1. Consider the data set occupationalStatus in the datasets package.

(a) What is the probability of a son having the same occupational status as his father? [Hint: investigate what diag(x) does if x is a matrix.]

(b) Renormalize the data so that each row sums to 1. In the new data set the ith row represents the conditional distribution of a son's occupational status given that his father has occupational status i.

(c) What is the probability that a son has occupational status between 1 and 3, given that his father has status 1?

What if the father has occupational status 8?

```
library(datasets)
data("occupationalStatus")

prob <- sum(diag(occupationalStatus)) / sum(occupationalStatus)
prob
> prob
[1] 0.2747

occupationalStatus_norm <- apply(occupationalStatus, 1, function(x) x/sum(x))
prob_1to3_given_1 <- occupationalStatus_norm[1,1:3] %*% matrix(1, nrow=3)/3
prob_1to3_given_1
     [,1]
[1,] 0.6981159

prob_1to3_given_8 <- occupationalStatus_norm[8,1:3] %*% matrix(1, nrow=3)/3
prob_1to3_given_8
> prob_1to3_given_8
       [,1]
[1,] 0.2243202
```

2. Create the following data frame, subsequently invert Gender for all individuals.

a) Name Age Height Weight Gender

Alex 25 177 57 M

Lilly 31 163 69 M

Mark 23 190 83 F

```
data <- data.frame(
  Name = c("Alex", "Lilly", "Mark"),
  Age = c(25, 31, 23),
  Height = c(177, 163, 190),
  Weight = c(57, 69, 83),
  Gender = c("M", "M", "F")
)

data$Gender <- ifelse(data$Gender == "M", "F", "M")

print(data)
```

```
   Name Age Height Weight Gender
1  Alex  25    177     57      F
2 Lilly  31    163     69      F
3  Mark  23    190     83      M
```

b) Create the below data frame

Name Working

Alex Yes

Lilly No

Mark No

```
data2 <- data.frame(
  Name = c("Alex", "Lilly", "Mark"),
  Working = c("Yes", "No", "No")
)

print(data2)

   Name Working
1  Alex     Yes
2 Lilly      No
3  Mark      No
```

c) Add the data frame column-wise to the previous one.

How many rows and columns does the new data frame have?

```
merged_data <- cbind(data, data2$Working)

print(merged_data)

   Name Age Height Weight Gender data2$Working
1  Alex  25    177     57      F           Yes
2 Lilly  31    163     69      F            No
3  Mark  23    190     83      M            No
```

3. A student recorded his/her scores on weekly R programming quizzes that were marked out of a possible 10 points. His/Herscores were as follows:

8, 5, 8, 5, 7, 6, 7, 7, 5, 7, 5, 5, 6, 6, 9, 8, 9, 7, 9, 9, 6, 8, 6, 6, 7

What is the mode of his/her scores on the weekly R programming quizzes?

```
scores <- c(8, 5, 8, 5, 7, 6, 7, 7, 5, 7, 5, 5, 6, 6, 9, 8, 9, 7, 9, 9, 6, 8, 6, 6, 7)

mode <- names(table(scores))[table(scores)==max(table(scores))]
```

```
print(mode)
```

```
[1] "7"
```

4. Construct the following data frame.

Countries	population_in_million	gdp_per_capita
A	100	2000
B	200	7000
C	120	15000

a) Write appropriate R code and reshape the above data frame from wide data format to long data format.

```
library(tidyr)

data <- data.frame(
  Countries = c("A", "B", "C"),
  population_in_million = c(100, 200, 120),
  gdp_per_capita = c(2000, 7000, 15000)
)

long_data <- gather(data, key = "variable", value = "value", -Countries)

print(long_data)
```

```
  Countries            variable  value
1         A population_in_million    100
2         B population_in_million    200
3         C population_in_million    120
4         A        gdp_per_capita   2000
5         B        gdp_per_capita   7000
```

```
6       C      gdp_per_capita  15000
```

b) Write R code and reshape from long to wide data format.

```
wide_data <- spread(long_data, key = "variable", value = "value")

print(wide_data)
```

```
  Countries gdp_per_capita population_in_million
1       A         2000                100
2       B         7000                200
3       C        15000                120
```

5. Consider the following data present. Create this file using windows notepad . Save the file as input.csv using the save As All files(*.*) option in notepad.

```
Name,Age,Country,Gender
John,25,USA,Male
Mary,31,Canada,Female
David,23,UK,Male
Samantha,27,Australia,Female
```