Assignment 9

Matt Jesser

October 20, 2018

Dataset1 <- read.csv("C:/Users/student/Documents/MATH421/Dataset1 .csv")

# 1

Missing <- function(x) {  
 return(sum(is.na(x)))  
}  
Missing(Dataset1)

## [1] 4

# 2

MissingNames <- function(x) {  
 return(names(which(colSums(is.na(x))>0)))  
}  
MissingNames(Dataset1)

## [1] "Age" "GPA"

# 3

SpecificMissingData <- function(x) {  
 Columns <- names(which(colSums(is.na(x))>0))  
 Numbers <- sapply(x, function(x) sum(is.na(x)))  
 list(Columns, Numbers)  
}  
SpecificMissingData(Dataset1)

## [[1]]  
## [1] "Age" "GPA"  
##   
## [[2]]  
## Age Gender GPA   
## 2 0 2

# 4

ReplaceData <- function(x) {  
 for (i in 1:ncol(x)){  
 if (is.numeric(x[,i])){  
 x[,i][is.na(x[,i])]=mean(x[,i], na.rm=TRUE)  
 }else{  
 levels=unique(x[,i])  
 x[,i][is.na(x[,i])]=levels[which.max(tabulate(match(x[,i], levels)))]  
 }  
 }  
 return (head(x))  
}  
ReplaceData(Dataset1)

## Age Gender GPA  
## 1 17 M 3.0000  
## 2 18 F 3.1000  
## 3 18 3.7000  
## 4 17 M 3.2125  
## 5 21 3.7000  
## 6 19 F 2.1000

# 5

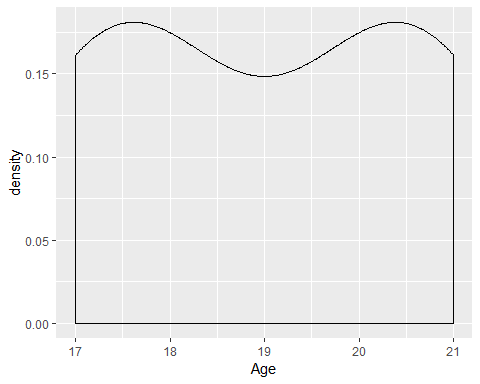
Q5 <- function(x){  
 num\_vars=c()  
 cat\_vars=c()  
 for (i in 1:ncol(x)){  
 if(is.numeric(x[,i])){  
 num\_vars=c(num\_vars,names(x)[i])  
 }else{  
 cat\_vars=c(cat\_vars,names(x)[i])  
 }  
 }  
 num\_data=x[num\_vars]  
 cat\_data=x[cat\_vars]  
 print(head(num\_data))  
 print(head(cat\_data))  
}  
Q5(Dataset1)

## Age GPA  
## 1 17 3.0  
## 2 18 3.1  
## 3 18 3.7  
## 4 17 NA  
## 5 21 3.7  
## 6 NA 2.1  
## Gender  
## 1 M  
## 2 F  
## 3   
## 4 M  
## 5   
## 6 F

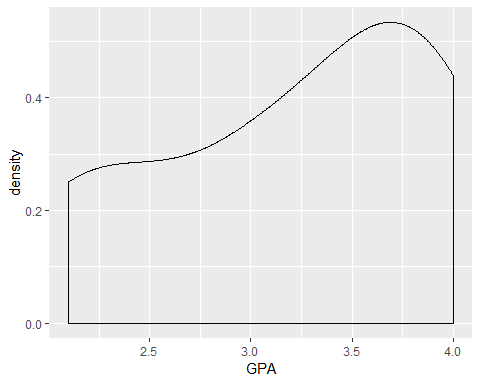
# 6

library(ggplot2)  
  
Q6 <- function(x){  
 for (i in 1:ncol(x)){  
 if (is.numeric(x[,i])){  
 print(ggplot(data=x)+geom\_density(mapping=aes(x=x[,i]))+xlab(names(x)[i]))  
 }  
 }  
}  
Q6(Dataset1)

## Warning: Removed 2 rows containing non-finite values (stat\_density).

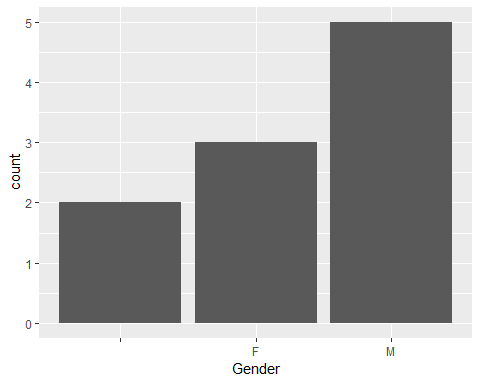


## Warning: Removed 2 rows containing non-finite values (stat\_density).



# 7

Q7 <- function(x){  
 for (i in 1:ncol(x)){  
 if (!is.numeric(x[,i])){  
 print(ggplot(data=x)+geom\_bar(mapping=aes(x=x[,i]))+xlab(names(x)[i]))  
 }  
 }  
}  
Q7(Dataset1)



# 8

Q8 <- function(x,i,j){  
 if(is.numeric(x[,i]) | is.numeric(x[,j])){  
 print("not categorical variables")  
 }else {  
 print(ggplot(data=x)+geom\_bar(mapping=aes(x=x[,i], fill=x[,j]),position = "dodge")+labs(x=names(x)[i],fill=names(x)[j]))  
 }  
 }  
  
Q8(Dataset1,2,3)

## [1] "not categorical variables"

Q8(Dataset1,1,3)

## [1] "not categorical variables"

# 9

Q9 <- function(x,i,j){  
 if(!is.numeric(x[,i]) | !is.numeric(x[,j])){  
 print("not numeric variables")  
 }else {  
 print(ggplot(data=x)+geom\_point(mapping=aes(x=x[,i], y=x[,j]))+labs(x=names(x)[i],y=names(x)[j]))  
 }  
}  
  
Q9(Dataset1,1,3)

## Warning: Removed 4 rows containing missing values (geom\_point).



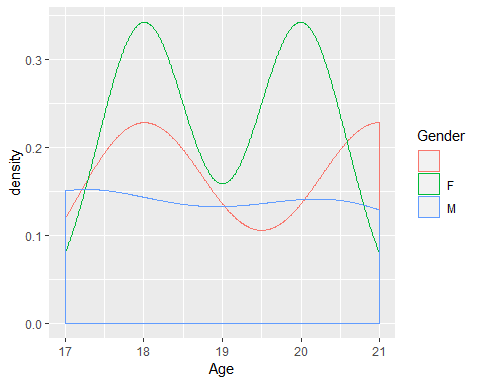
Q9(Dataset1,1,2)

## [1] "not numeric variables"

# 10

Q10 <- function(x,i,j){  
 if(!is.numeric(x[,i])){  
 print("not numeric")  
 }else if(is.numeric(x[,j])){  
 print("not categorical")  
 }else{  
 print(ggplot(data=x)+geom\_density(mapping = aes(x=x[,1],color=x[,j]))+labs(x=names(x)[i],color=names(x)[j]))  
 }  
}  
  
Q10(Dataset1,1,2)

## Warning: Removed 2 rows containing non-finite values (stat\_density).



Q10(Dataset1,2,3)

## [1] "not numeric"

Q10(Dataset1,1,3)

## [1] "not categorical"

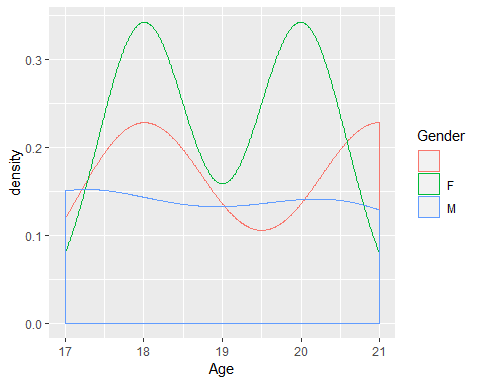
# 11

Q11 <- function(x){  
 for (i in 1:ncol(x)){  
 if (!is.numeric(x[,i])){  
 for(j in 1:ncol(x)){  
 if (!is.numeric(x[,j]) & names(x)[j]!=names(x)[i]){  
 print(ggplot(data=x)+geom\_bar(mapping=aes(x=x[,i],fill=x[,j]),position="dodge")+ labs(x=names(x)[i],fill=names(x)[j]))  
 }  
 }  
 }  
 }  
}  
Q11(Dataset1)

# 12

Q12 <- function(x){  
 for (i in 1:ncol(x)){  
 if (!is.numeric(x[,i])){  
 for(j in 1:ncol(x)){  
 if (!is.numeric(x[,j]) & names(x)[j]!=names(x)[i]){  
 print(ggplot(data=x)+geom\_bar(mapping=aes(x=x[,i],fill=x[,j]),position="dodge")+ labs(x=names(x)[i],fill=names(x)[j]))  
 }  
 }  
 }else{  
 for(j in 1:ncol(x)){  
 if (!is.numeric(x[,j])){  
 print(ggplot(data=x)+geom\_density(mapping=aes(x=x[,i],color=x[,j]))+labs(x=names(x)[i],color=names(x)[j]))  
 }  
 }  
 }  
 }  
}  
Q12(Dataset1)

## Warning: Removed 2 rows containing non-finite values (stat\_density).



## Warning: Removed 2 rows containing non-finite values (stat\_density).

