Final project r script

USA <- read.csv("~/workspace/SNHU/DAT-500/finalproject/USA.csv")

GBR <- read.csv("~/workspace/SNHU/DAT-500/finalproject/GBR.csv")

summary(USA)

summary(GBR)

#While alot is equivalent between the two data sets, USA has twice

#as many elements. GBR has a higher mean temp, while USA has a

#wider range of temps. Rainfall almost doubles in GBR versus USA.

#In contrast to temp, USA has a smaller range of rainfall, while

#GBR has over 3.5X the range in rainfall compared to USA

tempUSAavg <- mean(USA$Temperature..C.)

tempUSAmin <- min(USA$Temperature..C.)

tempUSAmax <- max(USA$Temperature..C.)

rainUSAavg <- mean(USA$Rainfall..mm.)

rainUSAmin <- min(USA$Rainfall..mm.)

rainUSAmax <- max(USA$Rainfall..mm.)

tempGBRavg <- mean(GBR$Temperature..C.)

tempGBRmin <- min(GBR$Temperature..C.)

tempGBRmax <- max(GBR$Temperature..C.)

rainGBRavg <- mean(GBR$Rainfall..mm.)

rainGBRmin <- min(GBR$Rainfall..mm.)

rainGBRmax <- max(GBR$Rainfall..mm.)

#the above variables are created to pull the exact values from specific

#data sets. This is important because you can create a variable for a specific

#number through the $ sign-just like in excel. To me, right now, that is

#invaluable.

rain\_nums <- c(rainGBRavg,rainGBRmax,rainGBRmin,

rainUSAavg,rainUSAmax,rainUSAmin)

rnums =rain\_nums

names(rnums)=c("Avg GB rain","Max GB rain","Min GB rain",

"Avg US rain","Max US rain","Min US rain")

rnums

temp\_nums <- c(tempGBRavg,tempGBRmax,tempGBRmin,

tempUSAavg,tempUSAmax,tempUSAmin)

tnums =temp\_nums

names(tnums)=c("Avg GB temp","Max GB temp","Min GB temp",

"Avg US temp","Max US temp","Min US temp")

tnums

matF = matrix(

c(rnums,tnums),

nrow=2,

ncol=6,

byrow = TRUE) #we want to fill the matrix by rows!

matF

dimnames(matF)=list(c("rain","temp"),

c("GB avg","GB max","GB min",

"US avg","US max","US min"))

matF

df=data.frame(rnums,tnums)

df

#so the difference between matrix matF and the data frame df is

#aesthetics. matF is a 2X6 matrix while the df is a 6X2 matrix.

#I do understand I can pair the averages together and so on to provide

#a cleaner look to the data. I acknowledge in business, it is

#imperative to yield clean, accurate, precise and good looking results

#for both the technical and non-technical audiences.

############################################################################

#######CLEAN R SCRIPT BELOW#################################################

USA <- read.csv("~/workspace/SNHU/DAT-500/finalproject/USA.csv")

GBR <- read.csv("~/workspace/SNHU/DAT-500/finalproject/GBR.csv")

summary(USA)

summary(GBR)

tempUSAavg <- mean(USA$Temperature..C.)

tempUSAmin <- min(USA$Temperature..C.)

tempUSAmax <- max(USA$Temperature..C.)

rainUSAavg <- mean(USA$Rainfall..mm.)

rainUSAmin <- min(USA$Rainfall..mm.)

rainUSAmax <- max(USA$Rainfall..mm.)

tempGBRavg <- mean(GBR$Temperature..C.)

tempGBRmin <- min(GBR$Temperature..C.)

tempGBRmax <- max(GBR$Temperature..C.)

rainGBRavg <- mean(GBR$Rainfall..mm.)

rainGBRmin <- min(GBR$Rainfall..mm.)

rainGBRmax <- max(GBR$Rainfall..mm.)

rain\_nums <- c(rainGBRavg,rainGBRmax,rainGBRmin,

rainUSAavg,rainUSAmax,rainUSAmin)

rnums =rain\_nums

names(rnums)=c("Avg GB rain","Max GB rain","Min GB rain",

"Avg US rain","Max US rain","Min US rain")

rnums

temp\_nums <- c(tempGBRavg,tempGBRmax,tempGBRmin,

tempUSAavg,tempUSAmax,tempUSAmin)

tnums =temp\_nums

names(tnums)=c("Avg GB temp","Max GB temp","Min GB temp",

"Avg US temp","Max US temp","Min US temp")

tnums

matF = matrix(

c(rnums,tnums),

nrow=2,

ncol=6,

byrow = TRUE)

matF

dimnames(matF)=list(c("rain","temp"),

c("GB avg","GB max","GB min",

"US avg","US max","US min"))

matF

df=data.frame(rnums,tnums)

df