Denzel

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)

venndenzel <- data.frame(movies, choices)
venn <- venneuler(venndenzel)
plot(venn)
```

ii. **(2 marks)** What can you conclude from this diagram about the relationship between these choices?
We can see that the intersection between facial hair and hat or facial hair and glasses are large comparing to the individual circle. We can conclude that when danzel has facial hair, it is likely that he's wearing hat or glasses. We cannot conclude anything about all three showing up.

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?

```
venn$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
```

From residuals, we can see that facial hair is best represented and facial hair$glasses is worest represented.

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`.

```
a <- margin.table(denzelTable, margin = c(3,1))
b <- margin.table(denzelTable, margin = c(3,2))
c <- margin.table(denzelTable, margin = c(1,2))

print(a)
```

```
##           hat
## facialhair no yes
##        no   5   7
##        yes  9   7
```

```
print(b)
```

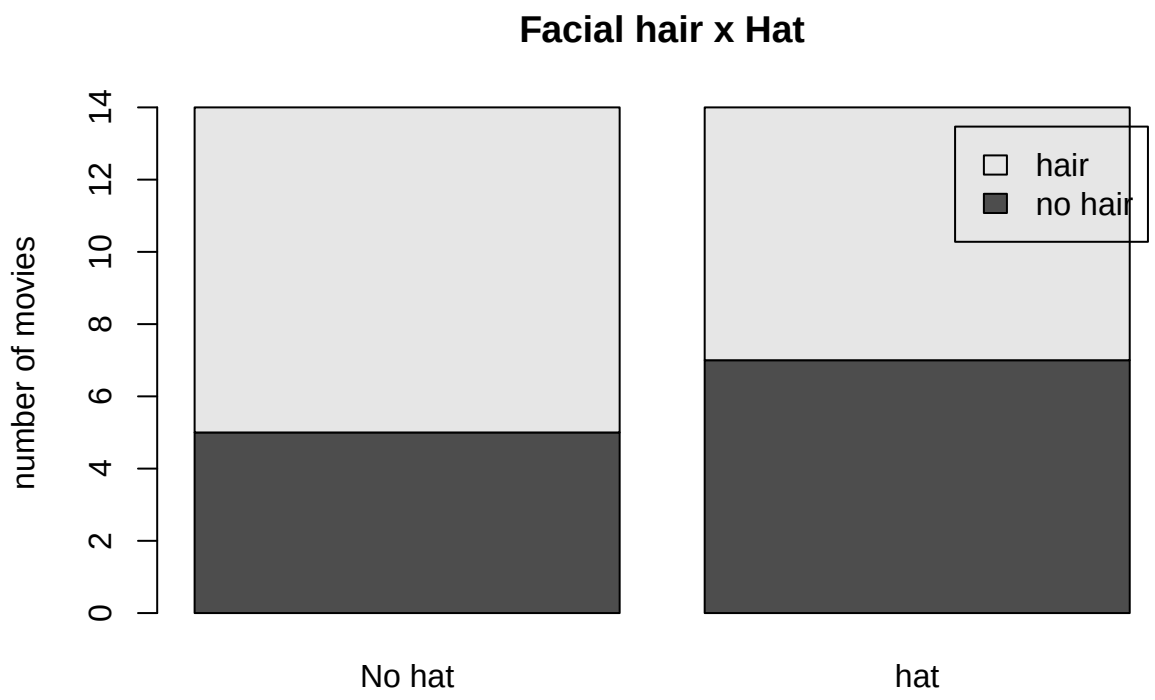
```
##           glasses
## facialhair no yes
##        no   6   6
##        yes 10   6
```

```
print(c)
```

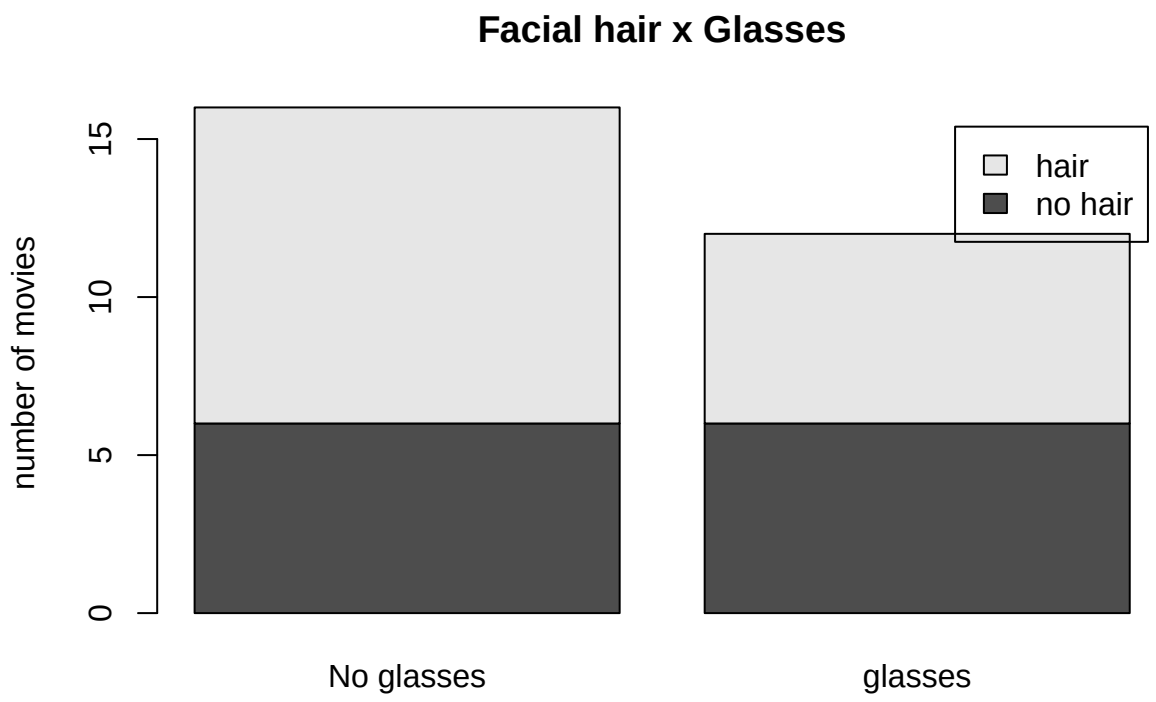
```
##      glasses
## hat   no yes
##   no   8   6
##   yes  8   6
```

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.

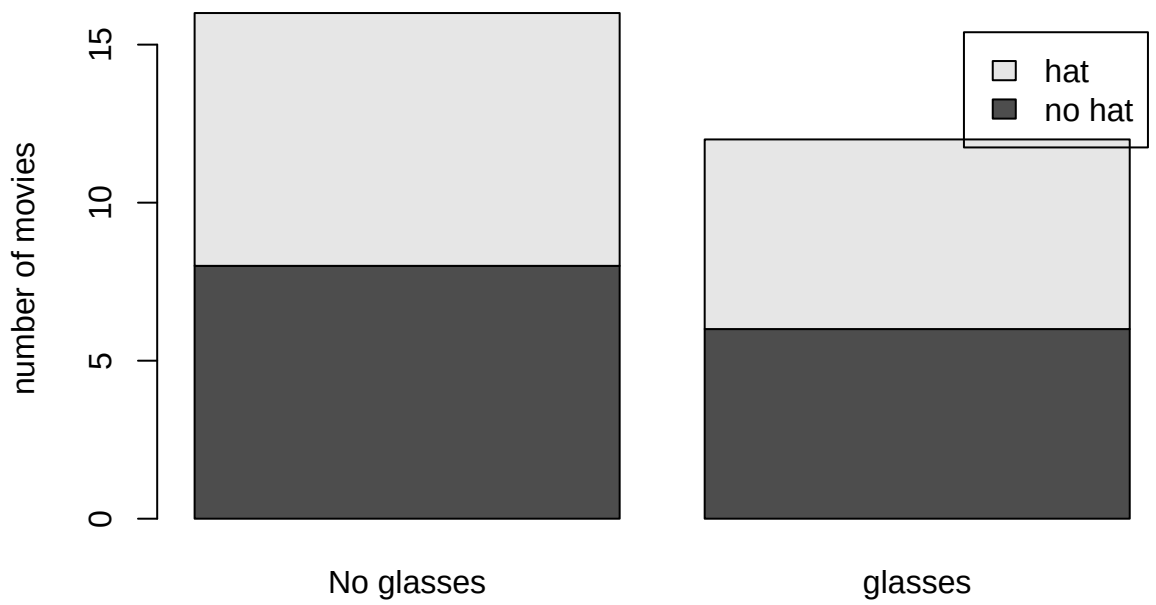
```
barplot(a, names.arg=c('No hat', 'hat'), ylab = 'number of movies',
        legend=c('no hair', 'hair'), main='Facial hair x Hat')
```




```
barplot(b, names.arg=c('No glasses', 'glasses'), ylab = 'number of movies',
        legend=c('no hair', 'hair'), main='Facial hair x Glasses')
```




```
barplot(c, names.arg=c('No glasses', 'glasses'), ylab = 'number of movies',
        legend=c('no hat', 'hat'), main='Hat x Glasses')
```



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?

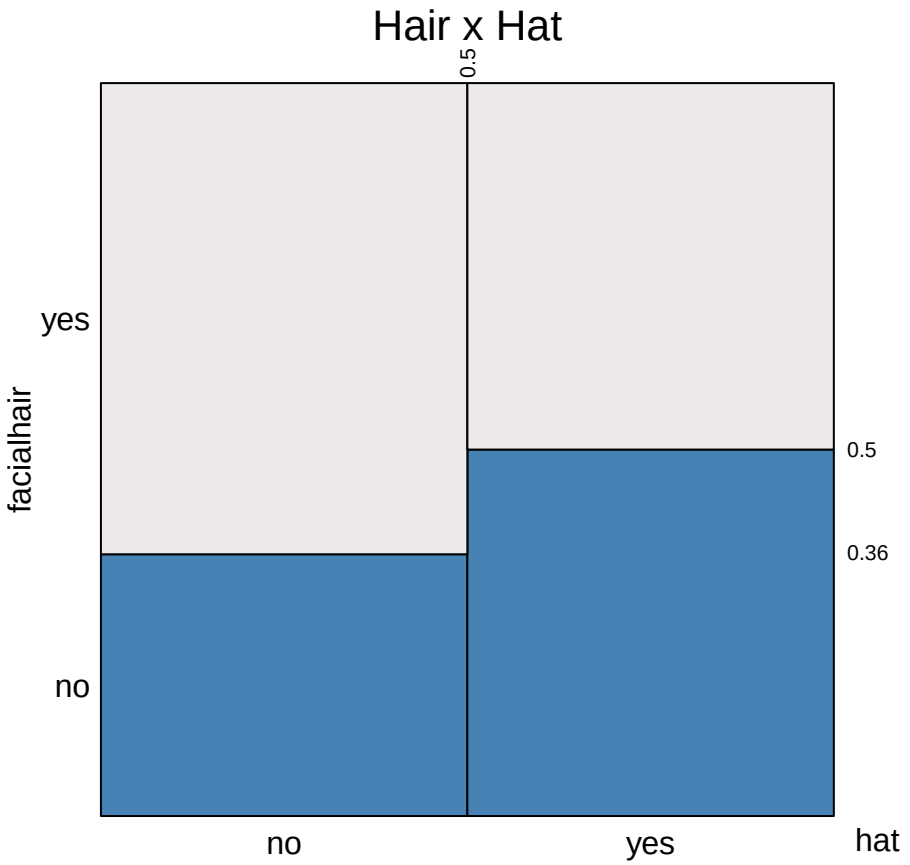
From first two graphs, Denzel has a higher probability of having facial hair when he does not wear hat or glasses comparing to wearing hat or glasses. From the last graph, we can see that probability of wearing hat for both cases are the same. This means that there is no direct relationship between hat and glasses.

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`.

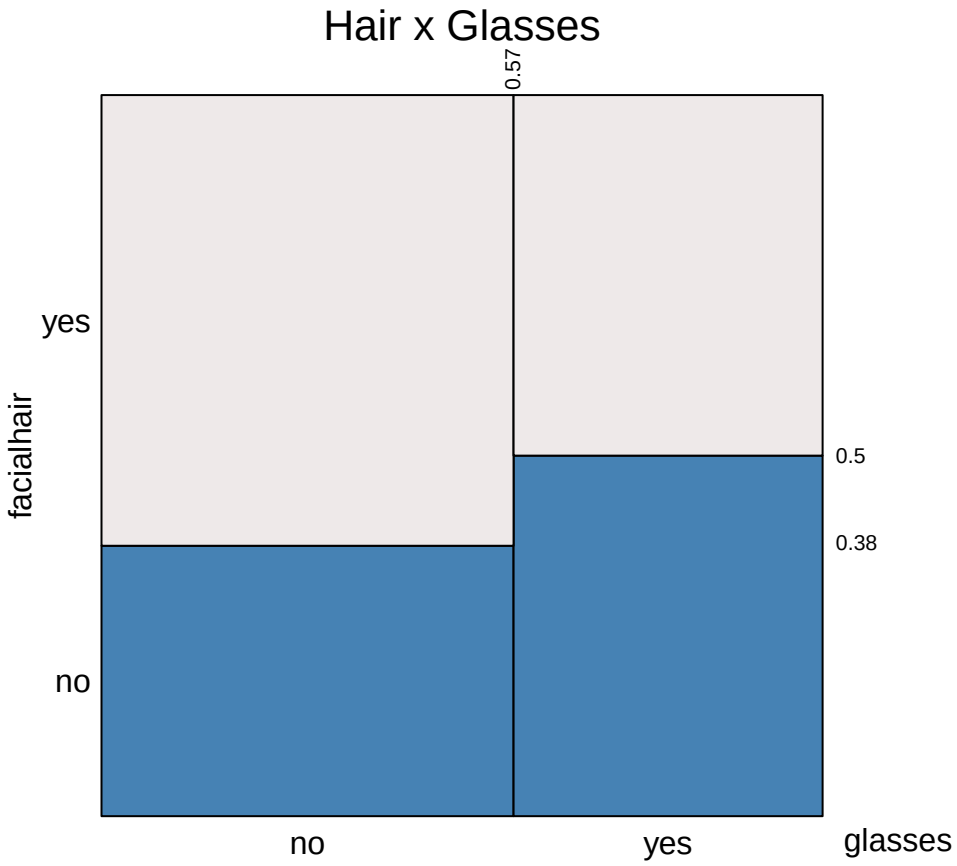
i. **(4 marks)** Produce the eikosogram for each the following combinations

- Facial hair conditional only on hat or not.
- Facial hair conditional only on glasses or not.
- Hat conditional on glasses or not.

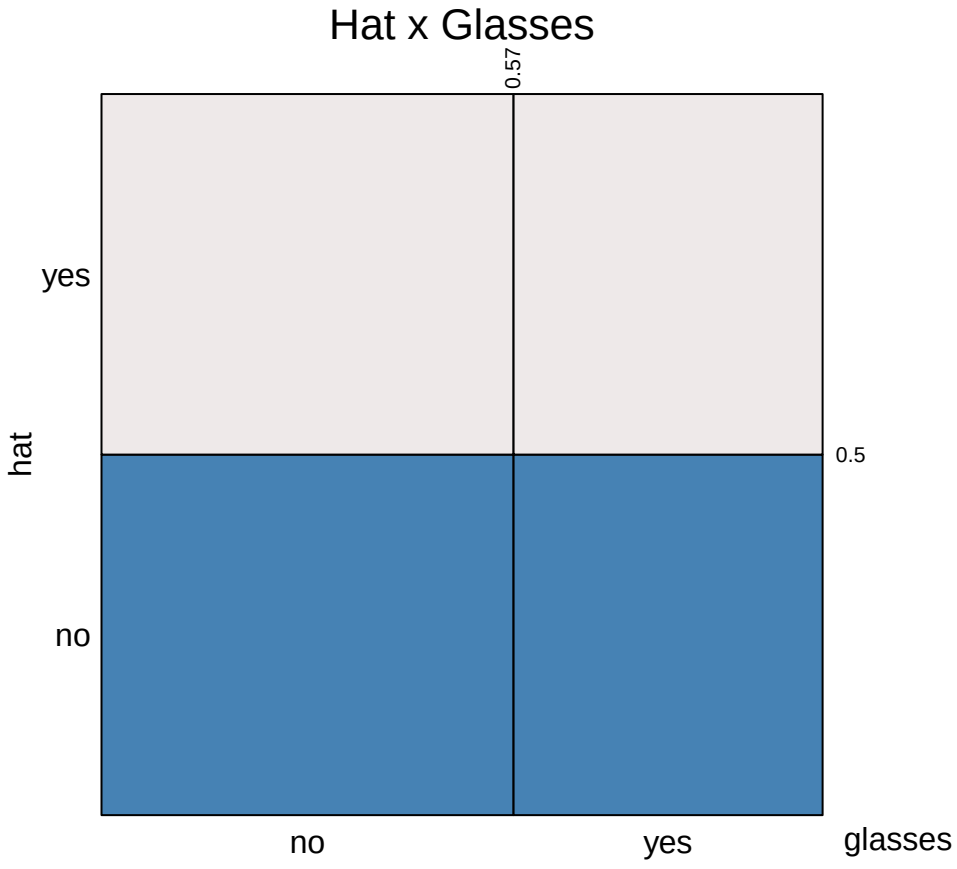
```
library(eikosograms)
eikos(facialhair~hat, data=denzel, main = 'Hair x Hat')
```




```
eikos(facialhair~glasses, data=denzel, main = 'Hair x Glasses')
```




```
eikos(hat~glasses, data=denzel, main = 'Hat x Glasses')
```



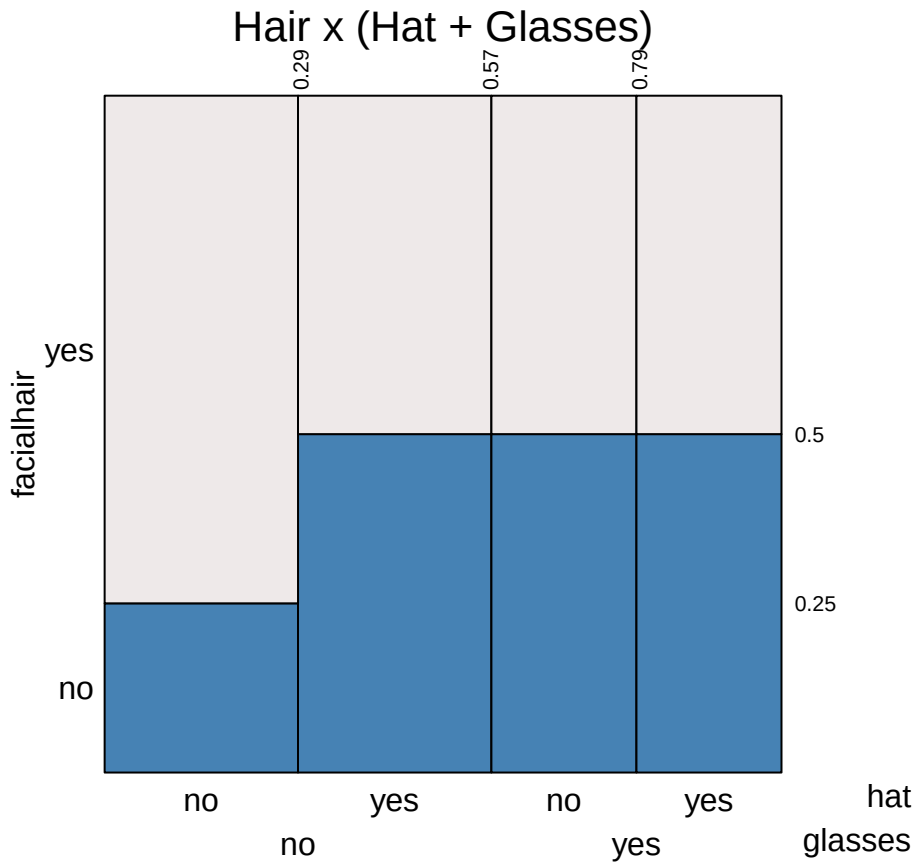
What conclusions do you draw about these various conditional distributions?

We can conclude that there is a nagative relationship between facial hair and hat or facial hair and glasses. Since the last graph is flat, this means that hat and glasses are indenpendent.

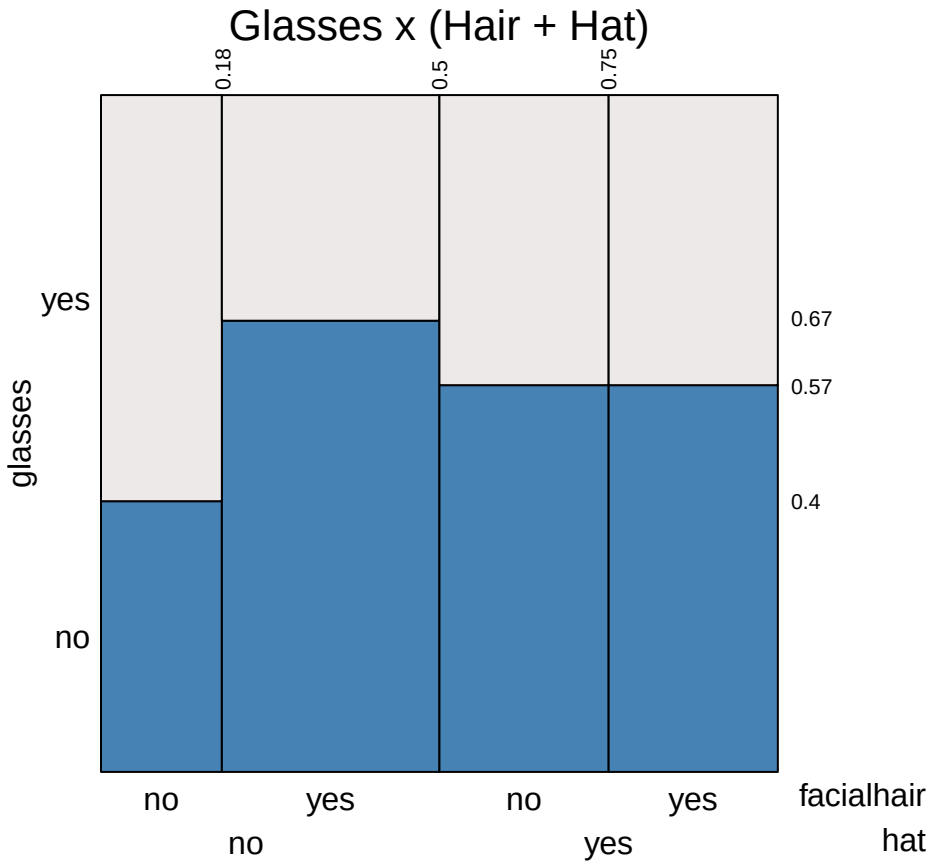
ii. **(6 marks)** Produce the eikosogram for each the following combinations

- 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

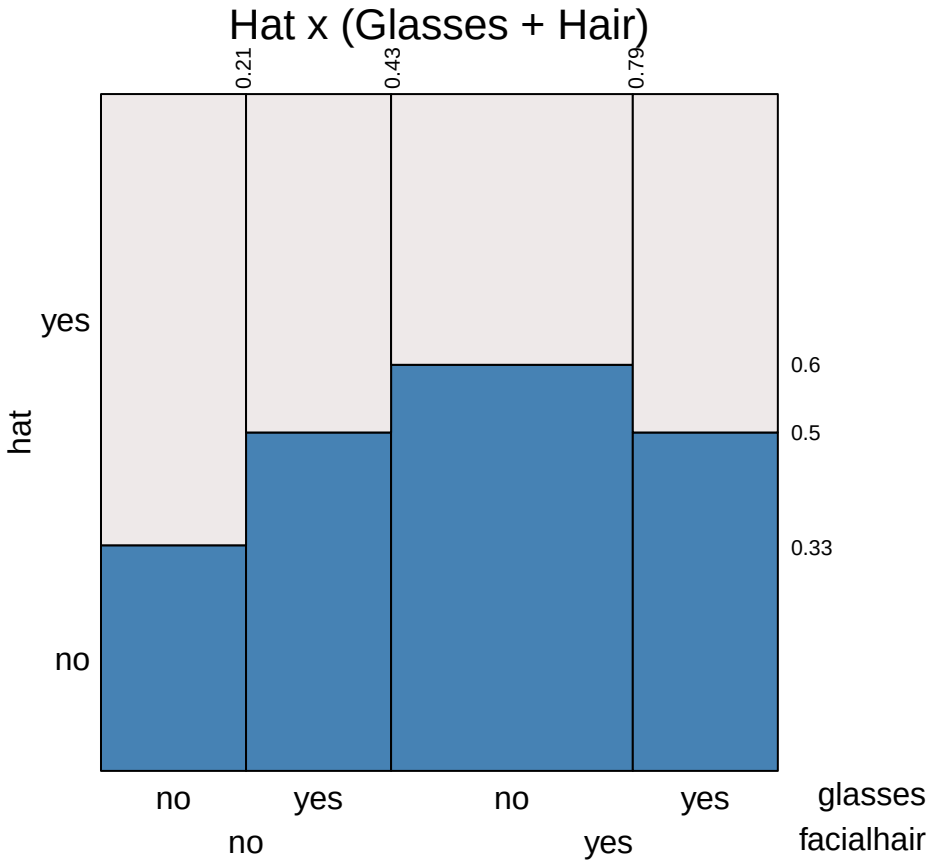
```
eikos(facialhair~hat+glasses, data=denzel, main = 'Hair x (Hat + Glasses)')
```




```
eikos(glasses~facialhair+hat, data=denzel, main = 'Glasses x (Hair + Hat)')
```




```
eikos(hat~glasses+facialhair, data=denzel, main = 'Hat x (Glasses + Hair)')
```



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

For first graph, hat is conditionally dependent on hair when glasses=no, otherwise conditionally independent. For second graph, hair is conditionally dependent on glasses when hat=no, otherwise independent. For third graph, hat is conditionally depedent on both glasses and hair