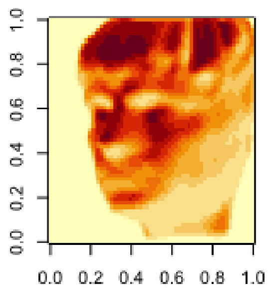
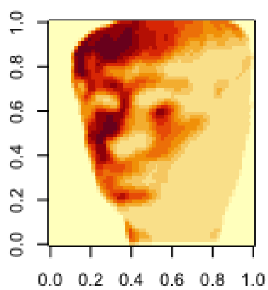
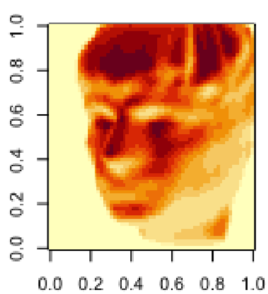
Homework 2

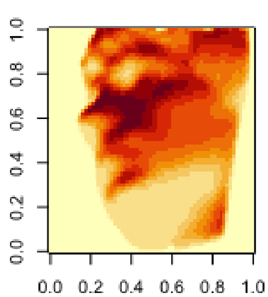
Yiwei He (yh9vhg), Da Lin (dl2de), Ziyue Jin (zj5qj)

Question 1

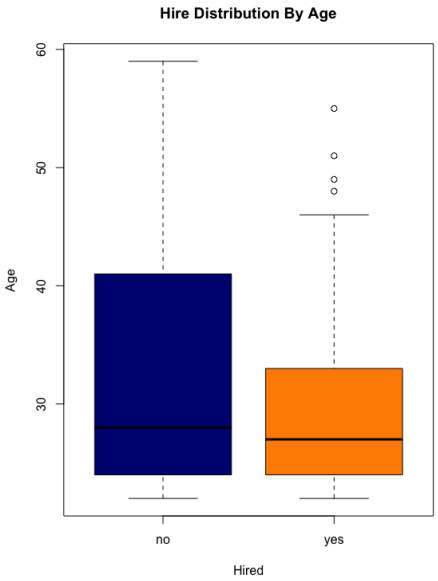


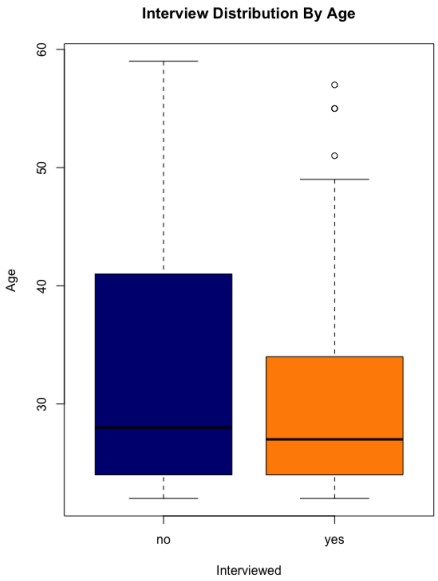






Question 2



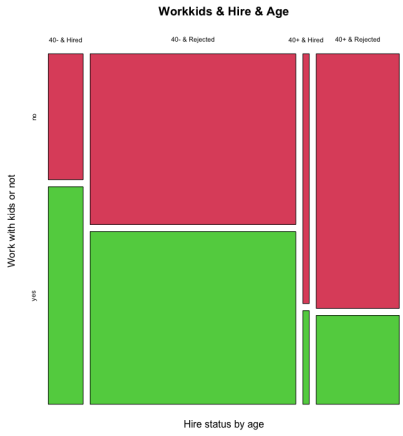


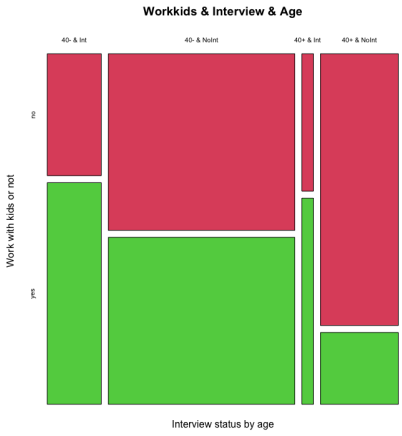
Introduction (Age vs. Interview/Hire):

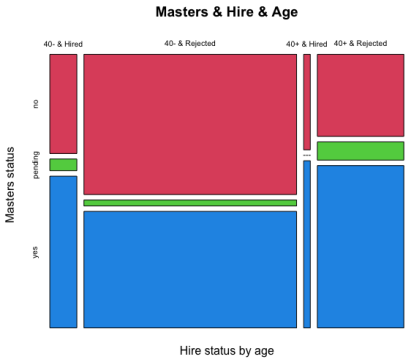
- There seems to be distinct discrimination against older age applicants in both interviews and hire.
- Older applicants seem more likely to be rejected, with a larger spread in ages than the accepted population.
- The accepted population has a smaller spread in ages but high outliers.
- Not sufficient for determining a causal relationship, as we need to explore more attributes.

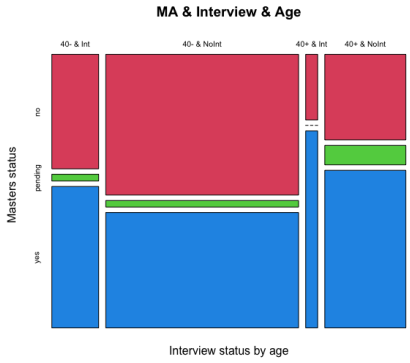
Attribute Analysis (Work With Kids):

- Regardless of age, applicants who have worked with kids seem more likely to receive interviews, which indicates that age might not solely impact offering interviews.
- Only around 30% of applicants aged 40+ have worked with kids as compared to more than 60% for 40- applicants; this might explain why younger applicants seem more likely to be hired.









Attribute Analysis (MA Status):

- Regardless of age, applicants with a MA or a pending MA seem more likely to be hired/interviewed

Conclusion:

Although the direct relationship from Introduction implies age discrimination, as we explore more attributes such as working with kids and MA status, we observe potential confounding variables that might explain why younger applicants are more preferable in the interviewing/hiring process. To back up our results, we might need further statistics modeling or inference procedures.

Appendix

```
#Question1
face <- read.table("/Users/a/Desktop/STAT 3280/HW2/face-data.txt")
face1=face[,1]
face2=face[,2]
face3=face[,3]
face4=face[,4]
mat1=matrix(face1,nrow=64,ncol=64)
mat2=matrix(face2,nrow=64,ncol=64)
mat3=matrix(face3,nrow=64,ncol=64)
mat4=matrix(face4,nrow=64,ncol=64)
mat11=apply(t(mat1),2,rev)
mat111=apply(t(mat11),2,rev)
mat1111=apply(t(mat111),2,rev)
mat22=apply(t(mat2),2,rev)
mat222=apply(t(mat22),2,rev)
mat2222=apply(t(mat222),2,rev)
mat22222=apply(t(mat2222),1,rev)
mat33=apply(t(mat3),2,rev)
mat333=apply(t(mat33),2,rev)
mat3333=apply(t(mat333),2,rev)
mat44=apply(t(mat4),2,rev)
mat444=apply(t(mat44),2,rev)
mat4444=apply(t(mat444),2,rev)
mat44444=apply(t(mat4444),1,rev)
par(mfrow=c(1,4))
image(mat1111)
image(mat22222)
image(mat3333)
image(mat44444)

#Question2
library(YaleToolkit)
x <- read.csv("/Users/a/Desktop/STAT 3280/HW2/TeacherHires.csv")
class(x$Hired)
x <- read.csv("/Users/a/Desktop/STAT 3280/HW2/TeacherHires.csv",stringsAsFactors = TRUE)
class(x$Hired)
##Analysis
par(mfrow=c(1,2))
mosaicplot(table(x$hired40,x$workkids),xlab="Hire status by age",
        ylab="Work with kids or not",main = "Workkids & Hire & Age",col=c(2,3))
mosaicplot(table(x$int40,x$workkids),xlab="Interview status by age",
        ylab="Work with kids or not",main = "Workkids & Interview & Age",col=c(2,3))
```

```
mosaicplot(table(x$hired40,x$MA),xlab="Hire status by age",
        ylab="Masters status",main = "Masters & Hire & Age",col=c(2,3,4))
mosaicplot(table(x$int40,x$MA),xlab="Interview status by age",
        ylab="Masters status",main = "MA & Interview & Age",col=c(2,3,4))
###
par(mfrow=c(1,2))
barplot(x=x$hired,y=x$age,xlab="Hired",ylab="Age",
    col=c("navy","darkorange"),main="Hire Distribution By Age")
plot(x=x$interviewed,y=x$age,xlab="Interviewed",ylab="Age",
    col=c("navy","darkorange"),main="Interview Distribution By Age")
```