

Assignment 1

Adithya Reddy Mettu

2024-02-04

#Data has been taken from ChatGPT

with the name of dataset 1

```
dataset.1 <- read.csv("C:/Users/adith/Downloads/dataset 1.csv")
dataset.1
```

```
##      Name Age Country
## 1   John  28     USA
## 2  Alice  24   Canada
## 3    Bob  32      UK
## 4    Eva  22 Germany
## 5   Sara  29   France
```

```
attach(dataset.1)
```

installed fbasics package for doing all the calculations automatically because in the summary function we are not able to find standard deviation which is included the fBasics library

#Task 1

```
library(fBasics)
options(scipen = 999) #used to round off the numbers
basicStats(data.frame(Age))
```

```
##              Age
## nobs          5.000000
## NAs           0.000000
## Minimum      22.000000
## Maximum      32.000000
## 1. Quartile  24.000000
## 3. Quartile  29.000000
## Mean        27.000000
## Median      28.000000
## Sum         135.000000
## SE Mean      1.788854
## LCL Mean     22.033344
## UCL Mean     31.966656
## Variance     16.000000
## Stdev        4.000000
## Skewness     -0.056250
## Kurtosis     -1.946875
```

Now we are going to perform descriptive statistics on the entire data set so that we can see what happens when we try to perform descriptive statistics on qualitative variables or categorical variables

```
summary(dataset.1)

##      Name      Age      Country
## Length:5      Min.   :22      Length:5
## Class :character 1st Qu.:24      Class :character
## Mode  :character Median :28      Mode  :character
##                  Mean  :27
##                  3rd Qu.:29
##                  Max.   :32
```

As we can see when we try to perform descriptive statistics on categorical variables or qualitative variables we get just get it as character

#Task 2

Now as calculating the descriptive statistics for quantitative values and categorical variables are done for the data in Age

We now move on to transforming one variable

```
library(caret)

## Loading required package: ggplot2
## Loading required package: lattice

Model_z_normalized <- preprocess(dataset.1,method = "range")
summary(predict(Model_z_normalized,dataset.1))

##      Name      Age      Country
## Length:5      Min.   :0.0      Length:5
## Class :character 1st Qu.:0.2      Class :character
## Mode  :character Median :0.6      Mode  :character
##                  Mean  :0.5
##                  3rd Qu.:0.7
##                  Max.   :1.0
```

#copying the values of the normalized data sets into new variable for further use

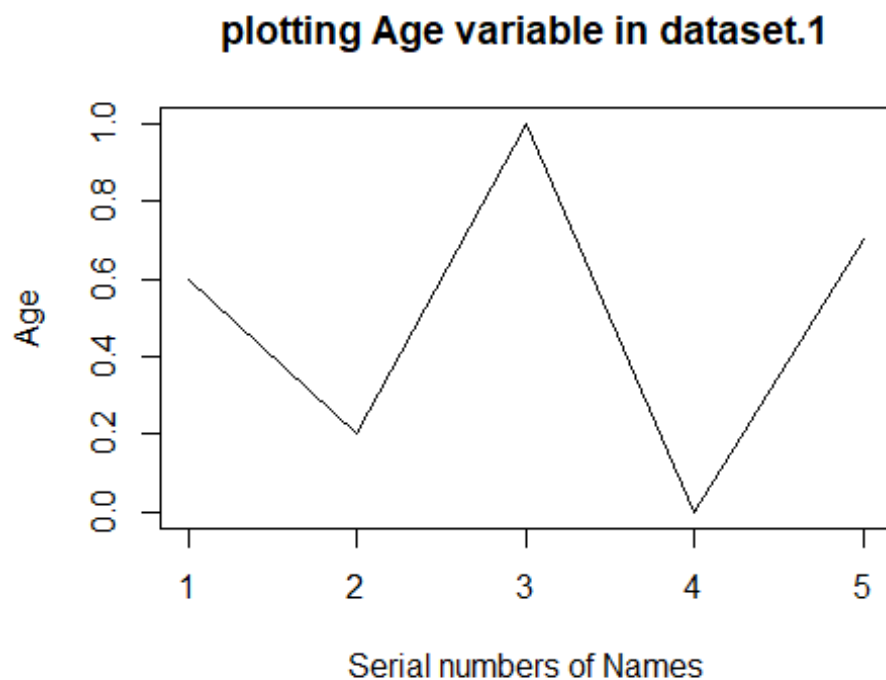
```
dataset1_replica <- predict(Model_z_normalized,dataset.1)
dataset1_replica

##   Name Age Country
## 1 John 0.6    USA
## 2 Alice 0.2  Canada
## 3 Bob 1.0    UK
## 4 Eva 0.0  Germany
## 5 Sara 0.7  France
```

Now as the normalization is done we move onto the plotting quantitative variable (for this instance Age)

#Task 3

```
#dataset_name $ column or row name is used for specifically selecting a row or a column in a dataset  
plot(dataset1_replica$Age,main="plotting Age variable in dataset.1", ylab =  
"Age", xlab = "Serial numbers of Names", type="l")
```



Now we are going to do Scatterplot

```
library(tidyverse)  
  
## — Attaching core tidyverse packages — tidyverse  
2.0.0 —  
## ✓ dplyr      1.1.4      ✓ readr      2.1.5  
## ✓ forcats    1.0.0      ✓ stringr    1.5.1  
## ✓ lubridate  1.9.3      ✓ tibble     3.2.1  
## ✓ purrr      1.0.2      ✓ tidyr      1.3.1  
## — Conflicts —  
tidyverse_conflicts() —  
## ✗ dplyr::filter() masks stats::filter()  
## ✗ dplyr::lag()     masks stats::lag()  
## ✗ purrr::lift()    masks caret::lift()  
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all  
conflicts to become errors
```

```
#c <- c(1,2,3,4,5) cmd to create a data
```

#plot(dataset1_replica\$Age) with just this we are not able to get scatter plot and in normal plot if we include the categorical variables for any axis it is showing an error where as in it is allowing all the qualitative variables as well

```
dataset1_replica %>% ggplot(aes(x=Name , y=Age)) +  
  geom_point()
```

