Chapter 2: Creating graphs from data

1.2 a)

0 to 2	2 to 4	4 to 6	6 to 8
Frequency: 4	Frequency: 5	Frequency: 3	Frequency: 3

1.2 b)
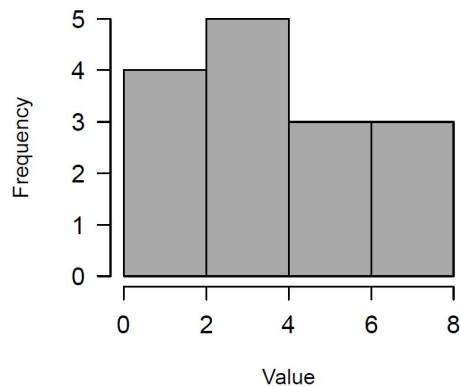
2.2 a) In the plot window you can see that a histogram for the values in the variable **dataset4** has been drawn by R.

```
dataset4 <- c(1.5, 5.5, 1.7, 7.2, 1.2, 7.9, 1.4, 3.6,
3.1, 3.8, 5.9, 3.6, 5.1, 3.2, 7.1)
hist(dataset4)
```

2.2 b) The histogram drawn by R has four bars for the ranges 1-2, 3-4, 5-6, and 7-8. The histogram from assignment 2.1b has four bars for the ranges 0-2, 2-4, 4-6, and 6-8. The difference in these two histograms is that they have a different number of bars for different ranges.

2.2 c)
```
# The breaks argument determines the number of bars in the histogram
# see ?hist for more help on the hist() function
hist(dataset4, breaks = 4)
```

2.2 d)
```
hist(dataset4, breaks = 4, col = 'blue', # col sets the bar color
xlab = 'x-axis', # xlab sets the x-axis name
ylab = 'Frequency', # ylab sets the y-axis name
main = 'My histogram') # main sets the title
```

2.2 e) These data are positively skewed.

Explanation: The mean is higher than the median, and so more values are concentrated on the left side (tail) of the distribution graph while the right tail of the distribution graph is longer.

```
mean(dataset4) # Mean: 4.12
median(dataset4) # Median: 3.6
```

2.3 a)
```
data(swiss) # Import the swiss data

education <- swiss$Education
agriculture <- swiss$Agriculture
```

2.3 b)
```
plot(x = education, y = agriculture,
     xlab = 'Percentage of education beyond primary school',
     ylab = 'Percentage of males involved in agriculture')
```

2.3 c) Looking at the scatter plot, there seems to be a tendency for provinces that have a high percentage of education beyond primary school to also have a low percentage of males involved in agriculture.

2.3 d)

```
plot(x = education, y = agriculture,
     xlab = 'Percentage of education beyond primary school',
     ylab = 'Percentage of males involved in agriculture',
     col = 'blue',
     main = 'Provinces in Switzerland',
     las = 1, # las sets the rotation of the axis labels
     bty = 'n') # bty = 'n'removes the outer border lines
```

2.4 a)
```
data(EuStockMarkets)
stockData <- data.frame(EuStockMarkets)

plot(stockData$DAX,
     type = 'l', # type = 'l'creates lines instead of dots
     xlab = 'Time',
     ylab = 'Price')
```

2.4 b)
```
# The lines() function adds a line to an existing plot
lines(stockData$SMI, col = 'red') # Add a line for the SMI stock
lines(stockData$CAC, col = 'blue') # Add a line for the CAC stock
lines(stockData$FTSE, col = 'green') # Add a line for the FTSE stock
```

2.5 a) Minimum: 1.2   Lower quartile: 35.3
        Median: 54.2   Upper quartile: 67.8
        Maximum: 89.7

```
quantile(agriculture, type = 6)
# Minimum: 1.2
# Lower quartile: 35.3
# Median: 54.1
# Upper quartile: 67.8
# Maximum: 89.7
```

2.5 b)
```
boxplot(agriculture)
```

2.5 c) The first line creates a vector called **educationLevel** that
        contains 47 times **'2.Medium'** . The second line changes the
        '2.Medium' to **'1.Low'** for the provinces that have a percentage
        of education beyond primary school lower than 6. The third line
        changes the **'2.Medium'** to **'3.High'** for the provinces that have
        a percentage higher than 12. The resulting table shows how many
        provinces had a percentage lower than 6, between 6 and 12, and
        higher than 12.

```
educationLevel <- rep('2.Medium', 47)
educationLevel[education <= 6] = '1.Low'
educationLevel[education >= 12] = '3.High'
table(educationLevel)
```

2.5 d) The code creates a box plot of the agriculture variable for each education level `'1.Low'` , `'2.Medium'` , and `'3.High'` .

```
boxplot(agriculture ~ educationLevel)
```