BA 64060 Assignment 1

Kyle Breeden

2025-09-07

#Use read.csv function to import dataset found on Kaggle.com  
#Use readr package for faster and more efficient reading per class slides  
#install.packages("readr") #not needed if already installed  
library(readr)

## Warning: package 'readr' was built under R version 4.5.1

assign1\_dataset <- read\_csv("C:/Users/kylej/OneDrive/Documents/BA 64060/Assignment 1/Data/WA\_Fn-UseC\_-HR-Employee-Attrition.CSV")

## Rows: 1470 Columns: 35  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (9): Attrition, BusinessTravel, Department, EducationField, Gender, Job...  
## dbl (26): Age, DailyRate, DistanceFromHome, Education, EmployeeCount, Employ...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

#Use Spec function to see column specifications  
spec(assign1\_dataset)

## cols(  
## Age = col\_double(),  
## Attrition = col\_character(),  
## BusinessTravel = col\_character(),  
## DailyRate = col\_double(),  
## Department = col\_character(),  
## DistanceFromHome = col\_double(),  
## Education = col\_double(),  
## EducationField = col\_character(),  
## EmployeeCount = col\_double(),  
## EmployeeNumber = col\_double(),  
## EnvironmentSatisfaction = col\_double(),  
## Gender = col\_character(),  
## HourlyRate = col\_double(),  
## JobInvolvement = col\_double(),  
## JobLevel = col\_double(),  
## JobRole = col\_character(),  
## JobSatisfaction = col\_double(),  
## MaritalStatus = col\_character(),  
## MonthlyIncome = col\_double(),  
## MonthlyRate = col\_double(),  
## NumCompaniesWorked = col\_double(),  
## Over18 = col\_character(),  
## OverTime = col\_character(),  
## PercentSalaryHike = col\_double(),  
## PerformanceRating = col\_double(),  
## RelationshipSatisfaction = col\_double(),  
## StandardHours = col\_double(),  
## StockOptionLevel = col\_double(),  
## TotalWorkingYears = col\_double(),  
## TrainingTimesLastYear = col\_double(),  
## WorkLifeBalance = col\_double(),  
## YearsAtCompany = col\_double(),  
## YearsInCurrentRole = col\_double(),  
## YearsSinceLastPromotion = col\_double(),  
## YearsWithCurrManager = col\_double()  
## )

#Use dplyr package per subsetting data in R reading.  
#Select quantitative and categorical variables from the above list to focus on.  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

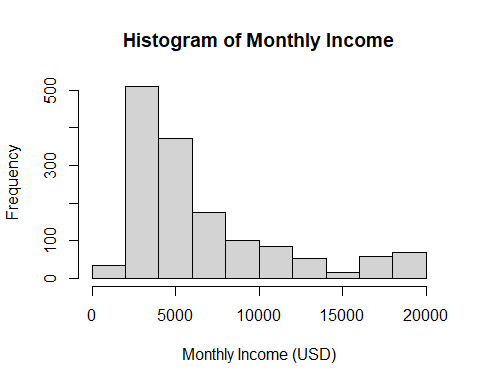
## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

subset\_assign1 <- assign1\_dataset %>%   
 select(Age, MonthlyIncome, YearsAtCompany, Attrition, Department, JobRole)  
#Then print descriptive stats for the subset  
summary(subset\_assign1)

## Age MonthlyIncome YearsAtCompany Attrition   
## Min. :18.00 Min. : 1009 Min. : 0.000 Length:1470   
## 1st Qu.:30.00 1st Qu.: 2911 1st Qu.: 3.000 Class :character   
## Median :36.00 Median : 4919 Median : 5.000 Mode :character   
## Mean :36.92 Mean : 6503 Mean : 7.008   
## 3rd Qu.:43.00 3rd Qu.: 8379 3rd Qu.: 9.000   
## Max. :60.00 Max. :19999 Max. :40.000   
## Department JobRole   
## Length:1470 Length:1470   
## Class :character Class :character   
## Mode :character Mode :character   
##   
##   
##

#Use mutate function to transform attrition variable to 0/1  
library(dplyr)  
subset\_assign1 <- subset\_assign1 %>% mutate(Attrition = if\_else(Attrition== "Yes", 1, 0))  
#check to make sure it worked  
View(subset\_assign1)

#Use hist function to plot one variable: monthly income  
hist(subset\_assign1$MonthlyIncome,  
 main = "Histogram of Monthly Income",  
 xlab = "Monthly Income (USD)")



#Use plot function to create scatterplot of Age against Monthly Income  
plot(subset\_assign1$Age, subset\_assign1$MonthlyIncome,  
 main = "Age vs Monthly Income",  
 xlab = "Age (years)",  
 ylab = "Monthly Income (USD)")

