Department of Information Systems and Technologies

CTIS 365: Applied Data Analysis

Semester: Spring 2025-2026

Lab Guide #2

OBJECTIVES: Data frame, Lists, Missing Values, Plot
Instructor: Dr. Syed Amjad ALI **Assistant:** Efe Mert SAHINKOC

Q1. **a.** Create ID, Name, Title and Salary Vectors of faculty of Bilkent University.
b. If an academic whose language grade is greater than 85, then s/he will get a 2.5% increase in his/her salary.

```
academician_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)

academician_Name <- c("Rick"    , "Dan"   ,   "Michelle" ,"Aaron"  ,  "Frank" ,   "Abbey" ,"Edward"  ,   "Abel"
,   "Reuben" ,  "Abelson","Abrams"  , "Ace"    ,   "Jane"     ,  "Mary"    ,  "James" ,   "Bertie" ,  "Rachel" ,
"Ross" ,    "Monica", "Chandler", "Phoebe",   "Joseph",   "Janice"  )

title               <-              c("Dr","Asst","Dr","Prof","Dr","Prof","Prof","Asst","Dr","Inst","Asst",
"Inst","Dr","Asst","Asst","Asst","Dr","Dr","Prof","Inst","Inst","Dr","Prof")

academician_Salary <- c(6230.3 , 5150.2 , 6110.0 , 7290.0 , 8430.0 , 8500.0 , 9540.0 ,4145.0 ,7410.0 , 6000.0
, 4510.0 , 6200.0 , 7500.0 , 4600.0 ,4621.0 ,5400.0 , 7410.0  ,8560.0 , 9740.0,  6870.0 , 5740.0 , 6410.0 ,8245.0)

academician_languageExamScore <- c( 65,  84,  86,  95,  91 , 40  ,35 , 85  , 38,  59 , 81 ,68 , 87,  84, 53
,94 , 85,  88,  79,  86,  53 , 45, 89)
```

- Calculate the new salaries.
- Create a data frame which should look like Figure1.
  (**Hint:** Use same salaries for the academicians who did not get 85 point from language exam).
- Rename the columns of the academician data frame by looking at Figure1.
- Convert the title column to a factor.
- Compare the means of old and new academician salaries.
- From mean comparison (you did in the previous item), state how many academicians have salary increase.
- Find and print the academician who earns the maximum amount of the raise.

```
[1] "The highest raise was made to  Ross  as  214 $"

> academicians
   Academician Id Academician Name Title Academician Salary Language Exam Point New Salary
1             111             Rick    Dr              6230.3                 65   6230.300
2             222              Dan  Asst              5150.2                 84   5150.200
3             333         Michelle    Dr              6110.0                 86   6262.750
4             444            Aaron  Prof              7290.0                 95   7472.250
5             555            Frank    Dr              8430.0                 91   8640.750
6             666            Abbey  Prof              8500.0                 40   8500.000
7             777           Edward  Prof              9540.0                 35   9540.000
8             888             Abel  Asst              4145.0                 85   4248.625
9             999           Reuben    Dr              7410.0                 38   7595.250
10            112           Abelson  Inst              6000.0                 59   6000.000
```

Figure 1

Q2. The following table shows first six rows of a data set from **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

- Find the number of missing values in each column of mydata.txt.
- Report the total number of missing values?
- Remove the missing values i.e., remove all rows containing a missing value.

Q3.

- Create a list containing a *vector*, a *matrix* and a *list*.
- Give names to the elements in the list.
- Access the first and second element of the list. (You can use the below given data)

Output:

```
$Color
[1] "Red"   "Green" "Black"

$`Odd numbers`
     [,1] [,2] [,3]
[1,]    1    5    9
[2,]    3    7   11

$`Language(s)`
$`Language(s)`[[1]]
[1] "Python"

$`Language(s)`[[2]]
[1] "PHP"

$`Language(s)`[[3]]
[1] "Java"
```

Q4a. Read the file named **temperature.csv** that includes weekly temperatures values for cities. Create a list then find and display the maximum and minimum temperature of the week for each cities. (Use lapply function for maximum and minimum).

Output:

```
Maximum temperature in Ankara is  26  degree.
Maximum temperature in Istanbul is  28  degree.
Maximum temperature in Izmir is  29  degree.
Minimum temperature in Ankara is  26  degree.
Minimum temperature in Istanbul is  28  degree.
Minimum temperature in Izmir is  29  degree.
```

Q4b. Using **temperature.csv**, draw a plot for the weekly temperature of Ankara. Your plot should look exactly like the below given screen shot.