FML assignment 1

Aavula Neetu

2024-02-04

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

##Importing the data

# Install and load the 'readxl' package if not already installed  
if (!requireNamespace("readxl", quietly = TRUE)) {  
 install.packages("readxl")  
}  
library(readxl)  
Employee\_Sample\_Data <- read\_excel("C:/Users/saikr/Downloads/Employee Sample Data.xlsx")  
View(Employee\_Sample\_Data)

##Print out descriptive statistics for a selection of quantitative and categorical variables

summary(Employee\_Sample\_Data)

## EEID Full Name Job Title Department   
## Length:1000 Length:1000 Length:1000 Length:1000   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## Business Unit Gender Ethnicity Age   
## Length:1000 Length:1000 Length:1000 Min. :25.00   
## Class :character Class :character Class :character 1st Qu.:35.00   
## Mode :character Mode :character Mode :character Median :45.00   
## Mean :44.38   
## 3rd Qu.:54.00   
## Max. :65.00   
##   
## Hire Date Annual Salary Bonus %   
## Min. :1992-01-09 00:00:00.0 Min. : 40063 Min. :0.00000   
## 1st Qu.:2007-02-14 00:00:00.0 1st Qu.: 71430 1st Qu.:0.00000   
## Median :2014-02-15 12:00:00.0 Median : 96557 Median :0.00000   
## Mean :2012-04-07 02:54:14.4 Mean :113217 Mean :0.08866   
## 3rd Qu.:2018-06-22 00:00:00.0 3rd Qu.:150782 3rd Qu.:0.15000   
## Max. :2021-12-26 00:00:00.0 Max. :258498 Max. :0.40000   
##   
## Country City Exit Date   
## Length:1000 Length:1000 Min. :1994-12-18 00:00:00.00   
## Class :character Class :character 1st Qu.:2014-12-25 00:00:00.00   
## Mode :character Mode :character Median :2019-05-23 00:00:00.00   
## Mean :2016-11-02 18:04:14.11   
## 3rd Qu.:2021-04-09 00:00:00.00   
## Max. :2022-08-17 00:00:00.00   
## NA's :915

mean(Employee\_Sample\_Data$Age)

## [1] 44.382

#Descriptive statistics for variables

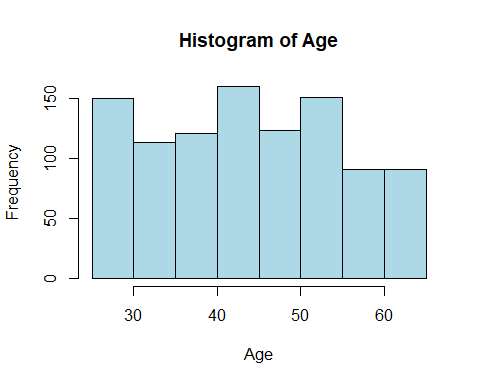
table(Employee\_Sample\_Data$Age)

##   
## 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50   
## 22 17 26 29 27 29 26 20 26 23 18 30 28 18 19 26 26 25 20 17 72 29 21 31 21 21   
## 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65   
## 24 30 22 25 50 16 19 15 19 22 18 16 19 23 15

#Transforming Variables

# Check for non-numeric values or zeros in 'mpg'  
if (any(!is.numeric(Employee\_Sample\_Data$Age)) || any(Employee\_Sample\_Data$Age == 0)) {  
 print("Cannot perform log transformation on non-numeric values or zeros in 'age'.")  
} else {  
 # Transform the 'age' variable using log transformation  
 Employee\_Sample\_Data$log\_age <- log(Employee\_Sample\_Data$Age)  
}

# Plot a histogram for 'Age'  
hist(Employee\_Sample\_Data$Age, main="Histogram of Age", xlab="Age", col="lightblue", border="black")



library(scatterplot3d)  
# Plot a scatterplot  
  
scatterplot3d(Employee\_Sample\_Data$Age, Employee\_Sample\_Data$`Annual Salary`, main="3D Scatterplot", xlab="AGE", ylab="Annual salary", color="blue")

