cda 第一次作业

林振炜

2019 年 3 月 19 日

- 第一题

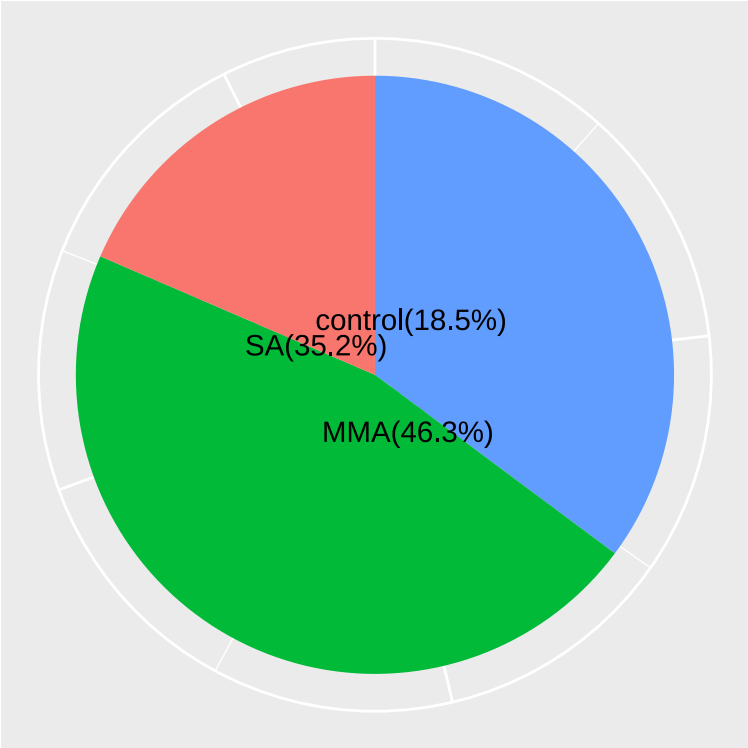
```
library(dplyr)
description <- read.csv('GDS5037/description.csv')
##calulate the frequency and percentage of each category and plot pie chart
s <- description %>%group_by(ID)%>%summarise(Counts=n())
#if"%>%" errors,maybe the reason is the function can't be found
library(ggplot2)
type <- s$ID
nums <- s$Counts
df <- data.frame(type = type, nums = nums)


#2.2 pie_chart
p <- ggplot(data = df, mapping = aes(x = 'Content', y = nums, fill = type)) + geom_bar(sta


label_value <- paste('(', round(df$nums/sum(df$nums) * 100, 1), '%)', sep = '')
label_value
```

```
## [1] "(18.5%)" "(46.3%)" "(35.2%)"
```

```
label <- paste(df$type, label_value, sep = '')
label
```

```
## [1] "control(18.5%)" "MMA(46.3%)"     "SA(35.2%)"
```

```
p+coord_polar(theta = 'y') + labs(x = '', y = '', title = '') + theme(axis.text = element
```




```
#2.3 barchart about gender
male_sample <-  'GSM1068462,GSM1068463,GSM1068465,GSM1068466,GSM1068467,
GSM1068469,GSM1068470,GSM1068471,GSM1068475,GSM1068480,GSM1068484,
GSM1068485,GSM1068489,GSM1068497,GSM1068501,GSM1068504,GSM1068509,
GSM1068511,GSM1068515,GSM1068516,GSM1068519,GSM1068523,GSM1068525,
GSM1068526,GSM1068529,GSM1068530,GSM1068534,GSM1068536,GSM1068541,
GSM1068553,GSM1068554,GSM1068558,GSM1068559,GSM1068564'
female_sample = 'GSM1068458,GSM1068459,GSM1068460,GSM1068461,GSM1068464,
GSM1068468,GSM1068472,GSM1068473,GSM1068474,GSM1068476,GSM1068477,
GSM1068478,GSM1068479,GSM1068481,GSM1068482,GSM1068483,GSM1068486,
GSM1068487,GSM1068488,GSM1068490,GSM1068491,GSM1068492,GSM1068493,
GSM1068494,GSM1068495,GSM1068496,GSM1068498,GSM1068499,GSM1068500,
GSM1068502,GSM1068503,GSM1068505,GSM1068506,GSM1068507,GSM1068508,
GSM1068510,GSM1068512,GSM1068513,GSM1068514,GSM1068517,GSM1068518,
GSM1068520,GSM1068521,GSM1068522,GSM1068524,GSM1068527,GSM1068528,
GSM1068531,GSM1068532,GSM1068533,GSM1068535,GSM1068537,GSM1068538,
GSM1068539,GSM1068540,GSM1068542,GSM1068543,GSM1068544,GSM1068545,
```

```
GSM1068546,GSM1068547,GSM1068548,GSM1068549,GSM1068550,GSM1068551,
GSM1068552,GSM1068555,GSM1068556,GSM1068557,GSM1068560,GSM1068561,
GSM1068562,GSM1068563,GSM1068565'
male_sample <- strsplit(male_sample,',')[[1]]
male_sample <- gsub("\n","",male_sample)
female_sample <- strsplit(female_sample,',')[[1]]
female_sample <- gsub("\n","",female_sample)
countmale <- length(male_sample)
countfemale <- length(female_sample)
gender <- c(rep('male',countmale),rep('female',countfemale))
ID_REF <- c(male_sample,female_sample)
gendata <- data.frame(ID_REF,gender)
genderdata <- merge(description,gendata,by = 'ID_REF')
su <- genderdata %>%group_by(ID,gender)%>%summarise(Counts=n())
genderdata
```

```
##         ID_REF
## 1   GSM1068458
## 2   GSM1068459
## 3   GSM1068460
## 4   GSM1068461
## 5   GSM1068462
## 6   GSM1068463
## 7   GSM1068464
## 8   GSM1068465
## 9   GSM1068466
## 10  GSM1068467
## 11  GSM1068468
## 12  GSM1068469
## 13  GSM1068470
## 14  GSM1068471
## 15  GSM1068472
## 16  GSM1068473
## 17  GSM1068474
## 18  GSM1068475
## 19  GSM1068476
## 20  GSM1068477
```

```
## 21  GSM1068478
## 22  GSM1068479
## 23  GSM1068480
## 24  GSM1068481
## 25  GSM1068482
## 26  GSM1068483
## 27  GSM1068484
## 28  GSM1068485
## 29  GSM1068486
## 30  GSM1068487
## 31  GSM1068488
## 32  GSM1068489
## 33  GSM1068490
## 34  GSM1068491
## 35  GSM1068492
## 36  GSM1068493
## 37  GSM1068494
## 38  GSM1068495
## 39  GSM1068496
## 40  GSM1068497
## 41  GSM1068498
## 42  GSM1068499
## 43  GSM1068500
## 44  GSM1068501
## 45  GSM1068502
## 46  GSM1068503
## 47  GSM1068504
## 48  GSM1068505
## 49  GSM1068506
## 50  GSM1068507
## 51  GSM1068508
## 52  GSM1068509
## 53  GSM1068510
## 54  GSM1068511
## 55  GSM1068512
## 56  GSM1068513
## 57  GSM1068514
```

```
## 58  GSM1068515
## 59  GSM1068516
## 60  GSM1068517
## 61  GSM1068518
## 62  GSM1068519
## 63  GSM1068520
## 64  GSM1068521
## 65  GSM1068522
## 66  GSM1068523
## 67  GSM1068524
## 68  GSM1068525
## 69  GSM1068526
## 70  GSM1068527
## 71  GSM1068528
## 72  GSM1068529
## 73  GSM1068530
## 74  GSM1068531
## 75  GSM1068532
## 76  GSM1068533
## 77  GSM1068534
## 78  GSM1068535
## 79  GSM1068536
## 80  GSM1068537
## 81  GSM1068538
## 82  GSM1068539
## 83  GSM1068540
## 84  GSM1068541
## 85  GSM1068542
## 86  GSM1068543
## 87  GSM1068544
## 88  GSM1068545
## 89  GSM1068546
## 90  GSM1068547
## 91  GSM1068548
## 92  GSM1068549
## 93  GSM1068550
## 94  GSM1068551
```

```
## 95  GSM1068552
## 96  GSM1068553
## 97  GSM1068554
## 98  GSM1068555
## 99  GSM1068556
## 100 GSM1068557
## 101 GSM1068558
## 102 GSM1068559
## 103 GSM1068560
## 104 GSM1068561
## 105 GSM1068562
## 106 GSM1068563
## 107 GSM1068564
## 108 GSM1068565
##
## 1   #GSM1068458 = Value for GSM1068458: Bronchial Epithelial Cells control biological 1
## 2   #GSM1068459 = Value for GSM1068459: Bronchial Epithelial Cells control biological 1
## 3   #GSM1068460 = Value for GSM1068460: Bronchial Epithelial Cells control biological 1
## 4   #GSM1068461 = Value for GSM1068461: Bronchial Epithelial Cells control biological 1
## 5   #GSM1068462 = Value for GSM1068462: Bronchial Epithelial Cells control biological 1
## 6   #GSM1068463 = Value for GSM1068463: Bronchial Epithelial Cells control biological 1
## 7   #GSM1068464 = Value for GSM1068464: Bronchial Epithelial Cells control biological 1
## 8   #GSM1068465 = Value for GSM1068465: Bronchial Epithelial Cells control biological 1
## 9   #GSM1068466 = Value for GSM1068466: Bronchial Epithelial Cells control biological 1
## 10  #GSM1068467 = Value for GSM1068467: Bronchial Epithelial Cells control biological 1
## 11  #GSM1068468 = Value for GSM1068468: Bronchial Epithelial Cells control biological 1
## 12  #GSM1068469 = Value for GSM1068469: Bronchial Epithelial Cells control biological 1
## 13  #GSM1068470 = Value for GSM1068470: Bronchial Epithelial Cells control biological 1
## 14  #GSM1068471 = Value for GSM1068471: Bronchial Epithelial Cells control biological 1
## 15  #GSM1068472 = Value for GSM1068472: Bronchial Epithelial Cells control biological 1
## 16  #GSM1068473 = Value for GSM1068473: Bronchial Epithelial Cells control biological 2
## 17  #GSM1068474 = Value for GSM1068474: Bronchial Epithelial Cells control biological 2
## 18  #GSM1068475 = Value for GSM1068475: Bronchial Epithelial Cells control biological 2
## 19  #GSM1068476 = Value for GSM1068476: Bronchial Epithelial Cells control biological 2
## 20  #GSM1068477 = Value for GSM1068477: Bronchial Epithelial Cells control biological 2
## 21      #GSM1068478 = Value for GSM1068478: Bronchial Epithelial Cells MMA biological 0
## 22      #GSM1068479 = Value for GSM1068479: Bronchial Epithelial Cells MMA biological 1
```

```
## 23      #GSM1068480 = Value for GSM1068480: Bronchial Epithelial Cells MMA biological 1
## 24      #GSM1068481 = Value for GSM1068481: Bronchial Epithelial Cells MMA biological 1
## 25      #GSM1068482 = Value for GSM1068482: Bronchial Epithelial Cells MMA biological 1
## 26      #GSM1068483 = Value for GSM1068483: Bronchial Epithelial Cells MMA biological 1
## 27      #GSM1068484 = Value for GSM1068484: Bronchial Epithelial Cells MMA biological 1
## 28      #GSM1068485 = Value for GSM1068485: Bronchial Epithelial Cells MMA biological 1
## 29      #GSM1068486 = Value for GSM1068486: Bronchial Epithelial Cells MMA biological 1
## 30      #GSM1068487 = Value for GSM1068487: Bronchial Epithelial Cells MMA biological 2
## 31      #GSM1068488 = Value for GSM1068488: Bronchial Epithelial Cells MMA biological 2
## 32      #GSM1068489 = Value for GSM1068489: Bronchial Epithelial Cells MMA biological 2
## 33      #GSM1068490 = Value for GSM1068490: Bronchial Epithelial Cells MMA biological 2
## 34      #GSM1068491 = Value for GSM1068491: Bronchial Epithelial Cells MMA biological 2
## 35      #GSM1068492 = Value for GSM1068492: Bronchial Epithelial Cells MMA biological 2
## 36      #GSM1068493 = Value for GSM1068493: Bronchial Epithelial Cells MMA biological 0
## 37      #GSM1068494 = Value for GSM1068494: Bronchial Epithelial Cells MMA biological 0
## 38      #GSM1068495 = Value for GSM1068495: Bronchial Epithelial Cells MMA biological 0
## 39      #GSM1068496 = Value for GSM1068496: Bronchial Epithelial Cells MMA biological 1
## 40      #GSM1068497 = Value for GSM1068497: Bronchial Epithelial Cells MMA biological 1
## 41      #GSM1068498 = Value for GSM1068498: Bronchial Epithelial Cells MMA biological 1
## 42      #GSM1068499 = Value for GSM1068499: Bronchial Epithelial Cells MMA biological 1
## 43      #GSM1068500 = Value for GSM1068500: Bronchial Epithelial Cells MMA biological 1
## 44      #GSM1068501 = Value for GSM1068501: Bronchial Epithelial Cells MMA biological 2
## 45      #GSM1068502 = Value for GSM1068502: Bronchial Epithelial Cells MMA biological 2
## 46      #GSM1068503 = Value for GSM1068503: Bronchial Epithelial Cells MMA biological 2
## 47      #GSM1068504 = Value for GSM1068504: Bronchial Epithelial Cells MMA biological 2
## 48      #GSM1068505 = Value for GSM1068505: Bronchial Epithelial Cells MMA biological 2
## 49      #GSM1068506 = Value for GSM1068506: Bronchial Epithelial Cells MMA biological 2
## 50      #GSM1068507 = Value for GSM1068507: Bronchial Epithelial Cells MMA biological 2
## 51      #GSM1068508 = Value for GSM1068508: Bronchial Epithelial Cells MMA biological 2
## 52      #GSM1068509 = Value for GSM1068509: Bronchial Epithelial Cells MMA biological 1
## 53      #GSM1068510 = Value for GSM1068510: Bronchial Epithelial Cells MMA biological 2
## 54      #GSM1068511 = Value for GSM1068511: Bronchial Epithelial Cells MMA biological 2
## 55      #GSM1068512 = Value for GSM1068512: Bronchial Epithelial Cells MMA biological 2
## 56      #GSM1068513 = Value for GSM1068513: Bronchial Epithelial Cells MMA biological 2
## 57      #GSM1068514 = Value for GSM1068514: Bronchial Epithelial Cells MMA biological 2
## 58      #GSM1068515 = Value for GSM1068515: Bronchial Epithelial Cells MMA biological 0
## 59      #GSM1068516 = Value for GSM1068516: Bronchial Epithelial Cells MMA biological 0
```

```
## 60      #GSM1068517 = Value for GSM1068517: Bronchial Epithelial Cells MMA biological 1
## 61      #GSM1068518 = Value for GSM1068518: Bronchial Epithelial Cells MMA biological 1
## 62      #GSM1068519 = Value for GSM1068519: Bronchial Epithelial Cells MMA biological 1
## 63      #GSM1068520 = Value for GSM1068520: Bronchial Epithelial Cells MMA biological 1
## 64      #GSM1068521 = Value for GSM1068521: Bronchial Epithelial Cells MMA biological 1
## 65      #GSM1068522 = Value for GSM1068522: Bronchial Epithelial Cells MMA biological 1
## 66      #GSM1068523 = Value for GSM1068523: Bronchial Epithelial Cells MMA biological 1
## 67      #GSM1068524 = Value for GSM1068524: Bronchial Epithelial Cells MMA biological 1
## 68      #GSM1068525 = Value for GSM1068525: Bronchial Epithelial Cells MMA biological 2
## 69      #GSM1068526 = Value for GSM1068526: Bronchial Epithelial Cells MMA biological 2
## 70      #GSM1068527 = Value for GSM1068527: Bronchial Epithelial Cells MMA biological 2
## 71       #GSM1068528 = Value for GSM1068528: Bronchial Epithelial Cells SA biological 0
## 72       #GSM1068529 = Value for GSM1068529: Bronchial Epithelial Cells SA biological 1
## 73       #GSM1068530 = Value for GSM1068530: Bronchial Epithelial Cells SA biological 1
## 74       #GSM1068531 = Value for GSM1068531: Bronchial Epithelial Cells SA biological 1
## 75       #GSM1068532 = Value for GSM1068532: Bronchial Epithelial Cells SA biological 1
## 76       #GSM1068533 = Value for GSM1068533: Bronchial Epithelial Cells SA biological 1
## 77       #GSM1068534 = Value for GSM1068534: Bronchial Epithelial Cells SA biological 1
## 78       #GSM1068535 = Value for GSM1068535: Bronchial Epithelial Cells SA biological 1
## 79       #GSM1068536 = Value for GSM1068536: Bronchial Epithelial Cells SA biological 1
## 80       #GSM1068537 = Value for GSM1068537: Bronchial Epithelial Cells SA biological 1
## 81       #GSM1068538 = Value for GSM1068538: Bronchial Epithelial Cells SA biological 1
## 82       #GSM1068539 = Value for GSM1068539: Bronchial Epithelial Cells SA biological 1
## 83       #GSM1068540 = Value for GSM1068540: Bronchial Epithelial Cells SA biological 1
## 84       #GSM1068541 = Value for GSM1068541: Bronchial Epithelial Cells SA biological 1
## 85       #GSM1068542 = Value for GSM1068542: Bronchial Epithelial Cells SA biological 1
## 86       #GSM1068543 = Value for GSM1068543: Bronchial Epithelial Cells SA biological 1
## 87       #GSM1068544 = Value for GSM1068544: Bronchial Epithelial Cells SA biological 1
## 88       #GSM1068545 = Value for GSM1068545: Bronchial Epithelial Cells SA biological 1
## 89       #GSM1068546 = Value for GSM1068546: Bronchial Epithelial Cells SA biological 1
## 90       #GSM1068547 = Value for GSM1068547: Bronchial Epithelial Cells SA biological 1
## 91       #GSM1068548 = Value for GSM1068548: Bronchial Epithelial Cells SA biological 1
## 92       #GSM1068549 = Value for GSM1068549: Bronchial Epithelial Cells SA biological 1
## 93       #GSM1068550 = Value for GSM1068550: Bronchial Epithelial Cells SA biological 1
## 94       #GSM1068551 = Value for GSM1068551: Bronchial Epithelial Cells SA biological 1
## 95       #GSM1068552 = Value for GSM1068552: Bronchial Epithelial Cells SA biological 1
## 96       #GSM1068553 = Value for GSM1068553: Bronchial Epithelial Cells SA biological 2
```

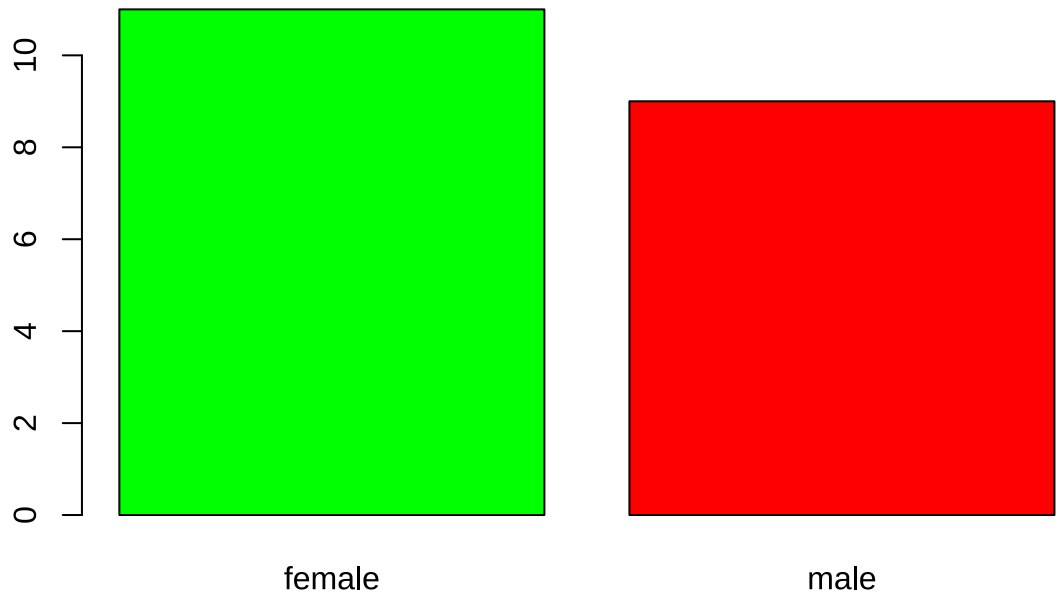
```
## 97       #GSM1068554 = Value for GSM1068554: Bronchial Epithelial Cells SA biological 2
## 98       #GSM1068555 = Value for GSM1068555: Bronchial Epithelial Cells SA biological 2
## 99       #GSM1068556 = Value for GSM1068556: Bronchial Epithelial Cells SA biological 2
## 100      #GSM1068557 = Value for GSM1068557: Bronchial Epithelial Cells SA biological 2
## 101      #GSM1068558 = Value for GSM1068558: Bronchial Epithelial Cells SA biological 2
## 102      #GSM1068559 = Value for GSM1068559: Bronchial Epithelial Cells SA biological 2
## 103      #GSM1068560 = Value for GSM1068560: Bronchial Epithelial Cells SA biological 2
## 104      #GSM1068561 = Value for GSM1068561: Bronchial Epithelial Cells SA biological 2
## 105      #GSM1068562 = Value for GSM1068562: Bronchial Epithelial Cells SA biological 2
## 106      #GSM1068563 = Value for GSM1068563: Bronchial Epithelial Cells SA biological 2
## 107      #GSM1068564 = Value for GSM1068564: Bronchial Epithelial Cells SA biological 2
## 108      #GSM1068565 = Value for GSM1068565: Bronchial Epithelial Cells SA biological 2
##          ID gender
## 1   control female
## 2   control female
## 3   control female
## 4   control female
## 5   control   male
## 6   control   male
## 7   control female
## 8   control   male
## 9   control   male
## 10  control   male
## 11  control female
## 12  control   male
## 13  control   male
## 14  control   male
## 15  control female
## 16  control female
## 17  control female
## 18  control   male
## 19  control female
## 20  control female
## 21      MMA female
## 22      MMA female
## 23      MMA   male
## 24      MMA female
```

```
## 25      MMA female
## 26      MMA female
## 27      MMA   male
## 28      MMA   male
## 29      MMA female
## 30      MMA female
## 31      MMA female
## 32      MMA   male
## 33      MMA female
## 34      MMA female
## 35      MMA female
## 36      MMA female
## 37      MMA female
## 38      MMA female
## 39      MMA female
## 40      MMA   male
## 41      MMA female
## 42      MMA female
## 43      MMA female
## 44      MMA   male
## 45      MMA female
## 46      MMA female
## 47      MMA   male
## 48      MMA female
## 49      MMA female
## 50      MMA female
## 51      MMA female
## 52      MMA   male
## 53      MMA female
## 54      MMA   male
## 55      MMA female
## 56      MMA female
## 57      MMA female
## 58      MMA   male
## 59      MMA   male
## 60      MMA female
## 61      MMA female
```

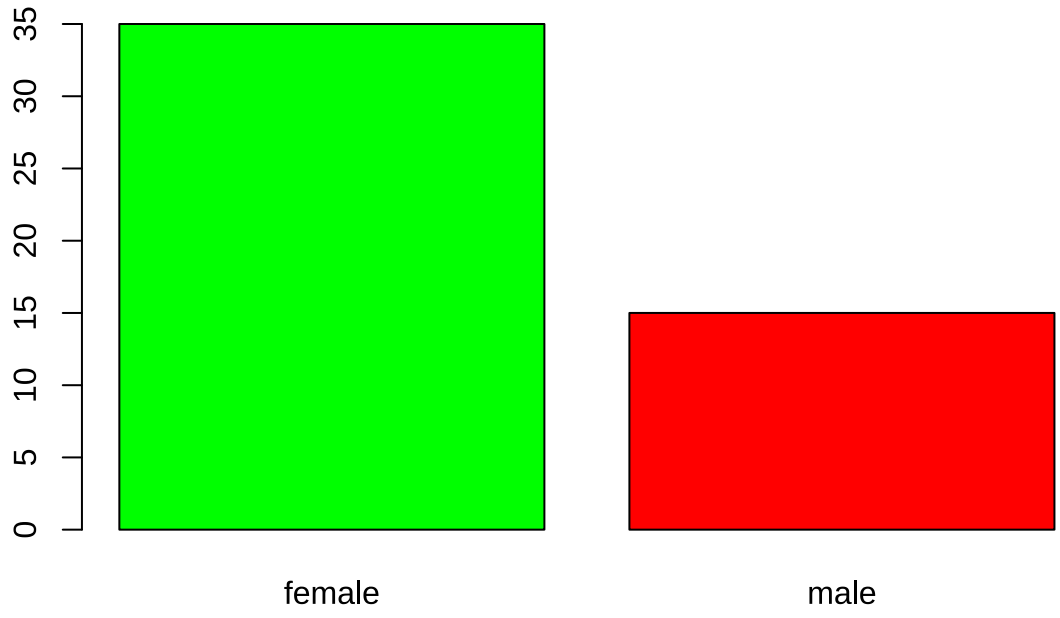
```
## 62      MMA   male
## 63      MMA female
## 64      MMA female
## 65      MMA female
## 66      MMA   male
## 67      MMA female
## 68      MMA   male
## 69      MMA   male
## 70      MMA female
## 71       SA female
## 72       SA   male
## 73       SA   male
## 74       SA female
## 75       SA female
## 76       SA female
## 77       SA   male
## 78       SA female
## 79       SA   male
## 80       SA female
## 81       SA female
## 82       SA female
## 83       SA female
## 84       SA   male
## 85       SA female
## 86       SA female
## 87       SA female
## 88       SA female
## 89       SA female
## 90       SA female
## 91       SA female
## 92       SA female
## 93       SA female
## 94       SA female
## 95       SA female
## 96       SA   male
## 97       SA   male
## 98       SA female
```

```
## 99       SA female
## 100      SA female
## 101      SA   male
## 102      SA   male
## 103      SA female
## 104      SA female
## 105      SA female
## 106      SA female
## 107      SA   male
## 108      SA female
```

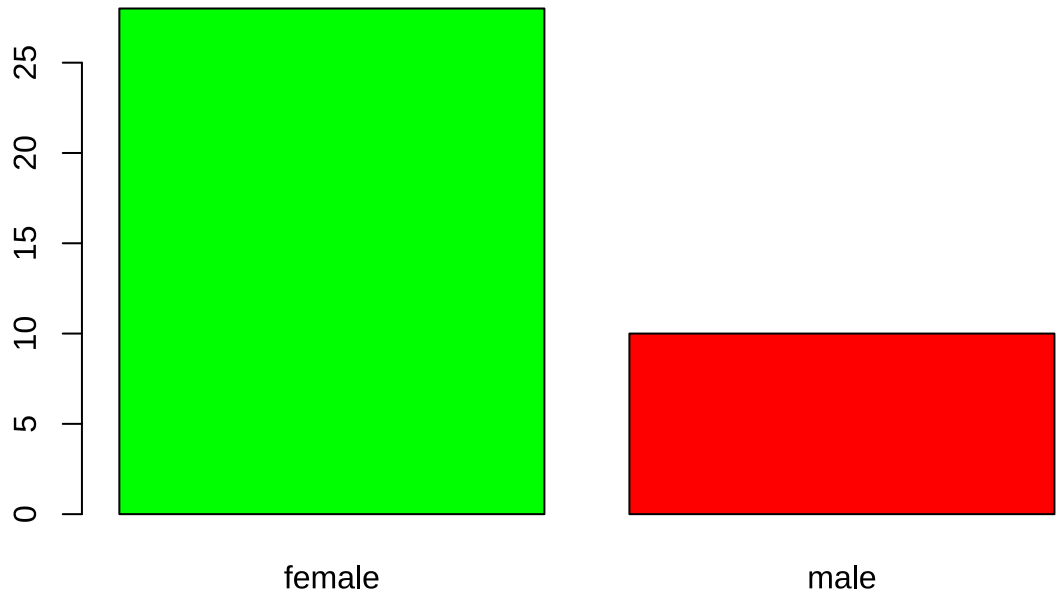
```r
write.csv(genderdata,file ='GDS5037/genderdata.csv' )

par(mforw=c(1,3))

row <- 1
value <- c(su[row,3]$Counts,su[row+1,3]$Counts)
gender <- c('female','male')
barplot(value,names.arg = gender,col = c('green','red'))
```



```
row <- 3
value <- c(su[row,3]$Counts,su[row+1,3]$Counts)
gender <- c('female','male')
barplot(value,names.arg = gender,col = c('green','red'))
```




```
row <- 5
value <- c(su[row,3]$Counts,su[row+1,3]$Counts)
gender <- c('female','male')
barplot(value,names.arg = gender,col = c('green','red'))
```




```
#2.4 group by patients' status,calculate the sample mean and
#variance of IDENTIFIER FAM174bB in each group
data_table <- read.csv('GDS5037/data_table.csv',nrows = 1);
#I know the data I need is the first
ttable <- t(data_table)
d <- ttable
names <- rownames(d)
rownames(d) <- NULL
ttable <- cbind(names,d);ttable <- ttable[-1,];ttable <- ttable[-1,]
ttable <- data.frame(ttable)
names(ttable) <- c('ID_REF','value')
data <- merge(ttable,description,by = 'ID_REF');
data <- data%>% mutate_at(.vars = vars('value'), .fun = as.numeric)
#data$value=as.data.frame(lapply(data$value,as.numeric))
mean <- aggregate(data[,2],list(data[,4]),mean);mean
```

```
##   Group.1        x
## 1 control 53.50000
## 2     MMA 58.74000
```

```
## 3      SA 47.76316
```

```
sd <- aggregate(data[,2],list(data[,4]),sd);sd
```

```
##   Group.1        x
## 1 control 26.44856
## 2     MMA 31.18268
## 3      SA 31.47838
```

```
n <- aggregate(data[,2],list(data[,4]),length);n
```

```
##   Group.1  x
## 1 control 20
## 2     MMA 50
## 3      SA 38
```

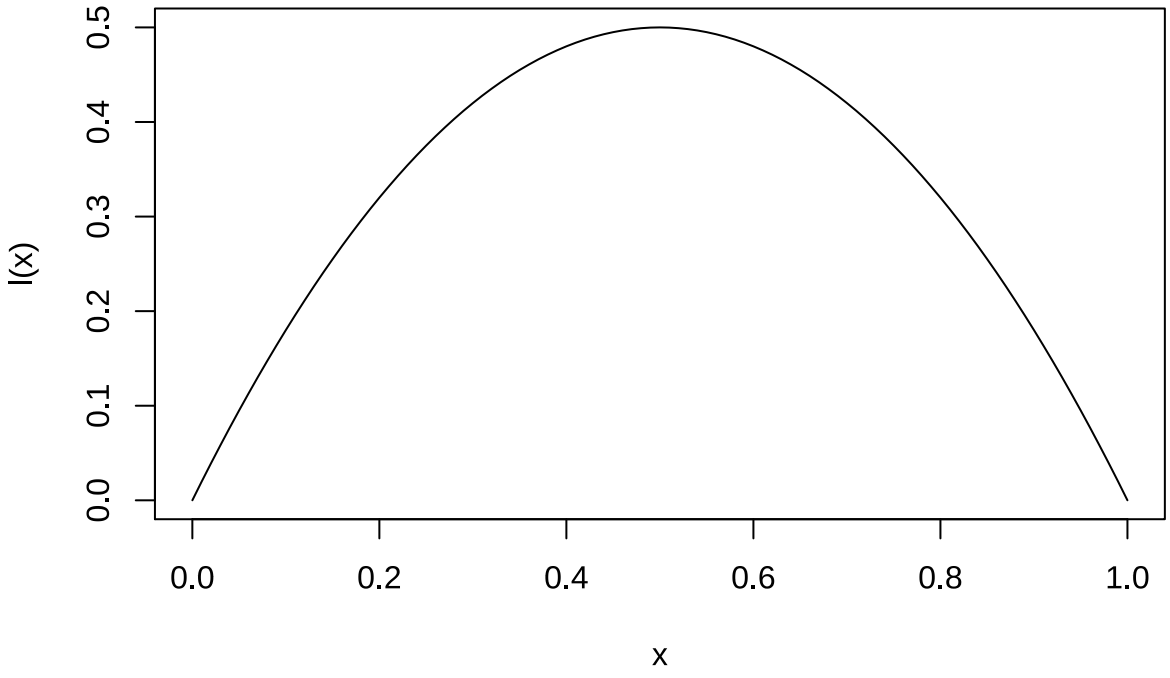
```
z <- (mean[1,2]-mean[3,2])/sqrt(sd[1,2]^2/n[1,2]+sd[3,2]^2/n[3,2])
p <- 1-pnorm(z);p
```

```
## [1] 0.2314096
```

```
#so the result show that the mean of the two groups are equal under 5% significant level
```

- 1.4 的 d 题有一个画图题

```
l <- function(x) 2*x*(1-x)
curve(l,0,1)
```




```
l <- function(x) -2*x*(1-x)
optimize(l,c(0,1))
```

```
## $minimum
## [1] 0.5
##
## $objective
## [1] -0.5
```

- so the result show that the minvalue is 0.5 when pi = 0.5
- 1.12

```
library(cdabookfunc)
binom_inference(0,25,0.5,method = 'wald')
```

```
## $z
## [1] -Inf
##
## $method
## [1] "wald"
```