Computational Analysis Assignment I

124384 Luycer Bosire

2/2/2021

Question One

1. Use the file “Women.txt” from the course website and read this into R. What is the dimension of the data you have just read in R?

#Reading the .txt file into R,

women<- read.table("Women.txt", header=TRUE)

Next, check the dimension of this table. We may also go further ahead and show the first and last three items from the table. We’ll realise that the table has 17 rows and 3 columns. This data can be easily viewed

dim(women)

## [1] 17 3

head(women,3)

## height weight age  
## 1 58 115 33  
## 2 59 117 34  
## 3 60 120 37

tail(women,3)

## height weight age  
## 15 72 164 37  
## 16 71 160 NA  
## 17 73 161 35

View(women)

1. Use the file “Women.txt” from the course website and read this into R. A new woman joined the study, she is 66” tall, 165lbs and 34 years. Append this information to your data.

new.woman<-data.frame(height="66",weight="165",age="34")  
women1<-rbind(women,new.woman)  
women1

## height weight age  
## 1 58 115 33  
## 2 59 117 34  
## 3 60 120 37  
## 4 61 123 31  
## 5 62 126 31  
## 6 63 129 34  
## 7 64 132 31  
## 8 65 135 39  
## 9 66 139 35  
## 10 67 142 34  
## 11 68 146 34  
## 12 69 150 36  
## 13 70 154 33  
## 14 71 159 30  
## 15 72 164 37  
## 16 71 160 <NA>  
## 17 73 161 35  
## 18 66 165 34

1. Use the file “Women.txt” from the course website and read this into R. How many women have a weight under 140?

with(women,women$weight<140) #From this, It returns the first 9 values as less than 140 and the rest as greater than.

## [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE FALSE FALSE FALSE  
## [13] FALSE FALSE FALSE FALSE FALSE

We can obtain similar results by suing the below function, as the above one is quite tedious ovr manually counting the TRUE values from the output;

sum(women$weight<140)

## [1] 9

1. Use the file “Women.txt” from the course website and read this into R. There is a correction to the woman in row D, her age should be 39. Change the age in row D to 39. You’re then required to sort your data by weight and store the results in a new data set.

women<- read.table("Women.txt", header=TRUE)  
women2<-replace(women$age,4,39)  
women2

## [1] 33 34 37 39 31 34 31 39 35 34 34 36 33 30 37 NA 35

1. Use apply to generate a summary report, with the mean, median, sd of height, weight and age.

The question requires the mean,median and standard deviation using apply function which is mainly used to avoid loops. It takes three arguments; x-dataframe Margin-1 is for row operations and 2 column operations FUN- the function to be applied,e.g mean,median and sd

Before applying the function, there is need to omit incomplete cases if they exist.

attach(women)  
women2<-na.omit(women)  
mean<-apply(women2,2,FUN = ("mean"))  
median<-apply(women2,2,FUN="median")  
sd<-apply(women2,2,FUN="sd")

1. Use the apply function to get the mean, median and sd of the columns and create a matrix with row names, mean, median and sd.

t(data.frame(mean,median,sd))

## height weight age  
## mean 65.500000 138.25000 34.000000  
## median 65.500000 137.00000 34.000000  
## sd 4.760952 16.15549 2.503331

1. Write a function to calculate BMI.

attach(women)

## The following objects are masked from women (pos = 3):  
##   
## age, height, weight

BMI = function(height,weight){(0.45455\*weight/(.0254\*height)^2)}  
Women\_BMI<- BMI(height,weight)  
Women\_BMI

## [1] 24.08552 23.68078 23.48513 23.28947 23.09412 22.89933 22.70535 22.51237  
## [9] 22.48232 22.28707 22.24586 22.19767 22.14312 22.22259 22.28913 22.36235  
## [17] 21.28602

1. Do the women have a BMI within a recommended range for their height (Normal =18.5-24.9)? You require an R code to answer this.

#Using the sum function and logic and (&) operator to get the count of values within the normal range.

sum(Women\_BMI>18.5 & Women\_BMI<24.90)

## [1] 17

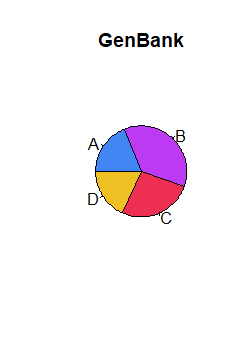
#This means all the 17 women are within the normal range.

Question 2 a. You’re required to load the library “ape” and then read the data bank as follows>bank<-table(read.GenBank(c(“X94991.1”), as.character=TRUE)) You’re then required to produce a pie char of the data bank in both 2D and 3D. Label your pie chart appropriately.

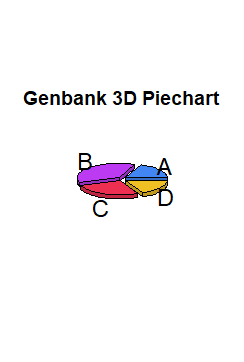
library(ape)  
library(plotrix)  
bank<-table(read.GenBank(c("X94991.1"), as.character=TRUE))  
bank

##   
## a c g t   
## 410 789 573 394

colors = c('#4286f4','#bb3af2','#ed2f52','#efc023')  
labels=c("A","B","C","D")  
pie(bank, labels, main='GenBank', col=colors, init.angle=180, clockwise=TRUE)



colors = c('#4286f4','#bb3af2','#ed2f52','#efc023')  
labels=c("A","B","C","D")  
pie3D(bank, labels=labels, explode=0.1, height=0.05, main='Genbank 3D Piechart', col=colors)



1. Use the file “gapminder.csv” from the course website and read this into R. You’re then required to perform the following analysis:

data\_set<-read.csv("gapminder.csv")  
str(data\_set)

## 'data.frame': 1704 obs. of 6 variables:  
## $ country : chr "Afghanistan" "Afghanistan" "Afghanistan" "Afghanistan" ...  
## $ continent: chr "Asia" "Asia" "Asia" "Asia" ...  
## $ year : int 1952 1957 1962 1967 1972 1977 1982 1987 1992 1997 ...  
## $ lifeExp : num 28.8 30.3 32 34 36.1 ...  
## $ pop : int 8425333 9240934 10267083 11537966 13079460 14880372 12881816 13867957 16317921 22227415 ...  
## $ gdpPercap: num 779 821 853 836 740 ...

head(data\_set)

## country continent year lifeExp pop gdpPercap  
## 1 Afghanistan Asia 1952 28.801 8425333 779.4453  
## 2 Afghanistan Asia 1957 30.332 9240934 820.8530  
## 3 Afghanistan Asia 1962 31.997 10267083 853.1007  
## 4 Afghanistan Asia 1967 34.020 11537966 836.1971  
## 5 Afghanistan Asia 1972 36.088 13079460 739.9811  
## 6 Afghanistan Asia 1977 38.438 14880372 786.1134

1. Obtain data set for only the year 1982

data\_1982<- subset(data\_set,subset = year==1982)  
head(data\_1982)

## country continent year lifeExp pop gdpPercap  
## 7 Afghanistan Asia 1982 39.854 12881816 978.0114  
## 19 Albania Europe 1982 70.420 2780097 3630.8807  
## 31 Algeria Africa 1982 61.368 20033753 5745.1602  
## 43 Angola Africa 1982 39.942 7016384 2756.9537  
## 55 Argentina Americas 1982 69.942 29341374 8997.8974  
## 67 Australia Oceania 1982 74.740 15184200 19477.0093

str(data\_1982)

## 'data.frame': 142 obs. of 6 variables:  
## $ country : chr "Afghanistan" "Albania" "Algeria" "Angola" ...  
## $ continent: chr "Asia" "Europe" "Africa" "Africa" ...  
## $ year : int 1982 1982 1982 1982 1982 1982 1982 1982 1982 1982 ...  
## $ lifeExp : num 39.9 70.4 61.4 39.9 69.9 ...  
## $ pop : int 12881816 2780097 20033753 7016384 29341374 15184200 7574613 377967 93074406 9856303 ...  
## $ gdpPercap: num 978 3631 5745 2757 8998 ...

1. Obtain data set for the American countries in 1997

data\_America<-subset(data\_set,continent == "Americas"& year==1997)  
head(data\_America)

## country continent year lifeExp pop gdpPercap  
## 58 Argentina Americas 1997 73.275 36203463 10967.282  
## 142 Bolivia Americas 1997 62.050 7693188 3326.143  
## 178 Brazil Americas 1997 69.388 168546719 7957.981  
## 250 Canada Americas 1997 78.610 30305843 28954.926  
## 286 Chile Americas 1997 75.816 14599929 10118.053  
## 310 Colombia Americas 1997 70.313 37657830 6117.362

1. Add a column to the data set loaded in R with the new column being gdp which is computed as gdp=popgdpPercap\*

attach(data\_set)  
GDP<-function(pop,gdpPercap){pop\*gdpPercap}  
gdp<-GDP(pop,gdpPercap)  
head(gdp)

## [1] 6567086330 7585448670 8758855797 9648014150 9678553274 11697659231

data\_set\_1<- cbind(data\_set,gdp)  
str(data\_set\_1)

## 'data.frame': 1704 obs. of 7 variables:  
## $ country : chr "Afghanistan" "Afghanistan" "Afghanistan" "Afghanistan" ...  
## $ continent: chr "Asia" "Asia" "Asia" "Asia" ...  
## $ year : int 1952 1957 1962 1967 1972 1977 1982 1987 1992 1997 ...  
## $ lifeExp : num 28.8 30.3 32 34 36.1 ...  
## $ pop : int 8425333 9240934 10267083 11537966 13079460 14880372 12881816 13867957 16317921 22227415 ...  
## $ gdpPercap: num 779 821 853 836 740 ...  
## $ gdp : num 6.57e+09 7.59e+09 8.76e+09 9.65e+09 9.68e+09 ...