Denzel

```
## Warning: package 'knitr' was built under R version 4.0.4
```

```
## Warning: package 'venneuler' was built under R version 4.0.3
```

```
## Loading required package: rJava
```

```
## Warning: package 'rJava' was built under R version 4.0.3
```

27 marks

The actor Denzel Washington has appeared in many movies. In some he appears with facial hair, or with a hat, or wearing glasses, or some combination of these.

Download the `denzel.R` file and load it directly into `R` using `source("denzel.R")` (you may have to give the full path to the file if this doesn't work, or just open the file in `R` and execute its contents).

This will provide data on twenty-one movies in which Mr. Washington has appeared and what sort of choice of accessory (i.e. hat, facial hair, glasses) he was sporting in each movie. The data is given in two forms, one as a data frame called `denzel` whose `row.names(denzel)` will give the movie titles. Its three variates are binary, each indicating whether Mr. Washington sported one or more of the three possibilities of facial hair, glasses, or a hat. The other contains the same information (absent the movie titles) but as a table of counts called `denzelTable`.

a. **Venn diagrams**. Install the package `venneuler` from the CRAN repository. (You will need `rJava` and hence `Java` installed. If this doesn't work, use the windows machines run by MFCF on the third floor of the MC building.)

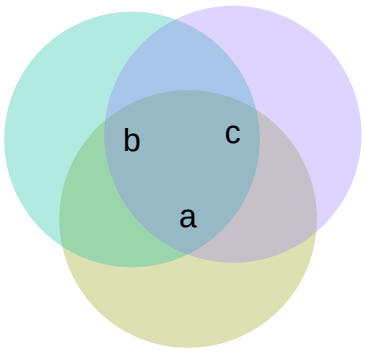
   i. **(4 marks)** Using `venneuler(...)`, draw a Venn diagram for the three sets "facial hair", "glasses", and "hat". **Hint:** you will need to determine an appropriate vector of `movies` and a matching vector of `choices` so that you can create an appropriate data frame `venndenzel` and produce a plot as

In constructing `movies` and `choices` it might be best to construct three separate pairs of vectors first, one for each choice and then concatenate them together. For example,

```
facialhair <- denzel[,"facialhair"] == "yes"
facialhairMovies <- row.names(denzel)[facialhair]
facialhairChoices <- rep("facial hair", length(facialhairMovies))

hat <- denzel[,"hat"] == "yes"
hatMovies <- row.names(denzel)[hat]
hatChoices <- rep("hat", length(hatMovies))

glasses <- denzel[,"glasses"] == "yes"
glassesMovies <- row.names(denzel)[glasses]
glassesChoices <- rep("glasses", length(glassesMovies))
# et cetera
movies <- c(facialhairMovies, hatMovies, glassesMovies)
choices <- c(facialhairChoices, hatChoices, glassesChoices)
ven = venneuler(data.frame(elements = movies, sets = choices))
plot(ven)
```

ii. **(2 marks)** What can you conclude from this diagram about the relationship between these choices?

```
# a uniform distributed 3 (independent) variable venn diagram looks like this:
M=10000
data  = 1:M
a = runif(M)>0.5
as <- data[a]
aname = rep("a", length(as))
b = runif(M)>0.5
bs <- data[b]
bname = rep("b", length(bs))
c = runif(M)>0.5
cs <- data[c]
cname = rep("c", length(cs))

counts = c(as,bs,cs)
names = c(aname,bname,cname)
par(mfrow=c(1,2))
plot( venneuler(data.frame(elements = counts, sets = names)))
plot( venneuler(data.frame(elements = movies, sets = choices)))
```



we conclude that Denzel seems to use facial hair most, and use glasses least,

the combination Denzel uses the most is facial hair + hat

iii. **(2 marks)** By looking at the residuals of the output from `venneuler(...)`, which set of choices is best represented by area in this diagram? Which is worst?

```
ven$residuals
```

```
##                 glasses                     hat             hat&glasses
##                3.907887                2.512107                -4.575347
##             facial hair     facial hair&glasses         facial hair&hat
##                1.461748                -4.775874                -3.613238
## facial hair&hat&glasses
##                3.988013
```

facial hair is best represented by area, facial hair & glasses is worst

b. **Barplots** Now using the data contained in `denzelTable`,

i. **(3 marks)** Construct and print each of the marginal tables: `facialhair` × `hat`, `facialhair` × `glasses`, and `hat` × `glasses`.

```
savepar = par(mar=c(5.1, 4.1, 4.1, 8.1), xpd=TRUE)

fht=margin.table(denzelTable, margin = c(1,3))
fgt=margin.table(denzelTable, margin = c(2,3))
hgt=margin.table(denzelTable, margin = c(2,1))
fht
```

```
##      facialhair
## hat    no yes
##   no    5   9
```

```
##   yes  7   7
```

```
fgt
```

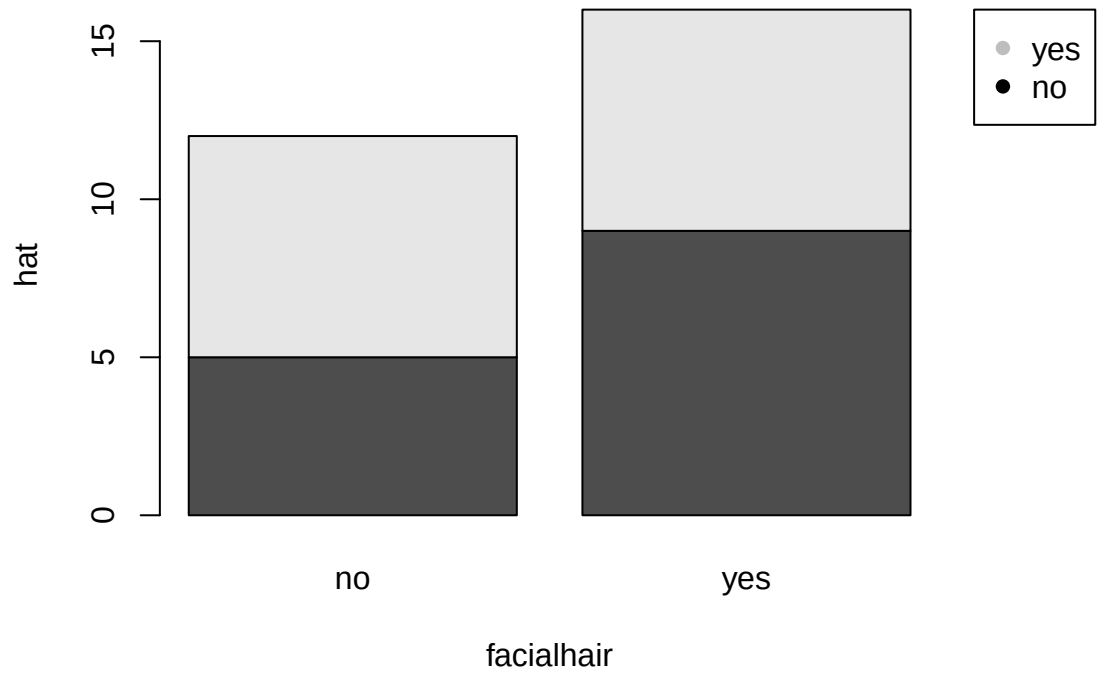
```
##        facialhair
## glasses no yes
##     no   6  10
##     yes  6   6
```

```
hgt
```

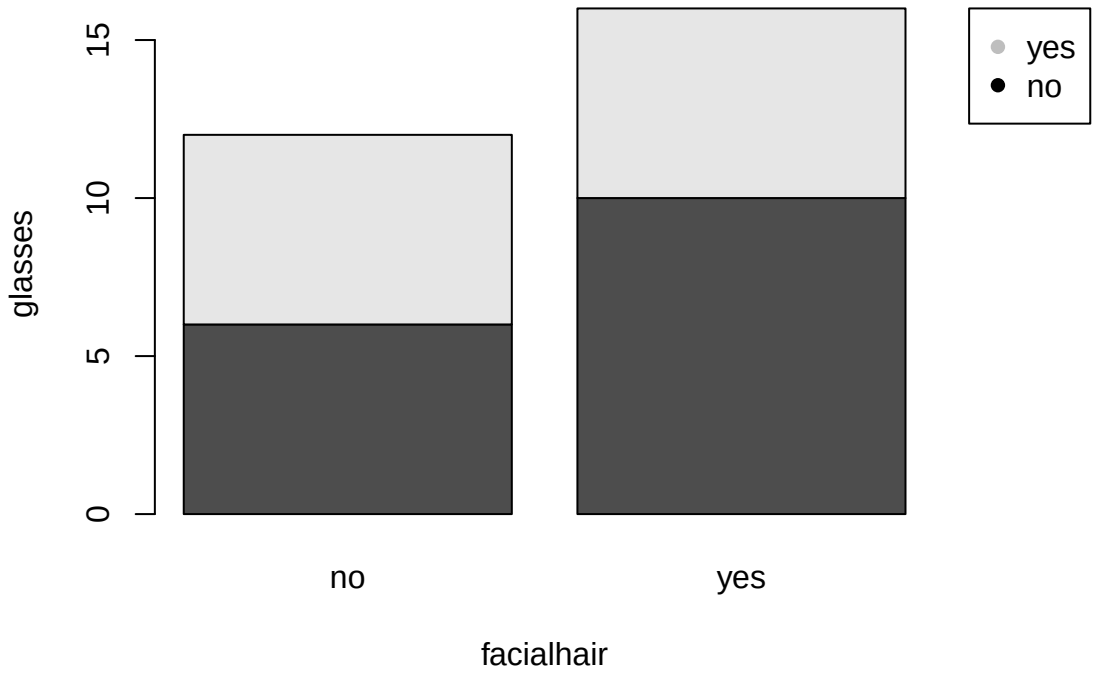
```
##        hat
## glasses no yes
##     no   8   8
##     yes  6   6
```

ii. **(3 marks)** Produce a barplot from each of the above tables showing `facialhair` by `hat`, `facialhair` by `glasses`, and `hat` by `glasses`, respectively. Have a legend attached to each.

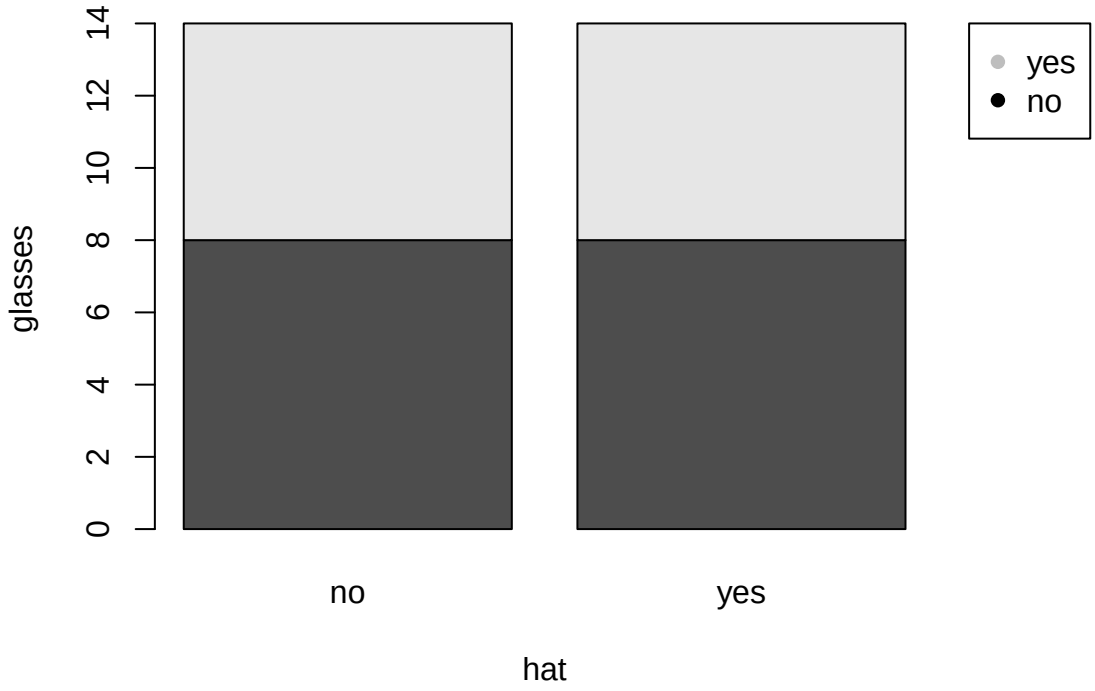
```
par(mar=c(5.1, 4.1, 4.1, 8.1), xpd=TRUE)
barplot(fht, xlab = "facialhair", ylab="hat")
legend("topright", legend = c("yes","no"), col = c("grey", "black"), pch=16, inset = c(-0.2,0))
```




```
par(mar=c(5.1, 4.1, 4.1, 8.1), xpd=TRUE)
barplot(fgt, xlab = "facialhair", ylab="glasses")
legend("topright", legend = c("yes","no"), col = c("grey", "black"), pch=16, inset = c(-0.2,0))
```




```
par(mar=c(5.1, 4.1, 4.1, 8.1), xpd=TRUE)
barplot(hgt, xlab = "hat", ylab="glasses")
legend("topright", legend = c("yes","no"), col = c("grey", "black"), pch=16, inset = c(-0.2,0))
```



iii. **(3 marks)** What conclusions can you draw about the relationship between the choices from these barplot displays? How do these displays compare to the Venn diagrams above in the information they provide?

choice of glasses seems to be independent with choice of hat.

If with facial hair, Denzel seems to prefer no glasses over with glasses. If without facial hair, choice of glasses seems randomly uniform.

He also seems to prefer facialhair with no hat AND no facial hair with hat over other combination of facialhair and hat

He usually has facial hair in movies. In half of his movies, he wears hat.

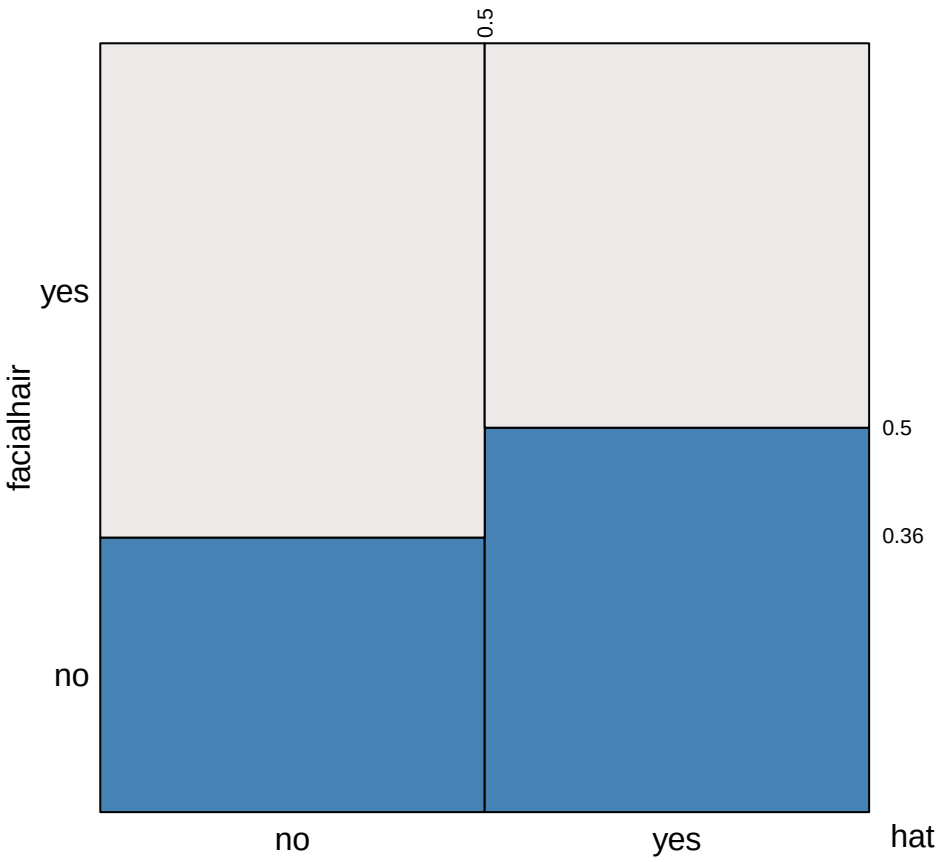
c. **Eikosograms**. You will need to first install the package `eikosograms` (from CRAN). For this question, you will be working with the data frame `denzel`.

i. **(4 marks)** Produce the eikosogram for each the following combinations
- Facial hair conditional only on hat or not.
- Facial hair conditional only on glasses or not.
- Hat conditional on glasses or not.

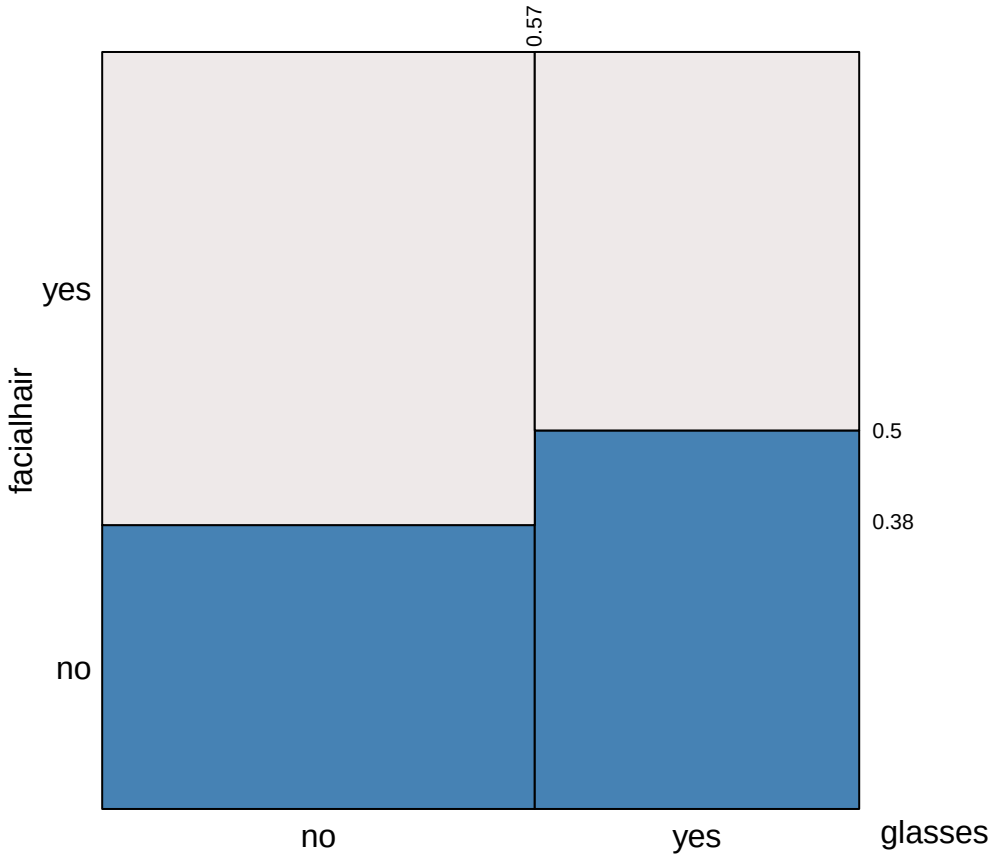
```
library("eikosograms")
```

```
## Warning: package 'eikosograms' was built under R version 4.0.4
```

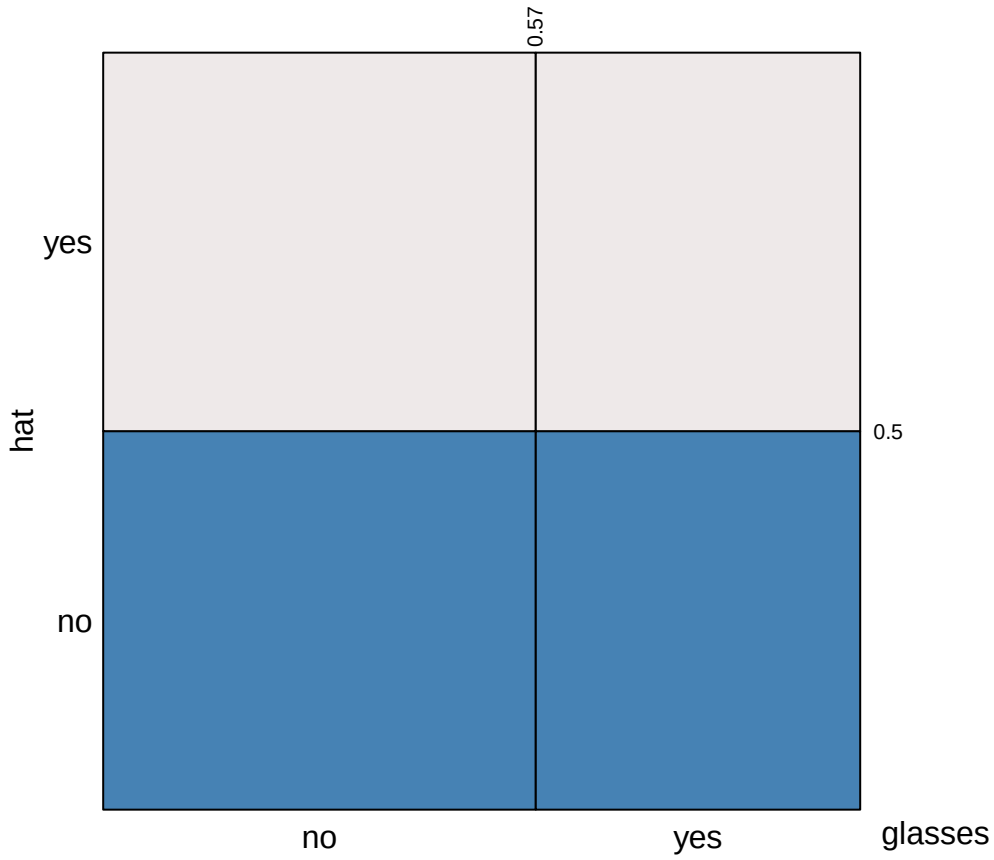
```
eikos(facialhair~hat , data = denzel)
```




```
eikos(facialhair~glasses, data = denzel)
```



```
eikos(hat~glasses, data = denzel)
```



What conclusions do you draw about these various conditional distributions?

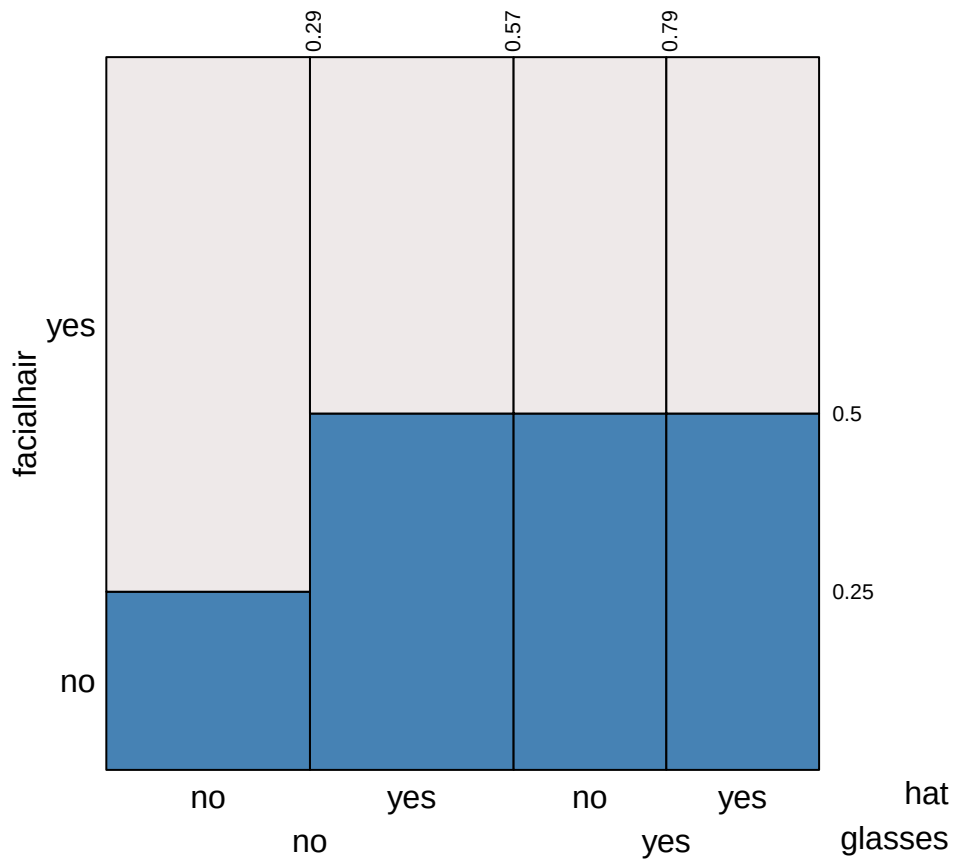
```
choice of glasses is independent with choice of hat.
Denzel choose facialhair more often when he choose no glasses, than when he choose glasses
Denzel choose facialhair more often when he choose no hat, than when he choose hat

Denzel has roughly same number movies with hat and without hat.
He has more movies without glasses than movies with glasses
```

ii. **(6 marks)** Produce the eikosogram for each the following combinations

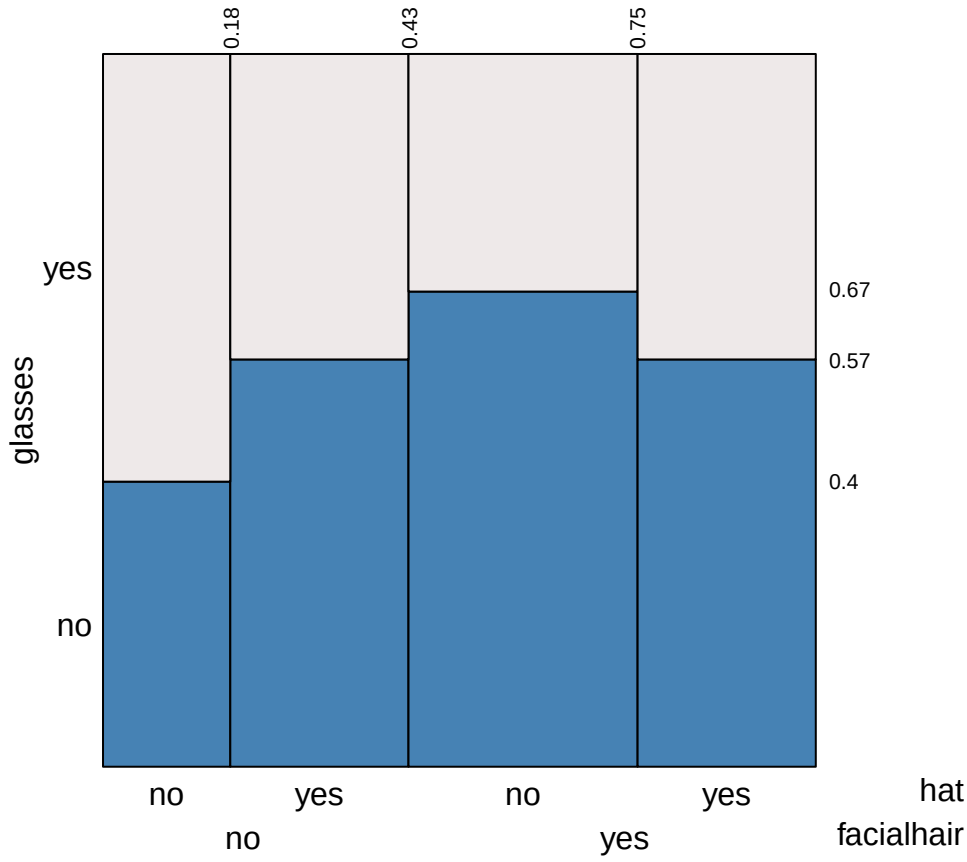
- Facial hair conditional on all possible choices of glasses and hat
- Glasses conditional on all possible choices of facial hair and hat
- Hat conditional on all possible choices of glasses and facial hair

What conclusions do you draw about each of these various conditional distributions?

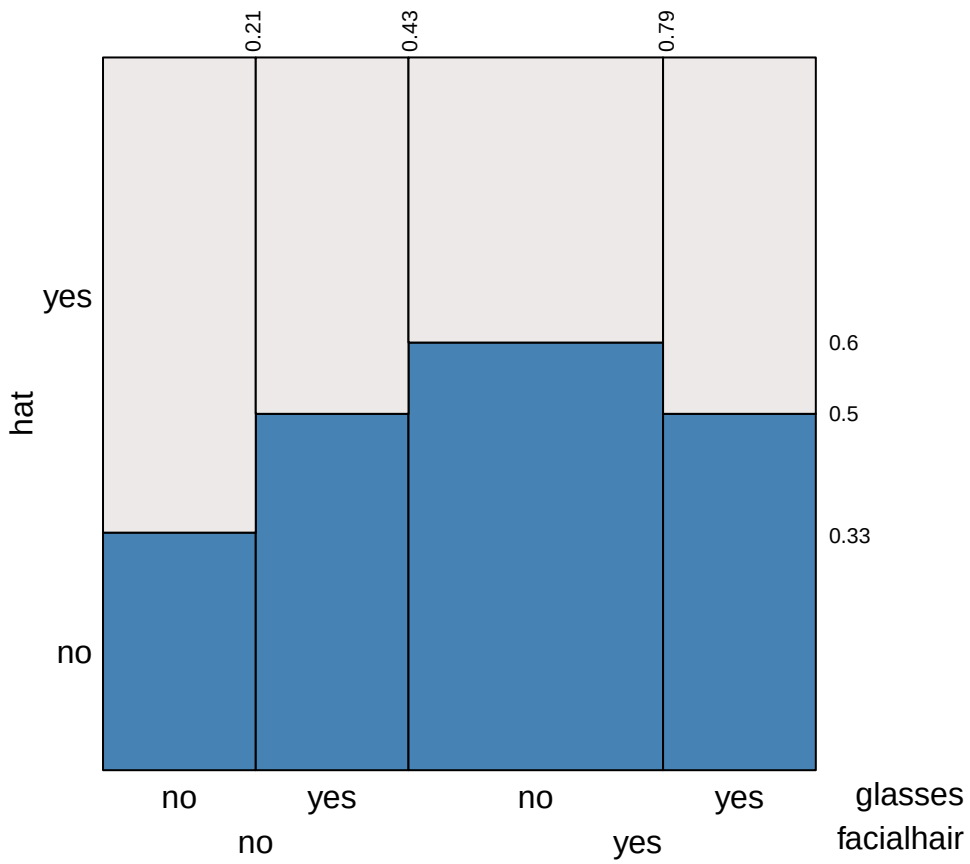
```
eikos(facialhair~. , data = denzel)
```




```
eikos(glasses~., data = denzel)
```



```
eikos(hat~., data = denzel)
```



Denzel clearly prefers to have facial hair when there is no glasses and no hat

Denzel always wear glasses when there is no facial hair and no hat

Denzel always wear hat when there is no glasses and no facialhair.

one variable alone could not explain the target well. (unless when target is facial hair, and we know Denzel has glass, then it does not matter whether he has hat or not.)