**Project 1**

Venkata Satya Nagendra Sai Krishna Mohan Kocherlakota

College of Professional Studies, Northeastern University

ALY6000: Introduction to Analytics

Prof. Roy Wada

September 26, 2023

**Introduction**

This is an introductory project session that helps us understand the foundations of R, including arithmetic and logic operators like '+', '\*', TRUE, FALSE, etc. I create vectors, datasets, and data frames to utilize operators and various functions such as 'hist' and 'ggplot' using the 'pacman' and 'tidyverse' libraries.

I have also used statistical functions for our knowledge for data analysis. Used min, max, mean, median & sd (minimum value, maximum value, mean, median and standard deviation of the vectors).

The built-in functions and operators used throughout the project are ‘c’, ‘seq’, ‘rep’, ‘range operator’, ‘sum’, ‘greater than/ less than’, ‘cumsum’, etc.

Used set.seed() function, runif and rnorm function for the first time, learnt there usage in R.

**Key Findings: -**

1. In the below mentioned snapshot, I have used the arithmetic operators like multiplication, squares, factorial and division of numbers. In addition to that I have used logical operator ‘AND’ and ‘OR’.

A screenshot of a computer

Description automatically generated

1. The below code is to increase, multiplication, greater than or equal to of a vector and showing the results of the vector based on the operators used respectivelyA screenshot of a computer

   Description automatically generated
2. The below vector gives us the TRUE values on the bases of the inform I provided for the first vector using logical operator.

**A screenshot of a computer

Description automatically generated**

1. The below operator gives us all the values greater than or equal to 20 of the second vector and ages\_vector which I have created using ‘seq’ function.

**A close-up of words

Description automatically generated**

1. set.seed() : - This is a function when random variable is created, this function is helped to reproduce this random variable which I have created to store it.

runif: - This function is used to create a random variable by giving the number of random variables with minimum value and maximum value of the vector.

rnorm: - Function is used to create random variables using normal distribution and the number of variables, mean and standard deviation is given as input in order to generate the random variable (random vector)

Used these functions to create two random variables/vectors as shown below.





1. Histogram is used in the below screenshot using a random vector which was created using ‘rnorm’ function and ‘hist’ function in R.



A screenshot of a computer

Description automatically generated

A graph of a graph

Description automatically generated

1. The below screenshots have shown various operators and functions. I have created a data frame using ‘read\_csv’ function and tidyverse library.

head(data\_frame): -The below function shows the first six rows of the data frame.

head(data\_frame, n=): - The below function shows the first seven rows of the data frame because I have number of rows to show using ‘n’ variable.

names(data\_frame): - The column names of the data frame.

select,arrange, filter, mutate, slice: - These are various functions which are used for sorting, selection, filtering, slicing respectively.

ggpolt: - This uses to plot the graph of given data frame and I have provided the titles, labels, etc.

A screenshot of a computer

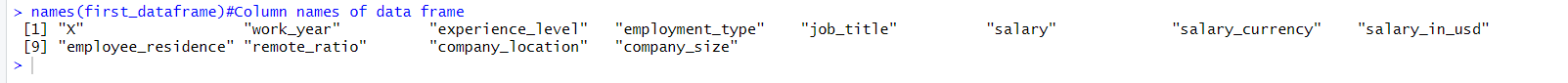
Description automatically generated

A close-up of a computer screen

Description automatically generated

A close-up of a computer screen

Description automatically generated



A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generatedA graph of a graph with text

Description automatically generated with medium confidence

**Conclusion: -**

In this module, I have learned the R language and how to use RStudio. I have also gained a basic understanding of arithmetic, logical, and statistical operators and functions. This has boosted my confidence in pursuit of my goal to become a data analyst. Learning about functions like 'runif,' 'rnorm,' and 'set.seed()' has been particularly valuable. Additionally, I have utilized various other operators such as 'range(),' 'mutate(),' 'slice(),' 'arrange(),' 'filter(),' etc. I have created numerous vectors, datasets, and data frames during this learning process.

I have successfully run all the test cases. Attaching screenshot for reference.

A screenshot of a computer

Description automatically generated

**Citations: -**

1. Zach. (2022). How (And When) to Use set.seed in R. Statology.

<https://www.statology.org/set-seed-in-r/#:~:text=The%20set.,time%20you%20run%20the%20code.Tu>

1. *What Is the Algorithm Used by the rnorm Function in R? | Saturn Cloud Blog*. (2023, September 9).

<https://saturncloud.io/blog/what-is-the-algorithm-used-by-the-rnorm-function-in-r/#:~:text=In%20R%2C%20the%20rnorm%20function%20is%20used%20to%20generate%20random,standard%20deviation%20of%20the%20distribution>.

1. H. T. (n.d.). Several ways to use runif.

<https://www.linkedin.com/pulse/several-ways-use-runif-henry-trunghieu-tu/>

1. Subset rows using their positions — slice. (n.d.).

<https://dplyr.tidyverse.org/reference/slice.html>

1. mutate function - RDocumentation. (n.d.). <https://www.rdocumentation.org/packages/dplyr/versions/0.5.0/topics/mutate>