Datenprojekt1

R.C. Wallner

2024-11-22

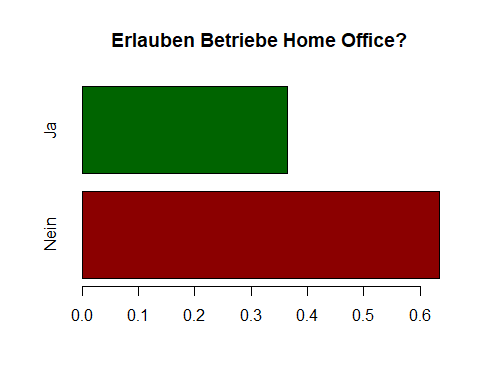
###1  
df1 <- data.frame(Betrieb = data$ID\_betnr,HO = data$bhomeoff)  
df2 <- data.frame(Betrieb = 1, HO = 0) #initialisieren mit Erstem Betrieb  
for(i in 1:nrow(df1)){  
 if( ! ( df1[i,1] %in% df2[,1] ) ){  
 df2 <- rbind(df2, df1[i,])  
 }  
}  
#Wenn die Betriebnummer noch nicht in der bisherigen Spalte 1 des neuen  
#Datenrahmens aufgenommen wurde, dann wird die Zeile hinzugefügt   
  
nrow(df2) == length(unique(df1[,1])) #TRUE heißt alle Betirb-Duplikate eliminert

## [1] TRUE

df2 <- df2[-nrow(df2), ] #Die NA-Betriebe kommen auch 1 mal ganz am Ende vor  
df2$HO <- factor(df2$HO, levels=c(0,1), labels = c("Nein","Ja"))  
Frequency\_Table <- prop.table(table(df2$HO))  
Share\_HO <- as.numeric(Frequency\_Table[2])  
Share\_HO

## [1] 0.3653137

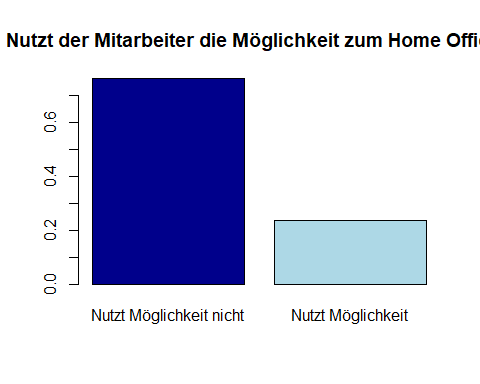
barplot(Frequency\_Table, main="Erlauben Betriebe Home Office?",  
 col=c("darkred","darkgreen"), horiz = T)



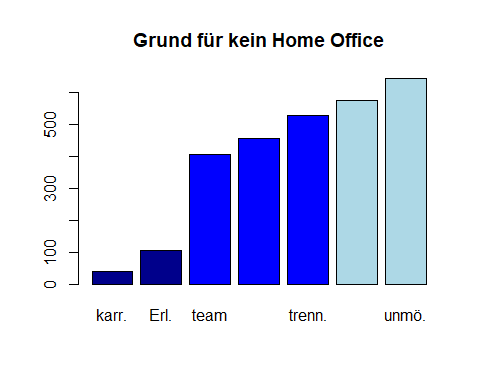
###2  
df3 <- cbind(df1, Nutzen = data$mheim)[df1$HO==1,]  
df4 <- data.frame(Betrieb = 6, HO = 1, Nutzen = 1)  
for(i in 2:nrow(df3)){  
 enthält <- F  
 for(j in 1:3){  
 if( is.na(df3[i,j]) ){enthält <- T}  
 }  
 if( !enthält){  
 df4 <- rbind(df4, df3[i,])  
 }  
}  
df4$Nutzen <- factor(df4$Nutzen, levels=c(0,1), labels= c(  
 "Nutzt Möglichkeit nicht", "Nutzt Möglichkeit"  
))  
Frequency\_Use <- prop.table(table(df4$Nutzen))  
Frequency\_Use #Knapp 1/4 genau 23.63% nutzen die HO Möglichkeit

##   
## Nutzt Möglichkeit nicht Nutzt Möglichkeit   
## 0.7636816 0.2363184

barplot(Frequency\_Use, col=c("darkblue", "lightblue"),  
 main= "Nutzt der Mitarbeiter die Möglichkeit zum Home Office?")

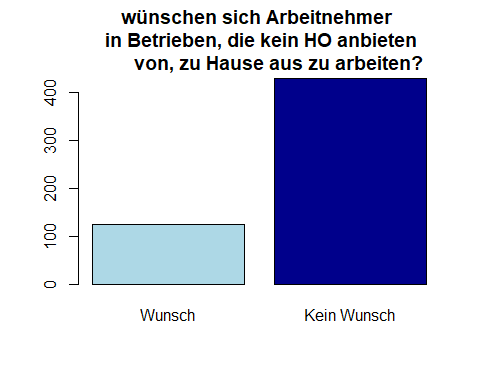


###3  
Relevante\_Variablen <- character()  
for(i in 1:length(colnames(data))){  
 if( grepl("mheimnein\_", colnames(data)[i] ) ){  
 Relevante\_Variablen <- c( Relevante\_Variablen, colnames(data)[i])  
 }  
} #alle Variablen, die mheimnein\_ enthalten, also hier zweckdienlich sind  
Index <- which(colnames(data) %in% Relevante\_Variablen )  
df5 <- na.omit( cbind(df1, data$mheim, data[,Index]) )  
nj <- numeric()  
for(j in 4:ncol(df5)){  
 nj <- c(nj, sum(df5[,j]) )  
}  
Relevante\_Variablen2 <- c("Erl.", "Tech.", "unmö.", "vorg.", "trenn.", "team", "karr.")  
Gründe <- data.frame(Relevante\_Variablen2, nj)[order(nj), ]  
barplot(Gründe$nj, names.arg= Gründe$Relevante\_Variablen2,  
 col=c(rep("darkblue",2),rep("blue",3),rep("lightblue",2)) ,  
 main="Grund für kein Home Office")



#Die meisten geben einfach nur an, dass es Ihnen nicht möglich ist,  
#am zweit häufigsten ist der Grund, dass Vorgesetzten die Anwesenheit "sehr wichtig" ist

###4  
df6 <- factor(na.omit(subset( data.frame(   
 wunsch3 = data$mheimwunsch,   
 HO = data$bhomeoff),  
 HO == 0))[,1],levels=1:3, labels = c(rep("Wunsch",2),"Kein Wunsch") )  
#df6 ist ein Vektor der alle mheimwunsch für HO = 0 herausfindet, und diese dann in  
#Wunscht besteht: Werte 1,2 und Wunsch besteht nicht: Wert 3 aufteilt  
barplot(table(df6), main = "wünschen sich Arbeitnehmer \n in Betrieben, die kein HO anbieten  
 von, zu Hause aus zu arbeiten?",  
 col=c("lightblue","darkblue"))



###5  
Wichtige\_Variablen <- character()  
for(i in 1:ncol(data)){  
 if( grepl("mheimwunsch\_", colnames(data)[i])){  
 Wichtige\_Variablen <- c(Wichtige\_Variablen, colnames(data)[i])  
 }  
}  
Index2 <- which(colnames(data) %in% Wichtige\_Variablen)  
df7 <- na.omit(data[, Index2])  
Gründe2 <- c("fahren","freiz.","fam","qual.","h\_erhöhen")  
Häufigkeiten <- apply(df7, 2, sum)  
barplot(Häufigkeiten, names.arg=Gründe2,  
 col="lightblue")

