Home Work 1

Ivan Ojiambo

September 16, 2017

Exercise 1

These are the things I need to do inorder to earn full points of the exercies:

- Am obliged to submit my exercise report to the gradescope in pdf format and on time
- My report ought to be reasonable and concisely
- In case of programming task, my report should show the approach I used to arrive at the answer, the value of the result and the conclusion.
- All plots, tables, graphs should be accompanied by explanation
- Am also obliged to answer all the question given in the exercise.
- **I will not plagiarize.**
- **I understand and follow the submission rule.**

Exercise 2

a. Column Names of the Database

```
dataset <- read.csv('https://courses.cs.ut.ee/MTAT.03.183/2017_fall/uploads/Main/abalone.csv',
                    TRUE, ',')
colnames(dataset)
```

```
## [1] "Gender"   "Length"   "Diameter" "Height"   "Weight"   "Rings"
```

b. Number of Observation in the Databas

```
nrow(dataset)
```

```
## [1] 1000
```

c. Print the first 4 lines from the dataset.

```
first_4Lines <- head(dataset,4)
first_4Lines
```

```
##   Gender Length Diameter Height Weight Rings
## 1      F  0.505    0.385  0.135 0.6185    12
## 2      F  0.650    0.475  0.165 1.3875     9
## 3      I  0.520    0.380  0.135 0.5395     8
## 4      F  0.550    0.425  0.140 0.9520     7
```

What are the values of feature rings of the printed observations?

```
first_4Lines[, c("Rings")]
```

```
## [1] 12  9  8  7
```

d. Extract the last 3 rows of the data frame. What is the weight of these abalones?

```
library(dplyr, warn.conflicts = FALSE)

last_3rows <- tail(dataset,3)
last_3rows
```

```
##      Gender Length Diameter Height Weight Rings
## 998       M  0.660    0.500  0.165 1.3195     9
## 999       I  0.525    0.400  0.130 0.6455     8
## 1000      M  0.515    0.395  0.135 1.0070     8
```

```
summarise(last_3rows,  Weight = sum(Weight))
```

```
##   Weight
## 1  2.972
```

e. What is the value of diameter in the row 577?

```
dataset[577,3]
```

```
## [1] 0.51
```

f. What is the mean of the height column?

```
summarise(dataset, Mean_Weight= mean(Height))
```

```
##   Mean_Weight
## 1      0.1411
```

g. Extract the subset of rows of the data frame where gender is M and weight values are below 0.75. What is the mean of diameter in this subset? Solve this subtask without using the dplyr package.

```
data_male  <- dataset[dataset$Gender =='M' , ]
final_data <- data_male[data_male$Weight <0.75 ,]
final_data
```

```
##     Gender Length Diameter Height Weight Rings
## 12       M  0.400    0.310  0.110 0.3140    11
## 13       M  0.460    0.360  0.125 0.5470     8
## 17       M  0.540    0.420  0.155 0.7385    12
## 19       M  0.345    0.255  0.090 0.2005     9
## 23       M  0.465    0.360  0.110 0.4955     7
```

```
## 37       M  0.490    0.385  0.125 0.6490     8
## 42       M  0.505    0.405  0.110 0.6250     9
## 59       M  0.430    0.310  0.130 0.6485     9
## 72       M  0.360    0.270  0.100 0.2170     6
## 74       M  0.470    0.375  0.130 0.5230     8
## 93       M  0.430    0.345  0.115 0.3045    11
## 94       M  0.565    0.425  0.100 0.7145    12
## 97       M  0.480    0.375  0.115 0.6765     6
## 127      M  0.485    0.410  0.150 0.6960    13
## 130      M  0.435    0.335  0.110 0.4385     7
## 135      M  0.465    0.360  0.115 0.5795     7
## 143      M  0.355    0.265  0.090 0.1680     8
## 148      M  0.475    0.385  0.120 0.5620     8
## 170      M  0.480    0.370  0.130 0.6430     8
## 172      M  0.460    0.345  0.110 0.4595     7
## 180      M  0.475    0.355  0.120 0.4800     8
## 188      M  0.235    0.160  0.060 0.0545     4
## 190      M  0.500    0.380  0.155 0.5955    12
## 198      M  0.480    0.365  0.120 0.6015     7
## 220      M  0.530    0.425  0.130 0.7020     9
## 221      M  0.345    0.270  0.095 0.1970     9
## 242      M  0.360    0.295  0.105 0.2410     8
## 263      M  0.505    0.380  0.130 0.6560    13
## 270      M  0.520    0.390  0.120 0.6435     7
## 278      M  0.415    0.315  0.115 0.3895     9
## 280      M  0.410    0.300  0.100 0.3010     9
## 290      M  0.455    0.350  0.120 0.4835    11
## 292      M  0.475    0.395  0.135 0.5920    13
## 302      M  0.360    0.270  0.090 0.2225     6
## 308      M  0.385    0.255  0.100 0.3175     8
## 312      M  0.445    0.340  0.120 0.4475     9
## 315      M  0.385    0.275  0.115 0.2685     8
## 329      M  0.460    0.375  0.135 0.4935    12
## 332      M  0.515    0.395  0.120 0.6460     9
## 335      M  0.530    0.415  0.120 0.7060     9
## 336      M  0.450    0.340  0.130 0.3715     9
## 338      M  0.505    0.395  0.135 0.5915    12
## 367      M  0.505    0.385  0.110 0.6550     9
## 376      M  0.555    0.435  0.140 0.7495     8
## 393      M  0.425    0.325  0.095 0.3785     7
## 398      M  0.310    0.245  0.095 0.1500     7
## 401      M  0.500    0.400  0.125 0.6725     7
## 403      M  0.495    0.375  0.115 0.6245     6
## 405      M  0.535    0.420  0.160 0.7465    10
## 406      M  0.455    0.355  0.135 0.4745    13
## 415      M  0.550    0.385  0.130 0.7275     8
## 419      M  0.290    0.230  0.075 0.1165     7
## 424      M  0.425    0.305  0.110 0.3590     9
## 464      M  0.395    0.295  0.115 0.3160    12
## 465      M  0.450    0.355  0.115 0.4790     8
## 466      M  0.355    0.265  0.085 0.2010     8
## 468      M  0.540    0.415  0.145 0.7400    12
## 483      M  0.385    0.300  0.115 0.3435     6
## 492      M  0.485    0.390  0.085 0.6435     8
```

```
## 500      M  0.350    0.265  0.090 0.2265     6
## 504      M  0.505    0.390  0.115 0.5585     8
## 523      M  0.370    0.290  0.090 0.2410    10
## 524      M  0.375    0.290  0.100 0.2760     9
## 525      M  0.335    0.260  0.075 0.2200     6
## 528      M  0.500    0.375  0.150 0.6360    10
## 536      M  0.485    0.385  0.125 0.4775    12
## 539      M  0.460    0.375  0.130 0.5735     9
## 541      M  0.375    0.300  0.100 0.2465    11
## 546      M  0.480    0.375  0.120 0.5895    11
## 551      M  0.335    0.250  0.090 0.1810     7
## 552      M  0.495    0.395  0.120 0.5530     8
## 560      M  0.510    0.390  0.125 0.6565    10
## 567      M  0.535    0.405  0.140 0.7315     7
## 580      M  0.235    0.170  0.055 0.0515     7
## 594      M  0.385    0.285  0.105 0.2905    12
## 597      M  0.515    0.380  0.135 0.6615    10
## 629      M  0.490    0.390  0.140 0.7070    13
## 632      M  0.350    0.260  0.090 0.1950     9
## 652      M  0.280    0.210  0.080 0.1085     7
## 653      M  0.530    0.425  0.130 0.7455    10
## 666      M  0.395    0.280  0.080 0.2660    12
## 670      M  0.495    0.400  0.135 0.6100     7
## 671      M  0.480    0.355  0.160 0.4640     8
## 675      M  0.505    0.365  0.115 0.5210     8
## 680      M  0.490    0.395  0.140 0.5490    11
## 685      M  0.525    0.385  0.100 0.5115     8
## 691      M  0.535    0.435  0.150 0.7170     9
## 696      M  0.575    0.445  0.140 0.7370    10
## 700      M  0.450    0.355  0.115 0.4780    10
## 703      M  0.440    0.335  0.110 0.3940     9
## 704      M  0.450    0.345  0.105 0.4115     7
## 719      M  0.515    0.455  0.135 0.7225     9
## 720      M  0.295    0.225  0.090 0.1385     9
## 742      M  0.465    0.340  0.105 0.4860     9
## 752      M  0.280    0.205  0.100 0.1165     5
## 755      M  0.505    0.385  0.130 0.6435     7
## 758      M  0.475    0.375  0.120 0.5630    10
## 764      M  0.515    0.390  0.120 0.6125     8
## 773      M  0.460    0.355  0.140 0.4910    10
## 778      M  0.350    0.260  0.090 0.1980    10
## 782      M  0.275    0.205  0.070 0.0940     5
## 785      M  0.340    0.255  0.095 0.2130     9
## 803      M  0.460    0.375  0.140 0.5105     9
## 805      M  0.350    0.275  0.110 0.2925     8
## 821      M  0.475    0.360  0.120 0.5780     8
## 829      M  0.540    0.420  0.190 0.6855    10
## 835      M  0.445    0.345  0.090 0.3795    10
## 847      M  0.325    0.240  0.085 0.1730     7
## 848      M  0.470    0.370  0.130 0.5225     7
## 852      M  0.440    0.350  0.110 0.4585     9
## 861      M  0.480    0.380  0.135 0.5280    14
## 869      M  0.365    0.295  0.080 0.2555     7
## 870      M  0.480    0.360  0.100 0.4390     8
```

```
## 885      M  0.490    0.395  0.120 0.6740     9
## 905      M  0.505    0.385  0.105 0.5525     9
## 943      M  0.495    0.385  0.135 0.7090    12
## 963      M  0.360    0.295  0.130 0.2765    10
## 980      M  0.490    0.385  0.125 0.6090     8
## 993      M  0.480    0.365  0.130 0.5305     8
```

```
mean(final_data$Diameter)
```

```
## [1] 0.3426471
```

h Now do the same as in the previous subtask (g) but use the %>% operator from dplyr package.

```
dataset %>%
  filter(Gender =='M', Weight <0.75) %>%
  summarise(Mean_Diameter = mean(Diameter))
```

```
##   Mean_Diameter
## 1     0.3426471
```

i. What is the minimum of length when rings is equal to 18?

```
data <- dataset[dataset$Rings ==18, ]
min(data$Length, na.rm = T)
```

```
## [1] 0.465
```

j. Is the weight of abalones related to how many rings they have? Please provide evidence and explanatory text with the conclusion. Hint: consider using the dplyr commands group_by and summarise.

```
library(dplyr, warn.conflicts = FALSE)

grouped_data <- dataset %>% select(Weight, Rings)

orderd_data <- grouped_data[order(grouped_data[,2]),]
cor(orderd_data$Weight, orderd_data$Rings)
```

```
## [1] 0.03393787
```

No, the weight of the balones is not related to the number of Rings. From the correlation coeficient (0.03397), its clear that the weight of the balones are not related to the increase in rings

Exercise3

a Was it interesting to you and what did you like or didn't like about this?

From talk **The best stats you've ever seen** by Hans Rosling, what I liked most was how he used data to prove his paradigms. I had never imagined that the size of the family can be used as an indicator for life

expectancy of any nation.

From TED talk **What makes a good life? Lessons from the longest study on happiness** by Robert Waldinger, what was intresting to me was to learn that what makes good life is not having alot of money or even being a celbrity but having quality relationship with people.

From the TED talk **How I hacked online Dating** by Amy Webber, what was intresting to me was to learn how she created 10 male profile to see which type of girls would get attracted to the type of man she wanted.

b What was the key message that you would tell about to your friend?

From **Hans Rosling's message**, I think one of the message I would love to share with my friends is how having a small familly is a key to having a high life expectancy of a nation. Also Comming from a third world country, I think one of the lesson I would love to share with my friend is how we can learn from countries such as singapore which moved from being a third world country in the 1960s to a first world country.

From **Robert Waldinger's** message, I would love to share with my friends how having good relationship with the people we love can keep us health and happier and not the money or fame.

From this message of **Amy Webber**, the key message I would share with my friend is to have take some market researcher on the things we are intrested in by collecting the relevant data that would help us make rational decisions.

c What was the best visualisation in your opinion and why was it so appealing to you?

From **Rosling's** message, In my opionion the best visaulisation was when he compared the life expectancy of Mauritious with other African countries. It was appealing to me because it was more less like prove of concept that indeed having a small family leads to high life expectancy.

From **Robert Waldinger**, what I visualised most was when he spoke of replacing screen time with people Time. This appealed to me because I spend most of my time on the computer and I have less time for interacting with people.

From **Amy Webber's** message the best visualisation was when she finally got mr Right after a series of data collection ad anlaysis of the type of man she wanted. It was appealing to me because am am one of those who believes in falling in love with the right person who fulfills atlest 80% of your expectations.

Exercise4

I have come to realise that I do not have proper plan for my incomes and expenditure and as a result I find myself spending anyhow I see fit. I have always tried to make a small budget of my income but I have always failed to keep to it since I did not have clear data of my spending behaviour and as a result I end up spending more sometimes on useless things that I didn't even intend to buy that month.

Its on this ground that I feel there is need for me to record all my data regarding my expenditures and incomes. I want to began by documenting all my receipts for the things that I buy for the period of three month. I also need to document all the money that I give out as gifts including the money that I give to my girl friends.

After recording my expenditures for the three month period, I will analyse the data and it will help me to budget well for my income. I hope to recomend the same approach to my friends who are having the same challenge as me.