Department of Computer Technology and Information Systems
CTIS 365: Applied Data Analysis
Semester: Fall 2019-2020

Lab Guide #2 – Week 03

OBJECTIVE: Lists, Missing Values, Plot

Instructor: Dr. Seyid Amjad ALI
Assistant: Burcu ALPER

Q1. You have been supplied with some data as vectors about the employees of an X Company (Id, Name, Salary, and Working Year). The X Company wants to give 7.5% raise to only those employees who have worked more than 15 years.

```
employee_Id <- c(111, 222, 333, 444, 555, 666, 777, 888, 999, 112, 113, 114, 115,
                 116, 117, 118, 119, 120, 121, 123, 124, 125, 126)
employee_Name <- c("Rick"    , "Dan"    ,   "Michelle" ,"Aaron"   ,  "Frank"  ,   "Abbey"  ,
                   "Edward"  , "Abel"   ,   "Reuben"  ,  "Abelson","Abrams"  ,  "Ace"    ,   "Jne"      ,
                   "Mary"    ,  "James"  ,   "Bertie"  ,  "Rachel"  ,  "Ross"  ,    "Monica"  , "Chandler",
                   "Phoebe",   "Joseph",    "Janice"  )
employee_Salary <- c(623.3 , 515.2 , 611.0 , 729.0 , 843.0 , 850.0 , 954.0 ,1145.0 ,
                     741.0 , 600.0 , 451.0 , 620.0 , 750.0 , 460.0 ,1621.0 ,1400.0 , 541.0  ,
                     856.0 , 974.0,  687.0 , 574.0 , 641.0 ,1245.0)
employee_workingYear <- c( 15,  14,   8,  15,   4,  14  ,19 ,12  ,2,   5 , 1 ,16 , 7,   4,
                           20 ,14 ,  5,  18,  19,   6,  13 , 4, 12)
```

    a. Calculate the new salaries.
    b. Create a data frame (or a tibble) which should look like Figure1.
       (**Hint:** Use same salaries for employees whose working years are lower than 5 year).
    c. Rename the columns of the employee data frame (or a tibble) by looking at Figure1.
    d. Convert the id column to a factor.
    e. Find and print the employee who earns the maximum amount of the raise.

```
"The highest raise to the  James  was made as  121.575 $"
```

Employee Id	Employee Name	Employee Salary	Employee Working Year	New Salary
111	Rick	623.3	15	623.300
222	Dan	515.2	14	515.200
333	Michelle	611.0	8	611.000
444	Aaron	729.0	15	729.000
555	Frank	843.0	4	843.000
666	Abbey	850.0	14	850.000

Figure 1

Q2. The following table shows first six rows of a data set from a file that has .txt extension **(mydata.txt).**

```
BodyWgt BrainWgt NonD Dream Sleep Span Gest Pred Exp Danger
6654.000   5712.0 12.3   NA   3.3 38.6  645    3   5      3
   1.000      6.6  6.3  2.0   8.3  4.5   42    3   1      3
   3.385     44.5   NA   NA  12.5 14.0   60    1   1      1
   0.920      5.7   NA   NA  16.5   NA   25    5   2      3
2547.000   4603.0  2.1  1.8   3.9 69.0  624    3   5      4
  10.550    179.5  9.1  0.7   9.8 27.0  180    4   4      4
```

    a. Read data from the given **.txt** file.
    b. Find the number of missing values in each column.
    c. How many missing values are there?
    d. Remove the missing values (remove rows having missing values).

Q3. Read the **academic.csv** file that includes the name of the advisors as headers and name of their students. Create a list as shown below.

```
> class_grades.list
$Cüneyt.Sevgi
[1] st1  st2  st3  st4  st5  st6  st18 st19 st20
Levels: st1 st18 st19 st2 st20 st3 st4 st5 st6

$Burcu.Liman
[1] st9  st10 st11 st12 st13 st14 st15 st21 st22
Levels: st10 st11 st12 st13 st14 st15 st21 st22 st9

$Ceren.Serim
[1] st23 st24 st25 st26 st27 st28 st29 s30  st31
Levels: s30 st23 st24 st25 st26 st27 st28 st29 st31

$Nese.Sahin
[1] st33 st34 st35 st36 st37 st38 st39 st40 st41
Levels: st33 st34 st35 st36 st37 st38 st39 st40 st41
```

Q4. Read **temperature.csv** file that includes the 3 cities weekly temperatures. Create a list and find the maximum and minimum temperature of the week for all cities.

Q5. Draw a plot of weekly temperature of Ankara. Your plot should look exactly like the below given screen shot.

Temperature of Ankara for One Week