

MSJemutai_PS_1

Mercy Salome Jemutai

9/4/2019

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

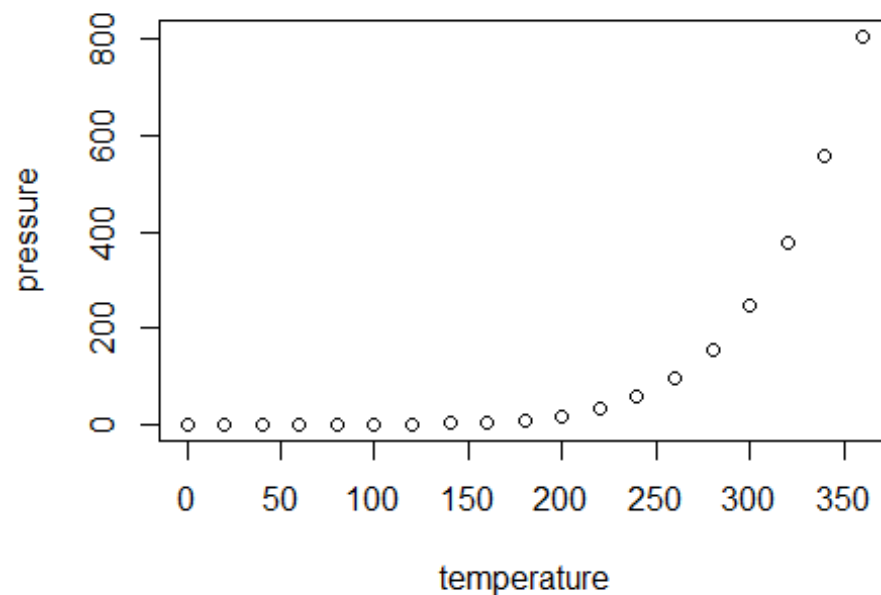
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   :  2.00
## 1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##   Mean  :15.4    Mean   : 42.98
## 3rd Qu.:19.0    3rd Qu.: 56.00
##   Max.  :25.0    Max.    :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

##PROBLEM SET 1, HW 1

##QUESTION 1 (A)

#Load data from the graduate school at Berkely

UCBAdmissions

, , Dept = A

##

Gender

Admit Male Female

Admitted 512 89

Rejected 313 19

##

, , Dept = B

##

Gender

Admit Male Female

Admitted 353 17

Rejected 207 8

##

, , Dept = C

##

Gender

```

## Admit      Male Female
##   Admitted  120    202
##   Rejected  205    391
##
## , , Dept = D
##
##           Gender
## Admit      Male Female
##   Admitted  138    131
##   Rejected  279    244
##
## , , Dept = E
##
##           Gender
## Admit      Male Female
##   Admitted   53     94
##   Rejected  138    299
##
## , , Dept = F
##
##           Gender
## Admit      Male Female
##   Admitted   22     24
##   Rejected  351    317

#confirm the data set is categorical, Table data set
#We cannot attach (works for lists, data frames, and environments only)
class(UCBAdmissions)

## [1] "table"

#Checking dimensions in the data set and how they classified
#The applicants are classified by gender, admit, and department
#There data forms a 3-way table!

str(UCBAdmissions)

## 'table' num [1:2, 1:2, 1:6] 512 313 89 19 353 207 17 8 120 205 ...
## - attr(*, "dimnames")=List of 3
## ..$ Admit : chr [1:2] "Admitted" "Rejected"
## ..$ Gender: chr [1:2] "Male" "Female"
## ..$ Dept  : chr [1:6] "A" "B" "C" "D" ...

# how many males versus females are in the set (use margin.table()function)
#2, is for Gender!
#https://rstudio-pubs-
static.s3.amazonaws.com/308591_6dfff566d383946c881cb5b6a735a79fd.html
margin.table(UCBAdmissions, 2)

```

```
## Gender
##   Male Female
##   2691   1835

# how many of them were admitted overall
#Admit - 1
#Also shows those rejected in total
margin.table(UCBAdmissions, 1)

## Admit
## Admitted Rejected
##      1755      2771

#how many were admitted to each of the departments
#Also shows total of those rejected by each department
margin.table(UCBAdmissions, c(1,3))

##           Dept
## Admit      A   B   C   D   E   F
## Admitted 601 370 322 269 147  46
## Rejected 332 215 596 523 437 668

#Let's find proportion of men admitted in each department vs women

round(prop.table(UCBAdmissions,1),2)

## , , Dept = A
##
##           Gender
## Admit      Male Female
## Admitted 0.29   0.05
## Rejected 0.11   0.01
##
## , , Dept = B
##
##           Gender
## Admit      Male Female
## Admitted 0.20   0.01
## Rejected 0.07   0.00
##
## , , Dept = C
##
##           Gender
## Admit      Male Female
## Admitted 0.07   0.12
## Rejected 0.07   0.14
##
## , , Dept = D
##
##           Gender
## Admit      Male Female
```

```
##   Admitted 0.08  0.07
##   Rejected 0.10  0.09
##
## , , Dept = E
##
##           Gender
## Admit      Male Female
##   Admitted 0.03  0.05
##   Rejected 0.05  0.11
##
## , , Dept = F
##
##           Gender
## Admit      Male Female
##   Admitted 0.01  0.01
##   Rejected 0.13  0.11
```

There seems to be a balance between number of departments that admitted more male and those that did not.

#Let's use a frequency table

```
fable(round(prop.table(UCBAdmissions, c(2,3)), 2),
      row.vars="Dept", col.vars = c("Gender", "Admit"))
```

```
##           Gender      Male      Female
##           Admit Admitted Rejected Admitted Rejected
## Dept
## A              0.62    0.38    0.82    0.18
## B              0.63    0.37    0.68    0.32
## C              0.37    0.63    0.34    0.66
## D              0.33    0.67    0.35    0.65
## E              0.28    0.72    0.24    0.76
## F              0.06    0.94    0.07    0.93
```

#Only 2 departments seem to have high proportion of men admitted than women

#Let's compare proportion of men and women applying to each department against admissions

```
round(prop.table(margin.table(UCBAdmissions,c(2,3)),1),2)
```

```
##           Dept
## Gender      A      B      C      D      E      F
##   Male  0.31 0.21 0.12 0.15 0.07 0.14
##   Female 0.06 0.01 0.32 0.20 0.21 0.19
```

#<http://users.stat.ufl.edu/~athienit/STA6505/ucbadmissions.pdf>

```
round(prop.table(margin.table(UCBAdmissions,c(1,3)),2),2)
```

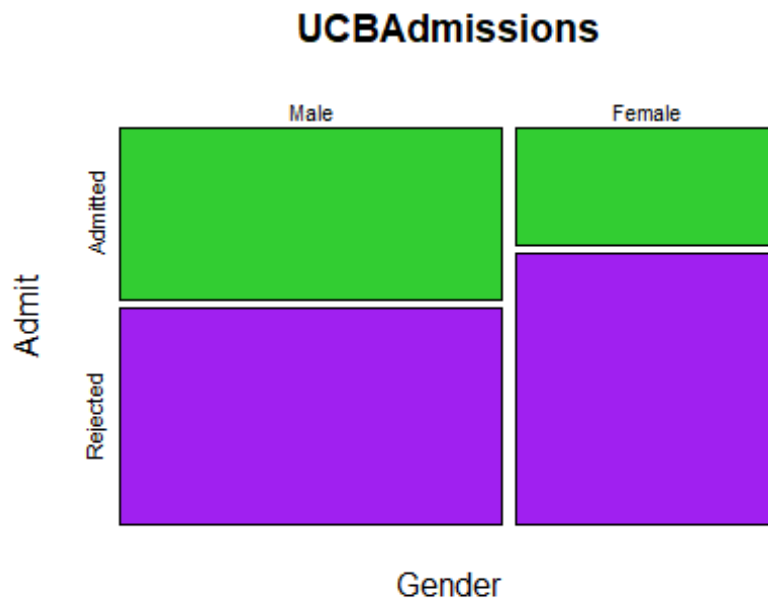
```
##           Dept
## Admit      A      B      C      D      E      F
##   Admitted 0.64 0.63 0.35 0.34 0.25 0.06
##   Rejected 0.36 0.37 0.65 0.66 0.75 0.94
```

#From the two, we can see that women seem to tend to apply to departments with low admission rates

##QUESTION 1(B)

#Mosaic plot for gender against admission

```
mosaicplot(~Gender+Admit, data=UCBAdmissions, col=c("Lime Green","Purple"))
```



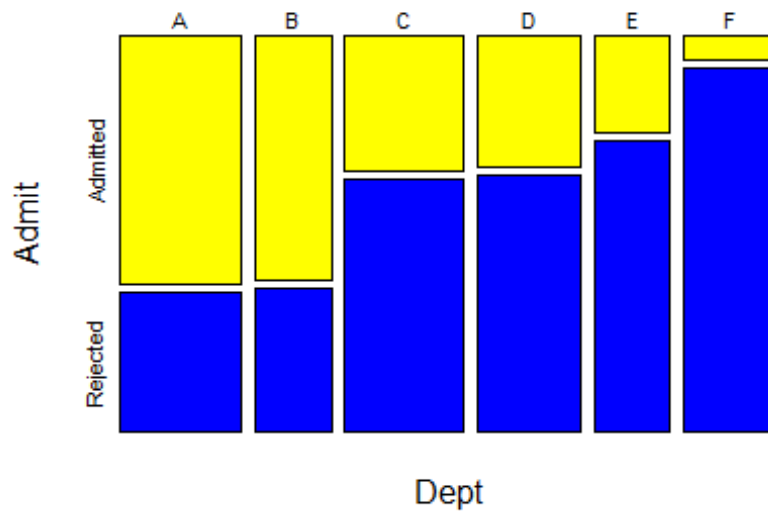
*#From the plot, among the admitted applicants, most are male!
#The plot does not seem to clearly show which gender forms the highest group among those rejected though****

#Let's plot another mosaic plot

#chance of being accepted vs the department?

```
mosaicplot(apply(UCBAdmissions, c(3,1), sum), col=c("Yellow","Blue"))
```

```
apply(UCBAdmissions, c(3, 1), sum)
```

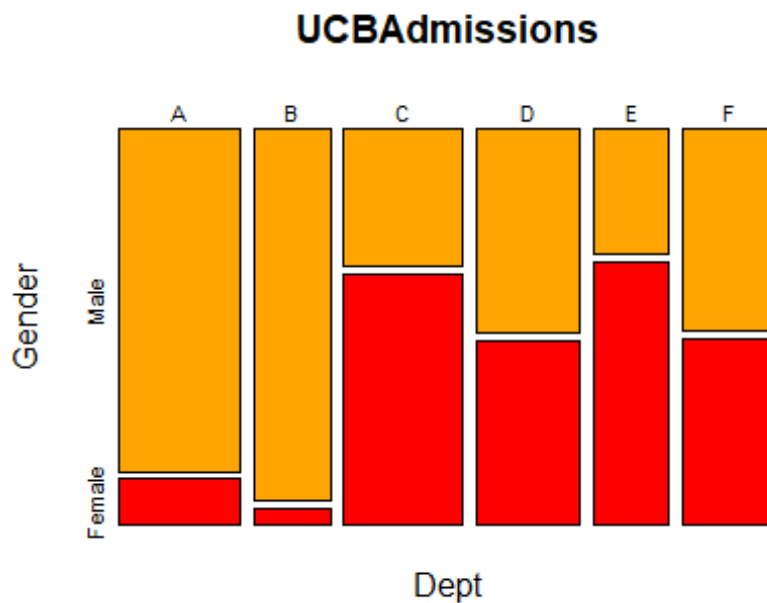


#Compare with the next plot

#Let's check gender against admission

```
#mosaicplot(apply(Titanic, c(3,2)), col=c("Pink", "Gold"))
```

```
mosaicplot(~Dept+Gender, data=UCBAdmissions, col=c("Orange", "Red"))
```



##From the last 2 plots, we can see that less women apply to departments that admit more people!

##QUESTION 2

```
#Download the file, read it into RStudio and plot either the anterior or
#Read in the file
hippocampus<-read.csv("C:\\Users\\jsang\\Downloads\\hippocampus_218P.csv")

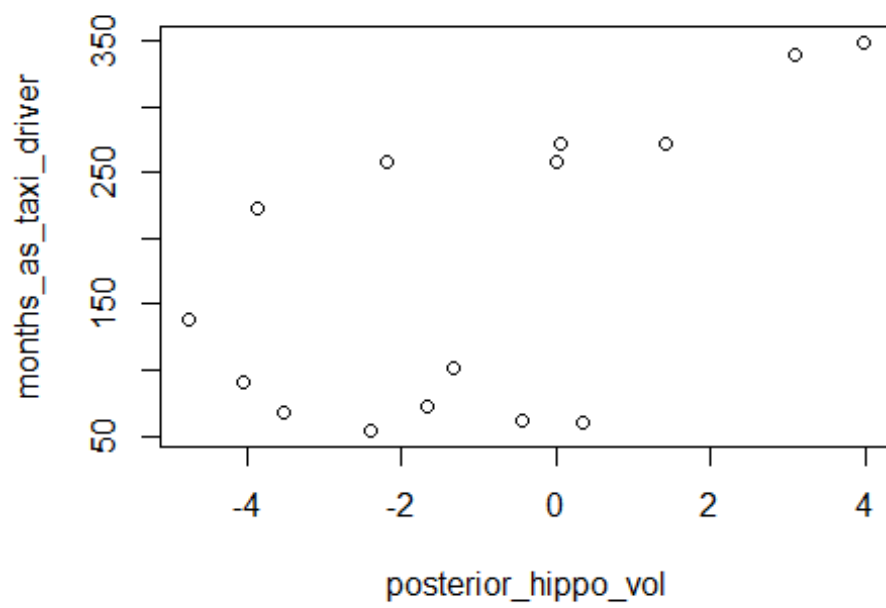
#Let's check names of columns
names(hippocampus)

## [1] "months_as_taxi_driver"      "posterior_hippo_vol"
## [3] "anterior_hippocampus_volume_mm3"

#attach it so you can call each column by name
attach(hippocampus)

#Plot the posterior hippocampus column against the months as a taxi driver
column in a scatterplot

plot(posterior_hippo_vol,months_as_taxi_driver)
```

```
#Detach!!  
detach(hippocampus)
```