Endsem

Shubhang Periwal 19201104

11/28/2019

```
setwd("D:/rprog")
library(ggplot2)
library(rio)
library(tidyverse)
```

```
## -- Attaching packages ----------------------------------------------------------------------------
```

```
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
## v purrr   0.3.3
```

```
## -- Conflicts ------------------------------------------------------------------------------------- tidy
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(plotly)
```

```
##
## Attaching package: 'plotly'
```

```
## The following object is masked from 'package:rio':
##
##     export
```

```
## The following object is masked from 'package:ggplot2':
##
##     last_plot
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
## The following object is masked from 'package:graphics':
##
##     layout
```

```
library(gganimate)
library(inline)
library(pixmap)
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:plotly':
##
##     select

## The following object is masked from 'package:dplyr':
##
##     select
```

```
library(dplyr)
```

R Markdown

question 1

```
ds1 <- import("HR021.csv",setclass = "tibble",header=TRUE)
ds2 = import("s50_1997.txt")
ds3 = import("s50_1995.txt")
ds2
```

```
##    alcohol drugs smoke sport
## 1        3     1     1     1
## 2        2     3     3     1
## 3        3     1     1     1
## 4        2     1     1     1
## 5        4     3     1     2
## 6        4     1     3     2
## 7        3     2     3     2
## 8        4     3     3     2
## 9        2     1     1     1
## 10       4     1     2     2
## 11       5     2     1     1
## 12       5     3     3     1
## 13       2     3     1     1
## 14       3     1     3     1
## 15       5     3     3     1
## 16       4     2     3     1
## 17       4     2     2     1
## 18       3     2     1     1
## 19       5     3     3     1
## 20       3     1     1     2
## 21       3     1     1     2
## 22       3     1     1     2
## 23       2     3     3     1
## 24       3     2     1     2
## 25       4     1     1     1
## 26       3     3     3     1
## 27       3     1     1     1
## 28       4     1     1     2
## 29       3     1     1     2
## 30       4     3     3     1
## 31       4     1     1     2
## 32       4     1     1     1
```

```
## 33       3     1     3     2
## 34       2     1     1     1
## 35       4     3     1     1
## 36       4     3     3     1
## 37       3     1     1     2
## 38       4     3     1     1
## 39       3     1     1     2
## 40       1     1     1     2
## 41       4     3     3     1
## 42       5     3     3     1
## 43       4     2     1     2
## 44       5     3     1     1
## 45       2     1     1     2
## 46       2     1     1     1
## 47       2     1     1     1
## 48       4     1     1     2
## 49       3     1     1     1
## 50       3     3     3     1
```

```
ds3
```

```
##    alcohol drugs smoke sport
## 1        3     1     2     2
## 2        2     2     3     1
## 3        2     1     1     1
## 4        2     1     1     2
## 5        3     1     1     2
## 6        4     1     1     2
## 7        4     3     1     1
## 8        4     3     3     2
## 9        2     1     1     2
## 10       4     1     1     2
## 11       5     2     3     2
## 12       5     3     3     2
## 13       3     3     1     1
## 14       3     1     1     1
## 15       4     1     2     2
## 16       4     2     2     2
## 17       2     1     1     1
## 18       4     1     1     1
## 19       3     1     1     2
## 20       2     1     1     2
## 21       1     1     1     2
## 22       3     1     1     1
## 23       4     4     3     2
## 24       3     1     1     2
## 25       3     1     1     2
## 26       4     3     3     2
## 27       2     1     1     2
## 28       2     1     1     2
## 29       3     2     1     2
## 30       1     1     1     2
## 31       4     1     1     2
## 32       4     1     1     2
```

```
## 33       3     1     1     2
## 34       2     1     1     2
## 35       3     2     1     2
## 36       4     1     1     2
## 37       2     1     1     2
## 38       3     1     1     2
## 39       2     1     1     2
## 40       1     1     1     2
## 41       4     1     2     2
## 42       4     3     3     1
## 43       2     1     1     2
## 44       5     3     2     1
## 45       2     1     1     2
## 46       2     1     1     1
## 47       2     1     1     1
## 48       2     1     1     2
## 49       1     1     1     1
## 50       1     2     1     2
```

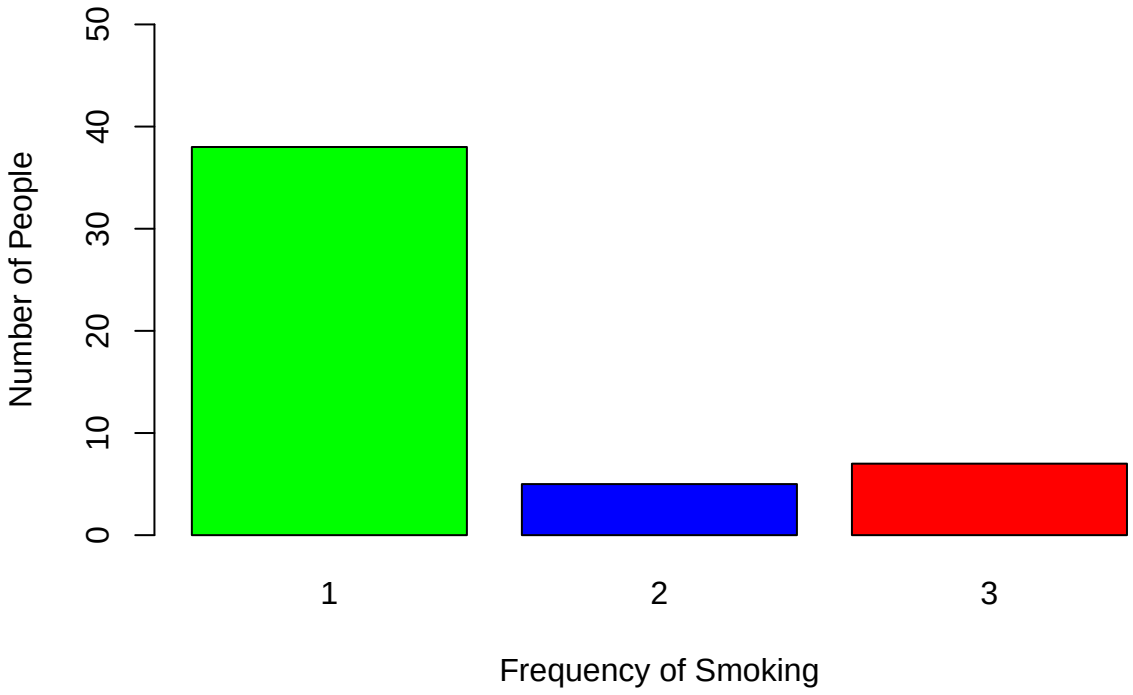
Question 2

```
#a
ds3$alcohol = factor(ds3$alcohol)
ds3$drugs = factor(ds3$drugs)
ds3$smoke = factor(ds3$smoke)
ds3$sport = factor(ds3$sport)
str(ds3)
```

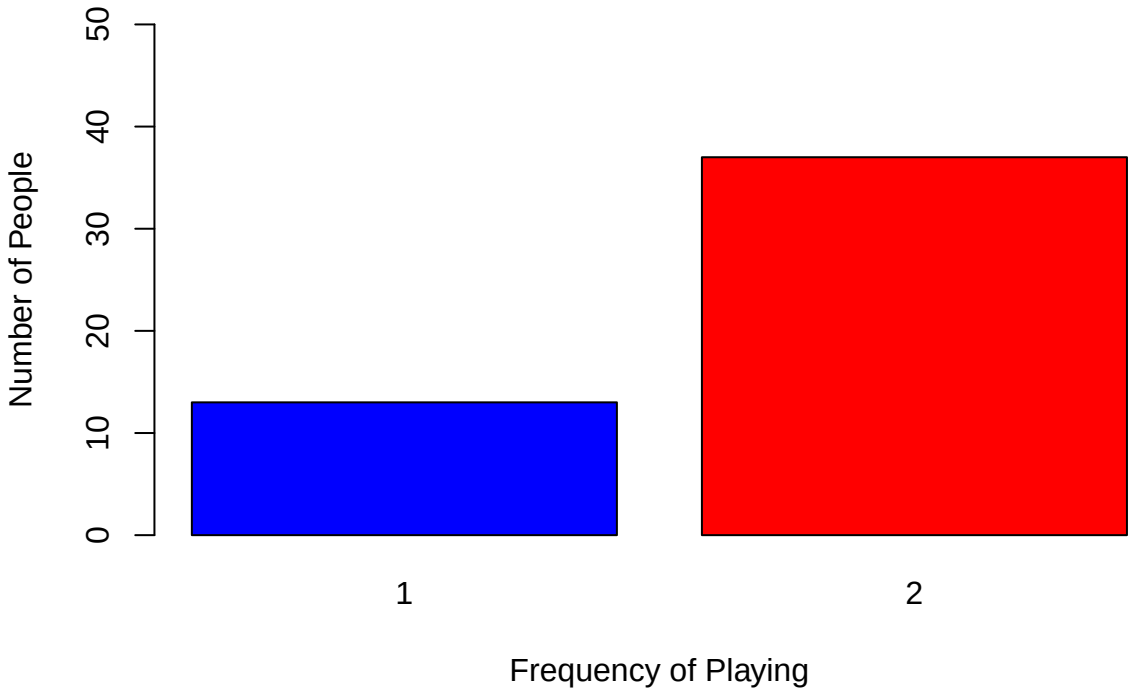
```
## 'data.frame':    50 obs. of  4 variables:
##  $ alcohol: Factor w/ 5 levels "1","2","3","4",..: 3 2 2 2 3 4 4 4 2 4 ...
##  $ drugs  : Factor w/ 4 levels "1","2","3","4": 1 2 1 1 1 1 3 3 1 1 ...
##  $ smoke  : Factor w/ 3 levels "1","2","3": 2 3 1 1 1 1 1 3 1 1 ...
##  $ sport  : Factor w/ 2 levels "1","2": 2 1 1 2 2 2 1 2 2 2 ...
```

```
#Q2 b
#,legend("topright",legend=c("Not","Occasional","Regular"),lty=c(1,1,1),lwd = c(3,3,3),col=c("green","b

t1 = table(ds3$smoke)
t2 = table(ds3$sport)
x = barplot(t1,xlab="Frequency of Smoking",ylab="Number of People",ylim = c(0,50),col=c("green","blue",
```




```
y =  barplot(t2,xlab="Frequency of Playing",ylab="Number of People",ylim = c(0,50),col=c("blue","red"))
```




```
par(mfrow=c(1,2))
x
```

```
##      [,1]
## [1,]  0.7
## [2,]  1.9
## [3,]  3.1
```

```
y
```

```
##      [,1]
## [1,]  0.7
## [2,]  1.9
```

The resulting plots show that there are high number of non smokers and low number of actual smokers, which is a good thing. The number of students who play regularly are almost 3 times the number of people who do not play. This shows that there maybe a positive correlation between students who play and students who don't smoke.

```
#c

y2 = length(ds3$smoke)
y1 = length(which(ds3$smoke=="3")) + length(which(ds3$smoke=="2"))
y1/y2
```

```
## [1] 0.24
```

```
y3 = length(which(ds3$smoke=="3" & ds3$sport=="2")) + length(which(ds3$smoke=="2" & ds3$sport=="2"))
y3/y2
```

```
## [1] 0.18
```

The proportion of pupils who smoke atleast occassionaly is 0.24 The proportion of pupils who smoke and play is 0.18

```
#d
class(ds3) <- c("s50survey")
summary.s50survey <- function(ds3){
  l1 = length(ds3$smoke)
  l2 = length(ds3$drugs)
  l3 = length(ds3$alcohol)
  l4 = length(ds3$sport)

  hsm = length(which(ds3$smoke=="3"))/l1
  lsm = length(which(ds3$smoke=="2"))/l1
  nsm = length(which(ds3$smoke=="1"))/l1

    hd = length(which(ds3$drugs=="4"))/l2
  ld = length(which(ds3$drugs=="2"))/l2
  nd = length(which(ds3$drugs=="1"))/l2
  md = length(which(ds3$drugs=="3"))/l2

  al1 = length(which(ds3$alcohol=="1"))/l3
  al2 = length(which(ds3$alcohol=="2"))/l3
  al3= length(which(ds3$alcohol=="3"))/l3
  al4 = length(which(ds3$alcohol=="4"))/l3
  al5 = length(which(ds3$alcohol=="5"))/l3

  sp1 = length(which(ds3$sport=="1"))/l4
  sp2 = length(which(ds3$sport=="2"))/l4

  cat("Smokers by Frequency\n ")
  cat("High Smokers: ")
  cat(hsm)
  cat("\n")
    cat("Low Smokers: ")
  cat(lsm)
  cat("\n")
    cat("Non Smokers: ")
  cat(nsm)
  cat("\n")

    cat("Frequency by Drug Consumption\n ")
  cat("High Users: ")
  cat(hd)
  cat("\n")
    cat("Occasional: ")
  cat(md)
```

```
  cat("\n")
    cat("Tried Once: ")
  cat(ld)
  cat("\n")
     cat("Non Users: ")
  cat(nd)
  cat("\n")

      cat("by Alcohol Consumption\n ")
  cat("High Users: ")
  cat(al5)
  cat("\n")
    cat("Occasional: ")
  cat(al4)
  cat("\n")
    cat("once a month: ")
  cat(al3)
  cat("\n")
     cat("once/twice a year: ")
  cat(al2)
  cat("\n")
  cat("Non Drinkers: ")
  cat(al1)
  cat("\n")

    cat("Sports by Frequency\n ")
  cat("High : ")
  cat(sp2)
  cat("\n")
    cat("Low : ")
  cat(sp1)
  cat("\n")
}
summary(ds3)
```

```
## Smokers by Frequency
##  High Smokers: 0.14
## Low Smokers: 0.1
## Non Smokers: 0.76
## Frequency by Drug Consumption
##  High Users: 0.02
## Occasional: 0.14
## Tried Once: 0.12
## Non Users: 0.72
## by Alcohol Consumption
##  High Users: 0.06
## Occasional: 0.28
## once a month: 0.24
## once/twice a year: 0.32
## Non Drinkers: 0.1
## Sports by Frequency
##  High : 0.74
## Low : 0.26
```

```
#(.72) who did not use cannabis or 72%
```

The proportion of students who do not use cannabis is 0.72 #e

```
ds2$alcohol = factor(ds2$alcohol)
ds2$drugs = factor(ds2$drugs)
ds2$smoke = factor(ds2$smoke)
ds2$sport = factor(ds2$sport)
str(ds2)
```

```
## 'data.frame':    50 obs. of  4 variables:
##  $ alcohol: Factor w/ 5 levels "1","2","3","4",..: 3 2 3 2 4 4 3 4 2 4 ...
##  $ drugs  : Factor w/ 3 levels "1","2","3": 1 3 1 1 3 1 2 3 1 1 ...
##  $ smoke  : Factor w/ 3 levels "1","2","3": 1 3 1 1 1 3 3 3 1 2 ...
##  $ sport  : Factor w/ 2 levels "1","2": 1 1 1 1 2 2 2 2 1 2 ...
```

```
class(ds2) <- c("s50survey")
summary(ds2)
```

```
## Smokers by Frequency
##  High Smokers: 0.34
## Low Smokers: 0.04
## Non Smokers: 0.62
## Frequency by Drug Consumption
##  High Users: 0
## Occasional: 0.34
## Tried Once: 0.14
## Non Users: 0.52
## by Alcohol Consumption
##  High Users: 0.12
## Occasional: 0.34
## once a month: 0.34
## once/twice a year: 0.18
## Non Drinkers: 0.02
## Sports by Frequency
##  High : 0.38
## Low : 0.62
```

```
temp = 50*(.74) - 50*(.38)
temp
```

```
## [1] 18
```

People who played sports have decreased from 37 to 18

#q3 a

```
ds1 <- import("HR021.csv",setclass = "tibble",header=TRUE)
colnames(ds1)
```

```
## [1] "Sex"        "Admissions" "CHO"        "2015"       "2016"
## [6] "2017"
```

```
#colnames(ds1) = c("Sex","Admissions","CHO","`2015`","`2016`","`2017`")

str(ds1)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    40 obs. of  6 variables:
##  $ Sex       : chr  "Male" "Male" "Male" "Male" ...
##  $ Admissions: chr  "First admissions" "First admissions" "First admissions" "First admissions" ...
##  $ CHO       : chr  "CHO 1" "CHO 2" "CHO 3" "CHO 4" ...
##  $ 2015      : int  225 371 222 505 407 254 476 358 450 25 ...
##  $ 2016      : int  246 323 202 441 356 185 455 440 452 35 ...
##  $ 2017      : int  249 292 226 447 392 187 467 502 437 21 ...
```

#q3 b

```
library(dplyr)
tem = filter(ds1,ds1$Admissions=="All admissions",Sex == "Male")
tef = filter(ds1,ds1$Admissions=="All admissions",Sex == "Female")
t1 = sum(tem$`2017`)
t2 = sum(tef$`2017`)
t1
```

```
## [1] 8522
```

```
t2
```

```
## [1] 8478
```

The number of males : 8522 The number of females : 8478

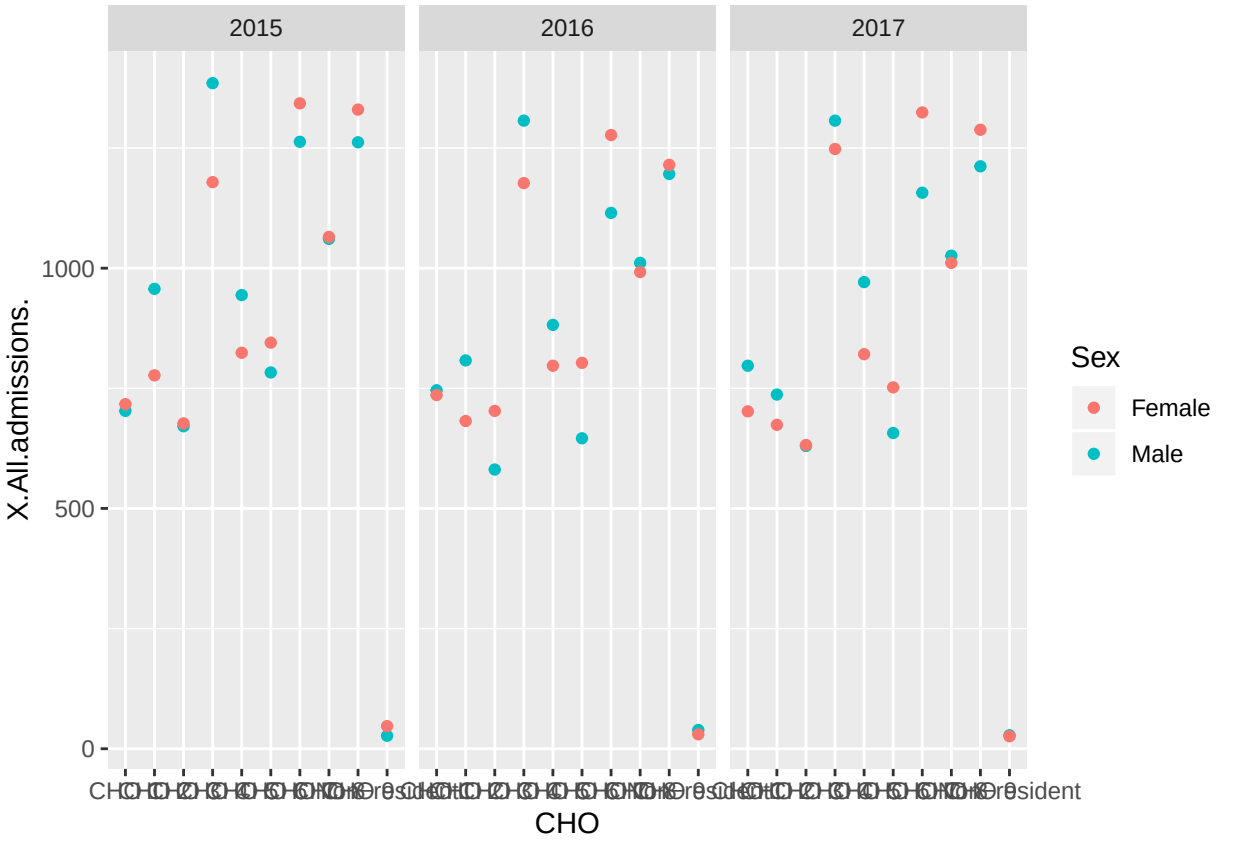
#q3 c

```
ds1new = filter(ds1,ds1$Admissions=="All admissions")
ds1new <- pivot_longer(ds1new, cols = `2015`:`2017`,
names_to = "Year", values_to = "`All admissions`")
dc = c('Admissions')
ds1new = ds1new[-2]

str(ds1new)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    60 obs. of  4 variables:
##  $ Sex             : chr  "Male" "Male" "Male" "Male" ...
##  $ CHO             : chr  "CHO 1" "CHO 1" "CHO 1" "CHO 2" ...
##  $ Year            : chr  "2015" "2016" "2017" "2015" ...
##  $ `All admissions`: int  703 746 797 957 808 737 671 581 630 1385 ...
```

```
ds1new = data.frame(ds1new)
ggplot(ds1new,aes(x=CHO,y=X.All.admissions.,color=Sex))+geom_point() + facet_wrap(~Year)
```


2015
2016
2017
X.All.admissions.
1000
500
0
CHO
Sex
Female
Male