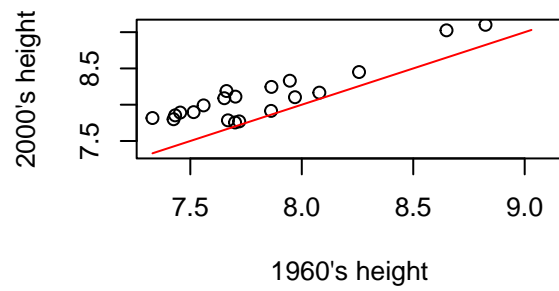
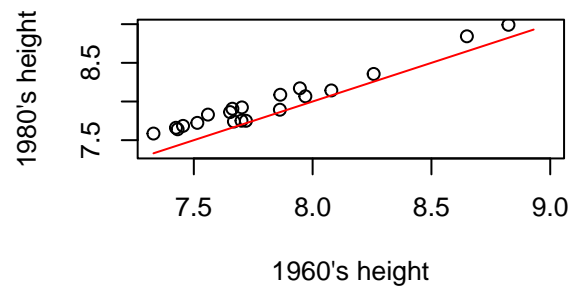
```
plot(log(top20_df$X1960, base = 10), log(top20_df$X1970, base = 10), type = 'p', xlab = "1960's height",
s = seq(log_min, log_max, 0.1)
lines(x = s, y = s, col = 'red', type = 'l')
```



All of top20 countries experienced a population growth.

5

```
par(mfrow = c(2, 2))
for (c in c('X1980', 'X2000', 'X2010', 'X2020')){
  c_log_min = log(min(top20_df[, c('X1960', c)]), base = 10)
  c_log_max = log(max(top20_df[, c('X1960', c)]), base = 10)
  plot(log(top20_df$X1960, base = 10), log(top20_df[, c], base = 10), type = 'p', xlab = "1960's height")
  s = seq(c_log_min, c_log_max, 0.1)
  lines(x = s, y = s, col = 'red', type = 'l')
}
```

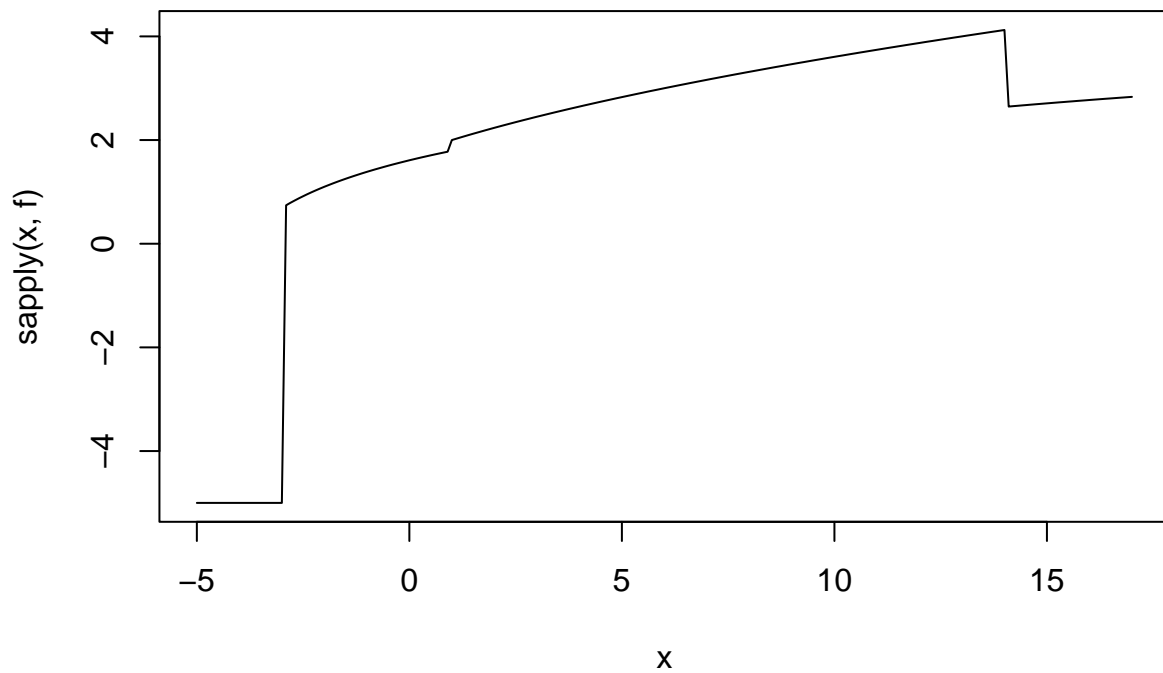


Exercise 4

1

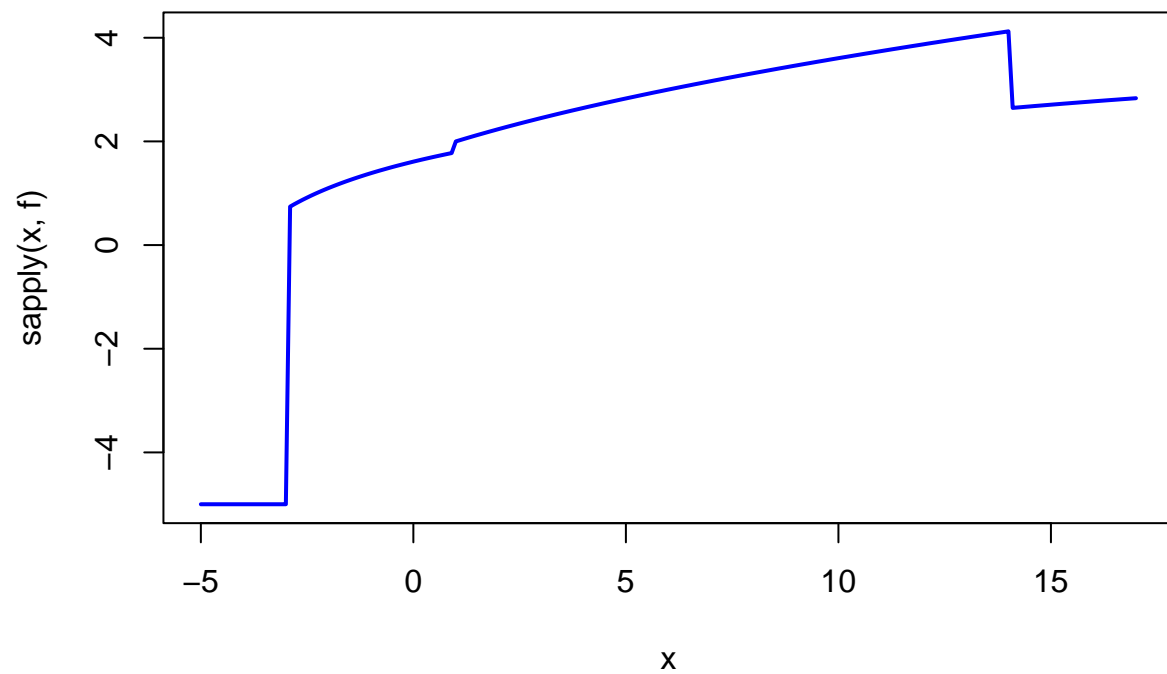
```
f = function (x){
  if (x <= -3){
    result = -5
  }
  else if((x > -3) & (x < 1)){
    result = log(x+5)
  }
  else if (x == 1){
    result = 2
  }
  else if ((x > 1) & (x <= 14)){
    result = (x+3)**0.5
  }
  else {
    result = log(x)
  }
  return(result)
}
```

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



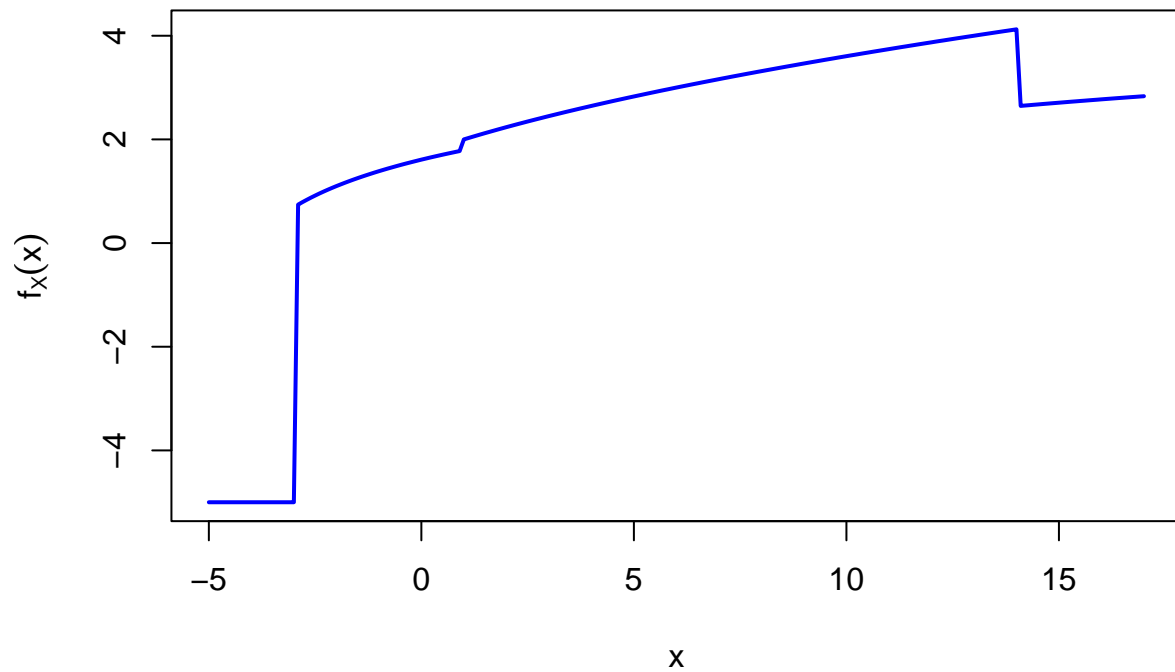
2

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



3

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

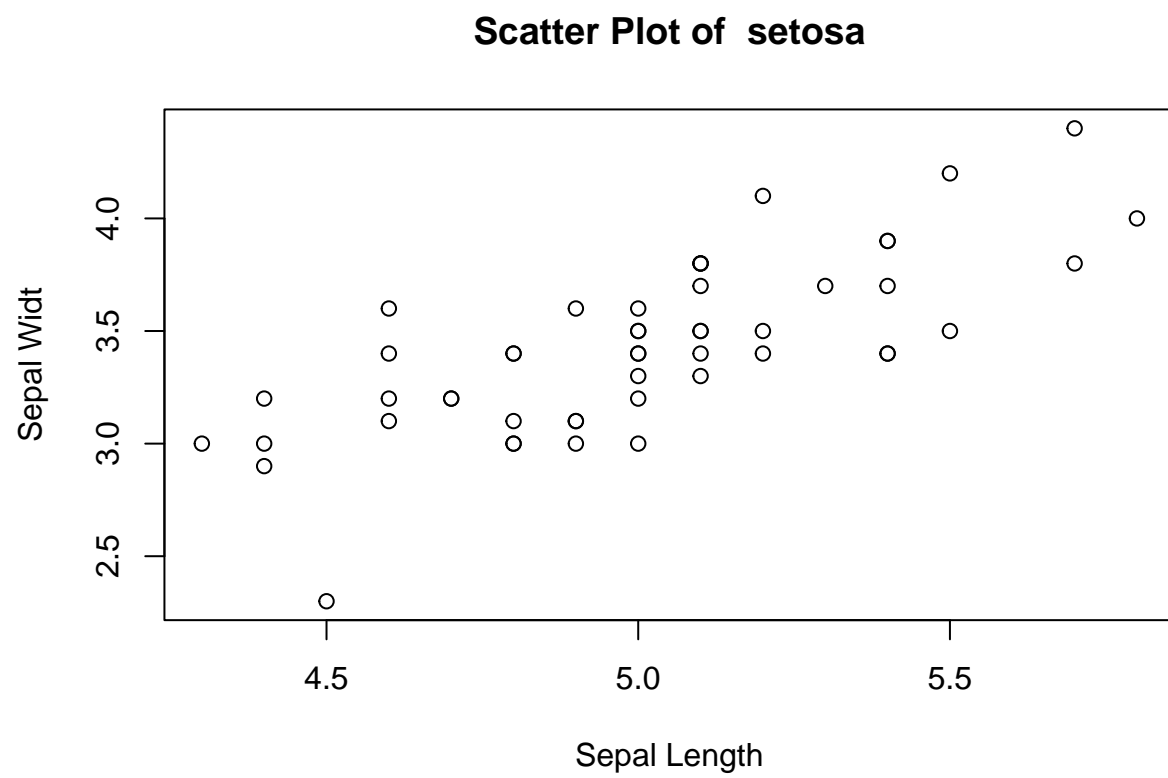
Exercise 5

1

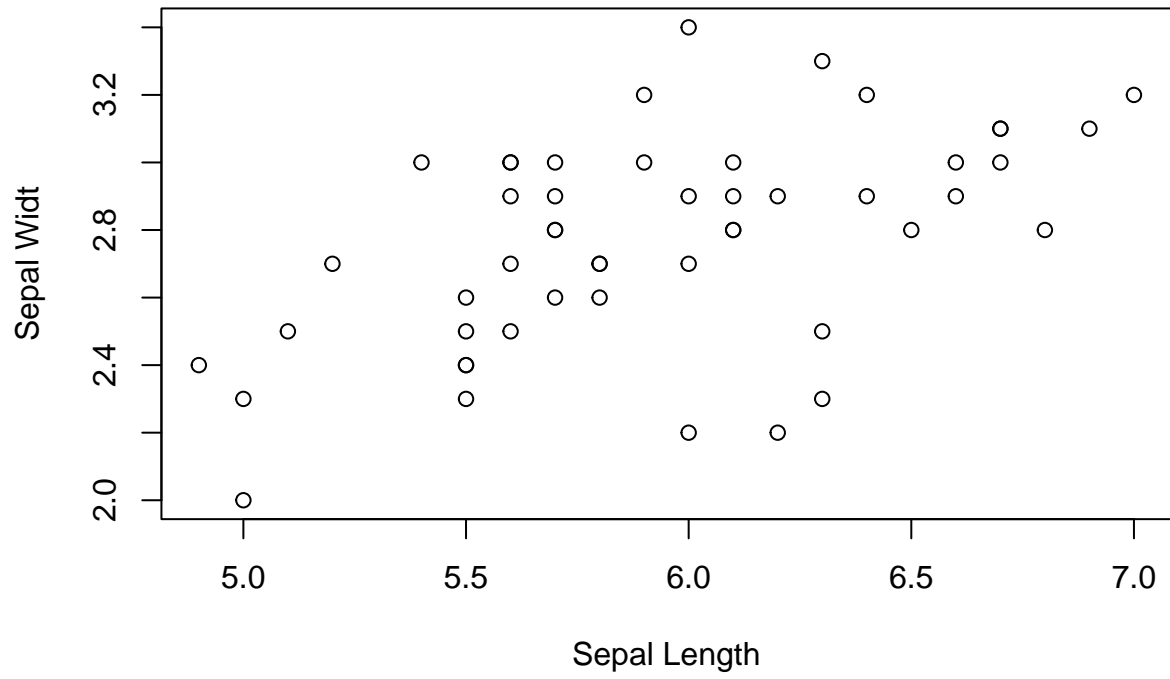
```
splitiris = split(iris, iris$Species)
```

2

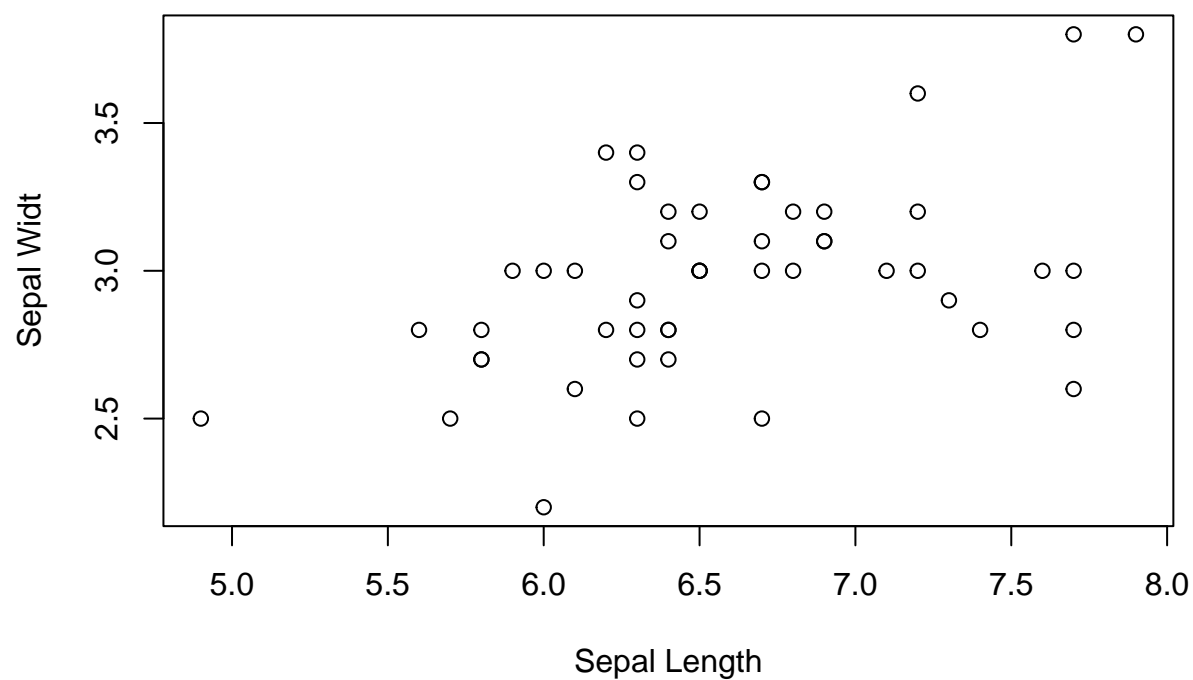
```
n = length(splitiris)
for (i in 1:n){
  i_df = splitiris[[i]]
  plot(x = i_df$Sepal.Length, y = i_df$Sepal.Width, xlab = 'Sepal Length', ylab = 'Sepal Width', main = )
}
```



Scatter Plot of versicolor

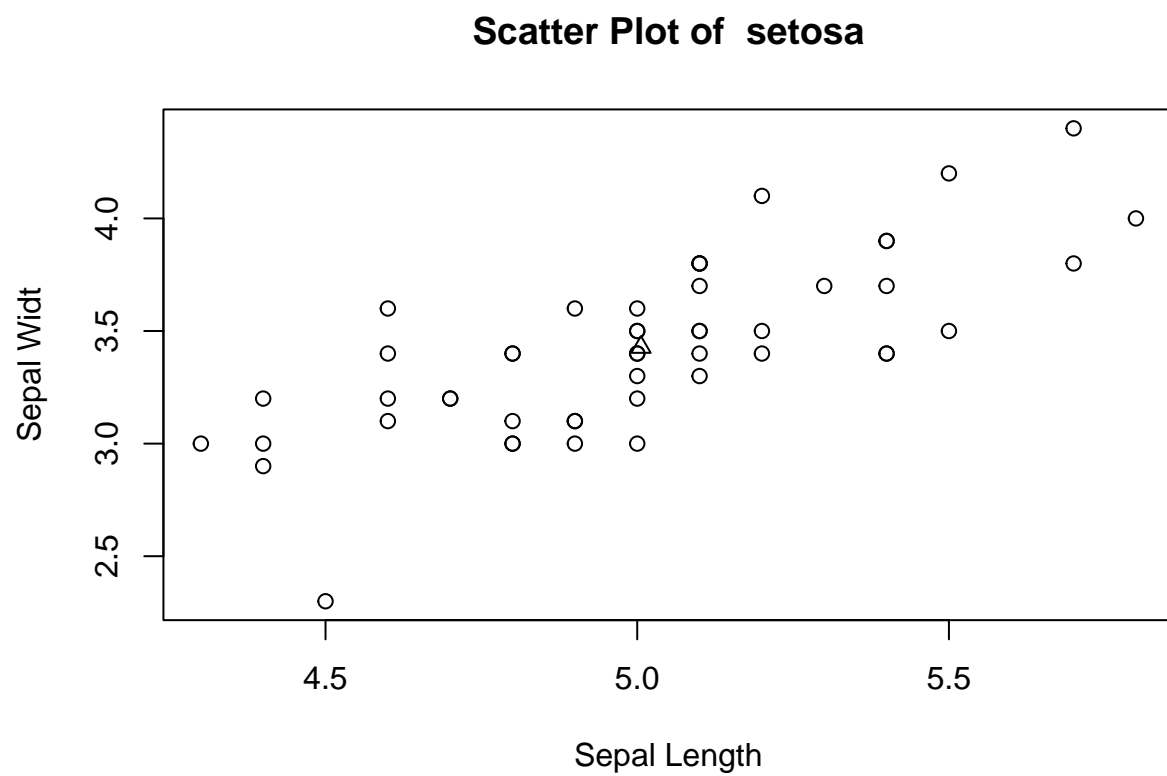


Scatter Plot of virginica

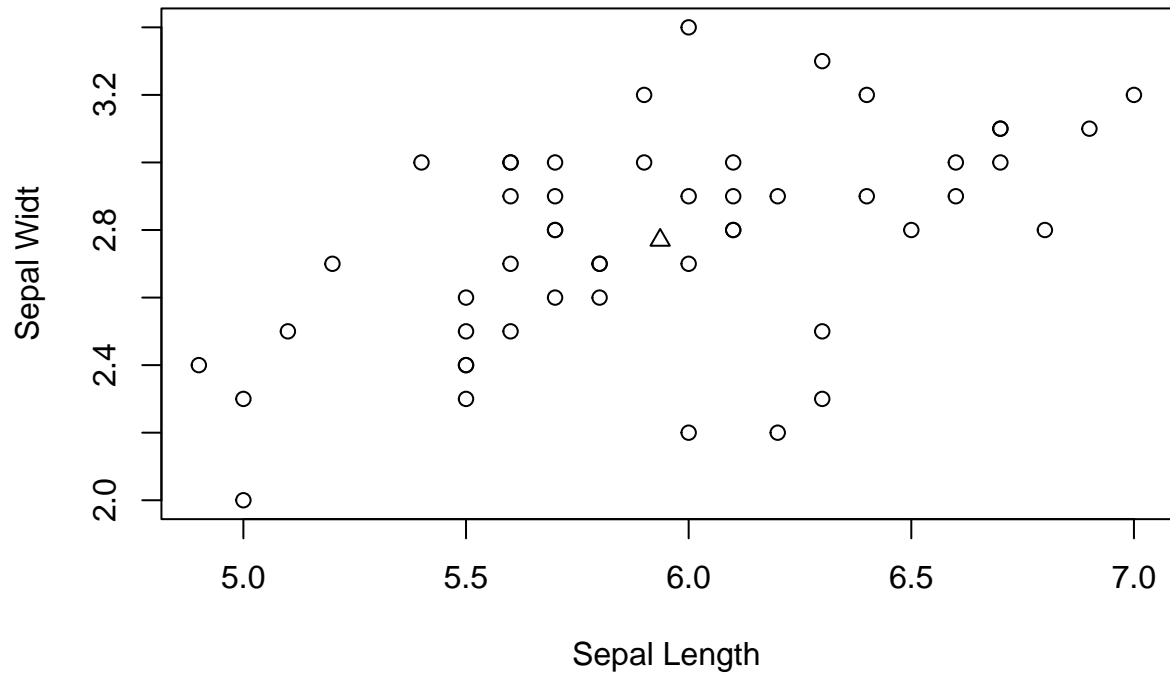


3

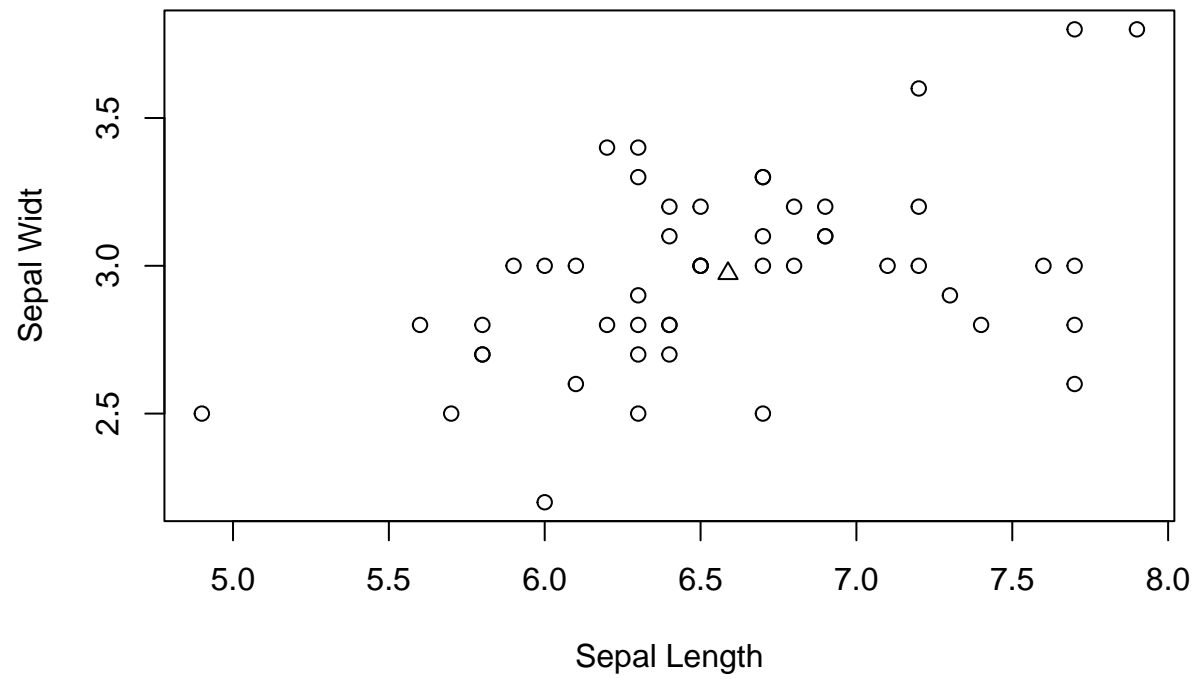
```
n = length(splitiris)
for (i in 1:n){
  i_df = splitiris[[i]]
  plot(x = i_df$Sepal.Length, y = i_df$Sepal.Width, xlab = 'Sepal Length', ylab = 'Sepal Width', main = 'Scatter Plot of virginica')
  points(x = mean(i_df$Sepal.Length), y = mean(i_df$Sepal.Width), pch = 24, col="black")
}
```



Scatter Plot of versicolor



Scatter Plot of virginica



4

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