ISM 4403 Homework Week 3

### Tasks:

Create a new Excel spreadsheet from the following table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| id | Height (inches) | gender | Hair color | Eye Color |
| 1 | 67 | male | brown | brown |
| 2 | 64 | female | brown | green |
| 3 | 74 | male | blond | brown |
| 4 | 73 | Male | brown | brown |
| 5 |  | female | red | green |
| 6 | 61 | female |  | green |
| 7 | 73 | female | blond | blue |
| 8 | 70 | female | brown | blue |
| 9 |  | female | blond |  |
| 10 |  | male | blond | brown |
| 11 | 74 |  | brown | brown |

Import the sheet into R as a dataset using the read package. **check**

**Recode the missing values in the height, hair color and eye color fields to be the mean, mode**, mode respectively. If gender is missing a value exclude the row from the dataset. **check**

To recode the height to be the mean for males and females respectively. In summary, you must enter the mean value for females into the height for any females with missing data. You must also enter the mean value for males into the height for any males with missing data. You MUST use R’s recode functionality to do this. **check**

For any missing data in the Hair color and eye color fields insert the mode of that field. **failedf**

Hint the following template can be used to accomplish this.

x[is.na(x)] <- mean(x, na.rm = TRUE)

x = na.omit(x)

Once you have filled in any missing data calculate the following items.

The mean height.

The sum of all heights.

The mode for gender, hair color, and eye color. **Check I think**

**Paste your code here**

**# importing data also taking care of nas**

**dataset = read.csv('lab3\_flat\_file.csv')**

**dataset <- dataset[-c(11), ]**

**head(dataset)**

**dataset$height\_in[which(is.na(dataset$height\_in))] <- mean(dataset$height\_in, na.rm = TRUE)**

**#install.packages("dplyr")**

**#install.packages("tidyverse")**

**install.packages("caret")**

**#load dplyr**

**library(dplyr)**

**library(caret)**

**# I have never tried to use mode on nominal or ordinal data without endcoing it**

**dataset$gender <- factor(dataset$gender, level = c("female", "male"), labels =c(1,2))**

**dataset$hair\_color <- factor(dataset$hair\_color, level = c("brown", "blond", "red"), labels =c(1,2,3))**

**dataset$eye\_color <- factor(dataset$eye\_color, level = c("brown", "green", "blue"), labels =c(1,2,3))**

**head(dataset)**

**# males**

**# go back into excel and change 1 obs. to male not Male**

**male = filter(dataset, gender == "male")**

**mean(male$height\_in)**

**# female**

**female = filter(dataset, gender == "female")**

**mean(female$height\_in)**

**# summary of data**

**mean(dataset$height\_in)**

**sum(dataset$height\_in)**

**summary(dataset)**

**count\_gender <- table(dataset$gender)**

**count\_hair <- table(dataset$hair\_color)**

**count\_eye <- table(dataset$eye\_color)**

**End of Paste**

**Paste your results here**

|  |
| --- |
| # importing data also taking care of nas  > dataset = read.csv('lab3\_flat\_file.csv')  > dataset <- dataset[-c(11), ]  > head(dataset)  ï..id height\_in gender hair\_color eye\_color  1 1 67 male brown brown  2 2 64 female brown green  3 3 74 male blond brown  4 4 73 male brown brown  5 5 NA female red green  6 6 61 female green  > dataset$height\_in[which(is.na(dataset$height\_in))] <- mean(dataset$height\_in, na.rm = TRUE)  > #install.packages("dplyr")  > #install.packages("tidyverse")  > install.packages("caret")  Error in install.packages : Updating loaded packages  > #load dplyr  > library(dplyr)  > library(caret)  >  > # I have never tried to use mode on nominal or ordinal data without endcoing it  > dataset$gender <- factor(dataset$gender, level = c("female", "male"), labels =c(1,2))  > dataset$hair\_color <- factor(dataset$hair\_color, level = c("brown", "blond", "red"), labels =c(1,2,3))  > dataset$eye\_color <- factor(dataset$eye\_color, level = c("brown", "green", "blue"), labels =c(1,2,3))  > head(dataset)  ï..id height\_in gender hair\_color eye\_color  1 1 67.00000 2 1 1  2 2 64.00000 1 1 2  3 3 74.00000 2 2 1  4 4 73.00000 2 1 1  5 5 68.85714 1 3 2  6 6 61.00000 1 <NA> 2  >  > # males  > # go back into excel and change 1 obs. to male not Male  > male = filter(dataset, gender == "male")  > mean(male$height\_in)  [1] NaN  >  > # female  > female = filter(dataset, gender == "female")  > mean(female$height\_in)  [1] NaN  >  > # summary of data  > mean(dataset$height\_in)  [1] 68.85714  > sum(dataset$height\_in)  [1] 688.5714  > summary(dataset)  ï..id height\_in gender hair\_color eye\_color  Min. : 1.00 Min. :61.00 1:6 1 :4 1 :4  1st Qu.: 3.25 1st Qu.:67.46 2:4 2 :4 2 :3  Median : 5.50 Median :68.86 3 :1 3 :2  Mean : 5.50 Mean :68.86 NA's:1 NA's:1  3rd Qu.: 7.75 3rd Qu.:72.25  Max. :10.00 Max. :74.00  > count\_gender <- table(dataset$gender)  > count\_hair <- table(dataset$hair\_color)  > count\_eye <- table(dataset$eye\_color)  > install.packages("caret")  WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:  https://cran.rstudio.com/bin/windows/Rtools/  Installing package into ‘C:/Users/Owner/Documents/R/win-library/3.6’  (as ‘lib’ is unspecified)  Warning in install.packages :  package ‘caret’ is in use and will not be installed |
|  |
| |  | | --- | | > | |

**End of Paste**

Using the Quantmod interface get data for Boeing, Apple, Nvida, AMD, and Intel. (See the following for more information <https://github.com/joshuaulrich/quantmod>) Yahoo finance should be used as your source. **check**

Print the mean stock price for each stock **check**

Chart it in a Bar chart **check**

**Paste your resulting code here**

|  |
| --- |
| > ## part 2 quantmod  >  > #install.packages("quantmod")  > library("quantmod")  >  > # apple stock AAPL  > getSymbols("AAPL", src = "yahoo")  [1] "AAPL"  Warning message:  'indexClass<-' is deprecated.  Use 'tclass<-' instead.  See help("Deprecated") and help("xts-deprecated").  > # boeing stock BA  > getSymbols("BA", src = "yahoo")  [1] "BA"  Warning message:  'indexClass<-' is deprecated.  Use 'tclass<-' instead.  See help("Deprecated") and help("xts-deprecated").  > # nvida stock  > getSymbols("NVDA", src = "yahoo")  [1] "NVDA"  Warning message:  'indexClass<-' is deprecated.  Use 'tclass<-' instead.  See help("Deprecated") and help("xts-deprecated").  > # AMD stock  > getSymbols("AMD", src = "yahoo")  [1] "AMD"  Warning message:  'indexClass<-' is deprecated.  Use 'tclass<-' instead.  See help("Deprecated") and help("xts-deprecated").  > # intel stock  > getSymbols("INTC", src = "yahoo")  [1] "INTC"  Warning message:  'indexClass<-' is deprecated.  Use 'tclass<-' instead.  See help("Deprecated") and help("xts-deprecated").  >  > # apple stock AAPL  > getSymbols("AAPL", src = "yahoo")  [1] "AAPL"  Warning message:  'indexClass<-' is deprecated.  Use 'tclass<-' instead.  See help("Deprecated") and help("xts-deprecated").  > # use chart series  > addMACD()  > addBBands()  >  > mean\_apple <- mean(AAPL$AAPL.Close)  > mean\_amd <- mean(AMD$AMD.Close)  > mean\_ba <- mean(BA$BA.Close)  > mean\_intc <- mean(INTC$INTC.Close)  > df2 <- data.frame(mean\_apple, mean\_amd, mean\_ba, mean\_intc)  > # you need this make the df a matrix before you can plot the data  > df2 = as.matrix(df2)  > merge(mean\_apple, mean\_amd, mean\_ba, mean\_intc, all = TRUE)  Error in fix.by(by.x, x) : 'by' must match numbers of columns  >  > barplot(df2, main = "Mean Closing Price",  + xlab = "Stock Ticker", ylab = "Price in USD", las = 1)  > help(barplot) |
|  |
| |  | | --- | | > | |

**END OF Paste**

**Paste your resulting graph here**



****

**END of Paste**

**Rubric:**

30 points for pasting correct summary information.

35 For connecting to the GDI system by installing the module.

35 for generating a Graph from the Quantmod system.