R Markdown Demo

BUSN 32100

Oct 5 2019

In this demo we continue to work with the starwars data

##Packages We start with loading the packages we'll use.

```
#1
#a)
library(tidyverse)
state.df = data.frame(state.x77, Region=state.region, Division=state.division)
summary(state.df)
```

```
##    Population        Income       Illiteracy       Life.Exp
##  Min.   :  365   Min.   :3098   Min.   :0.500   Min.   :67.96
##  1st Qu.: 1080   1st Qu.:3993   1st Qu.:0.625   1st Qu.:70.12
##  Median : 2838   Median :4519   Median :0.950   Median :70.67
##  Mean   : 4246   Mean   :4436   Mean   :1.170   Mean   :70.88
##  3rd Qu.: 4968   3rd Qu.:4814   3rd Qu.:1.575   3rd Qu.:71.89
##  Max.   :21198   Max.   :6315   Max.   :2.800   Max.   :73.60
##
##      Murder          HS.Grad          Frost             Area
##  Min.   : 1.400   Min.   :37.80   Min.   :  0.00   Min.   :  1049
##  1st Qu.: 4.350   1st Qu.:48.05   1st Qu.: 66.25   1st Qu.: 36985
##  Median : 6.850   Median :53.25   Median :114.50   Median : 54277
##  Mean   : 7.378   Mean   :53.11   Mean   :104.46   Mean   : 70736
##  3rd Qu.:10.675   3rd Qu.:59.15   3rd Qu.:139.75   3rd Qu.: 81162
##  Max.   :15.100   Max.   :67.30   Max.   :188.00   Max.   :566432
##
##            Region                Division
##  Northeast    : 9   South Atlantic    : 8
##  South        :16   Mountain          : 8
##  North Central:12   West North Central: 7
##  West         :13   New England       : 6
##                     East North Central: 5
##                     Pacific           : 5
##                     (Other)           :11
```

```
state.df$Abbr<-state.abb
print(state.df[1:3,1:11])
```

```
##         Population Income Illiteracy Life.Exp Murder HS.Grad Frost   Area
## Alabama       3615   3624        2.1    69.05   15.1    41.3    20  50708
## Alaska         365   6315        1.5    69.31   11.3    66.7   152 566432
## Arizona       2212   4530        1.8    70.55    7.8    58.1    15 113417
##         Region           Division Abbr
## Alabama  South East South Central   AL
## Alaska    West            Pacific   AK
## Arizona   West           Mountain   AZ
```

```
#b)
print(state.df[1:3,-c(9)])
```

```
##         Population Income Illiteracy Life.Exp Murder HS.Grad Frost   Area
## Alabama       3615   3624        2.1    69.05   15.1    41.3    20  50708
## Alaska         365   6315        1.5    69.31   11.3    66.7   152 566432
## Arizona       2212   4530        1.8    70.55    7.8    58.1    15 113417
##                    Division Abbr
## Alabama East South Central   AL
## Alaska              Pacific   AK
## Arizona            Mountain   AZ
```

```
#Region gone!!
#c)
state.df$Center.x<-state.center$x
state.df$Center.y<-state.center$y
print(state.df[1:3,1:12])
```

```
##         Population Income Illiteracy Life.Exp Murder HS.Grad Frost   Area
## Alabama       3615   3624        2.1    69.05   15.1    41.3    20  50708
## Alaska         365   6315        1.5    69.31   11.3    66.7   152 566432
## Arizona       2212   4530        1.8    70.55    7.8    58.1    15 113417
##         Region           Division Abbr  Center.x
## Alabama  South East South Central   AL  -86.7509
## Alaska    West             Pacific   AK -127.2500
## Arizona   West            Mountain   AZ -111.6250
```

```
#d)
longless <- state.df[state.df$Center.x < -100,]
longless.data<- data.frame(longless)
#e
longless9 <- data_frame(state.df$Center.x < -100 & state.df$Murder > 9)
which(longless9==TRUE)
```

```
## [1]  2  5 28 31
```

```
#[1]  2  5 28 31
#these four states meet the -100 longitude +9% murder rate criteria
badstates<-which(longless9==TRUE)
state.df$Life.Exp[badstates]
```

```
## [1] 69.31 71.71 69.03 70.32
```

```
# 69.31 71.71 69.03 70.32
#State #5 has the highest life expectancy.
```

First, let's try to `knit` this R Markdown file to a html file. Click on the `knit` button on top. It should open a html page.

In class, we have seen these two plots. Now each member of your group can choose one block and modified the code with in that block to show a plot with only two genders, male and female. You can also choose

between `knit to html` or `knit to pdf`. Compare your knitted html/pdf with your other member of your group. Let us know if you encounter any error.

==========

```
rio = read.csv("https://raw.githubusercontent.com/BUSN32100/data_files/master/rio.csv")
#2a. What kind of object is rio? What are its dimensions and columns names of rio? What does each row r
summary(rio)
```

```
##        id                          name        nationality       sex
##  Min.   :    18347   Ahmed Mohamed :    2   USA    : 567   female:5205
##  1st Qu.:245099667   Ben Saxton    :    2   BRA    : 485   male  :6333
##  Median :500201062   Carli Lloyd   :    2   GER    : 441
##  Mean   :499988509   Daniel Vargas :    2   AUS    : 431
##  3rd Qu.:753987424   David Graf    :    2   FRA    : 410
##  Max.   :999987786   Felipe Aguilar:    2   CHN    : 404
##                      (Other)       :11526   (Other):8800
##     date_of_birth       height          weight             sport
##  1988-03-05:    9   Min.   :1.210   Min.   : 31.00   athletics:2363
##  1990-12-20:    9   1st Qu.:1.690   1st Qu.: 60.00   aquatics :1445
##  1993-02-18:    9   Median :1.760   Median : 70.00   football : 611
##  1988-04-03:    8   Mean   :1.766   Mean   : 72.07   rowing   : 547
##  1988-04-29:    8   3rd Qu.:1.840   3rd Qu.: 81.00   cycling  : 525
##  1989-03-01:    8   Max.   :2.210   Max.   :170.00   hockey   : 432
##  (Other)   :11487   NA's   :330     NA's   :659      (Other)  :5615
##       gold             silver            bronze        year_of_birth
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :1954
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:1986
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :1990
##  Mean   :0.05772   Mean   :0.05677   Mean   :0.06102   Mean   :1989
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:1993
##  Max.   :5.00000   Max.   :2.00000   Max.   :2.00000   Max.   :2002
##
```

```
#Looks like rio includes information on all the athletes participating in the Rio olympics.
typeof(rio)
```

```
## [1] "list"
```

```
dim(rio)
```

```
## [1] 11538    12
```

```
#[1] 11538    12
#each row is an individual athlete.
is.null(rio)
```

```
## [1] FALSE
```

```
#FALSE, looks like there is no missing data.
#2.b)
#11,538 athlets competed in the olympics
class(rio$nationality)
```

```
## [1] "factor"
```

```
levels(rio$nationality)
```

```
##   [1] "AFG" "ALB" "ALG" "AND" "ANG" "ANT" "ARG" "ARM" "ARU" "ASA" "AUS"
##  [12] "AUT" "AZE" "BAH" "BAN" "BAR" "BDI" "BEL" "BEN" "BER" "BHU" "BIH"
##  [23] "BIZ" "BLR" "BOL" "BOT" "BRA" "BRN" "BRU" "BUL" "BUR" "CAF" "CAM"
##  [34] "CAN" "CAY" "CGO" "CHA" "CHI" "CHN" "CIV" "CMR" "COD" "COK" "COL"
##  [45] "COM" "CPV" "CRC" "CRO" "CUB" "CYP" "CZE" "DEN" "DJI" "DMA" "DOM"
##  [56] "ECU" "EGY" "ERI" "ESA" "ESP" "EST" "ETH" "FIJ" "FIN" "FRA" "FSM"
##  [67] "GAB" "GAM" "GBR" "GBS" "GEO" "GEQ" "GER" "GHA" "GRE" "GRN" "GUA"
##  [78] "GUI" "GUM" "GUY" "HAI" "HKG" "HON" "HUN" "INA" "IND" "IOA" "IRI"
##  [89] "IRL" "IRQ" "ISL" "ISR" "ISV" "ITA" "IVB" "JAM" "JOR" "JPN" "KAZ"
## [100] "KEN" "KGZ" "KIR" "KOR" "KOS" "KSA" "LAO" "LAT" "LBA" "LBR" "LCA"
## [111] "LES" "LIB" "LIE" "LTU" "LUX" "MAD" "MAR" "MAS" "MAW" "MDA" "MDV"
## [122] "MEX" "MGL" "MHL" "MKD" "MLI" "MLT" "MNE" "MON" "MOZ" "MRI" "MTN"
## [133] "MYA" "NAM" "NCA" "NED" "NEP" "NGR" "NIG" "NOR" "NRU" "NZL" "OMA"
## [144] "PAK" "PAN" "PAR" "PER" "PHI" "PLE" "PLW" "PNG" "POL" "POR" "PRK"
## [155] "PUR" "QAT" "ROT" "ROU" "RSA" "RUS" "RWA" "SAM" "SEN" "SEY" "SIN"
## [166] "SKN" "SLE" "SLO" "SMR" "SOL" "SOM" "SRB" "SRI" "SSD" "STP" "SUD"
## [177] "SUI" "SUR" "SVK" "SWE" "SWZ" "SYR" "TAN" "TGA" "THA" "TJK" "TKM"
## [188] "TLS" "TOG" "TPE" "TTO" "TUN" "TUR" "TUV" "UAE" "UGA" "UKR" "URU"
## [199] "USA" "UZB" "VAN" "VEN" "VIE" "VIN" "YEM" "ZAM" "ZIM"
```

```
summary(rio$nationality)
```

```
##     USA     BRA     GER     AUS     FRA     CHN     GBR     JPN     CAN
##     567     485     441     431     410     404     374     346     321
##     ESP     ITA     RUS     NED     POL     ARG     KOR     NZL     UKR
##     313     312     286     249     242     223     213     208     205
##     SWE     COL     HUN     RSA     DEN     MEX     BLR     CUB     IND
##     164     154     154     146     128     126     124     123     123
##     EGY     BEL     CZE     SUI     KAZ     SRB     TUR     ROU     POR
##     122     108     104     104     103     103     103      98      95
##     GRE     CRO     VEN     IRL     KEN     NGR     AUT     UZB     ALG
##      93      88      88      80      80      78      71      70      68
##     LTU     IRI     SLO     NOR     TUN     JAM     AZE     TPE     FIJ
##      67      64      63      62      61      57      56      56      54
##     FIN     THA     SVK     BUL     MAR     ISR     EST     MGL     CHI
##      54      54      51      50      49      47      46      43      42
##     GEO     PUR     QAT     ECU     ETH     HKG     MNE     ZIM     BRN
##      40      40      39      38      38      38      35      35      34
##     ARM     LAT     MAS     TTO     PRK     BAH     HON     DOM     PER
##      32      32      32      32      31      30      30      29      29
##     INA     ANG     IRQ     SIN     CMR     MDA     VIE     SEN     GUA
##      28      26      26      25      24      23      23      22      21
##     UGA     KGZ     URU     CYP     GHA     PHI     UAE     BOL     BOT
##      21      19      17      16      16      13      13      12      12
## (Other)
##     663
```

```r
#The US has the most athletes at 567
summary(rio$sex)
```

```
## female   male
##   5205   6333
```

```r
#female   male
#  5205   6333
subset.Male <- rio[rio$sex == 'male',]
#Brazil has the most male athlets at 269
summary(subset.Male$nationality)
```

```
##     BRA     USA     GER     FRA     AUS     GBR     JPN     ITA     ESP
##     269     264     237     234     211     205     179     168     167
##     CHN     ARG     RUS     POL     CAN     NED     KOR     NZL     RSA
##     153     146     144     142     130     110     107     103      97
##     CUB     HUN     DEN     UKR     EGY     MEX     COL     BEL     SWE
##      88      88      87      87      85      82      78      74      73
##     CRO     IND     POR     VEN     CZE     ALG     SRB     SUI     GRE
##      69      67      66      63      62      58      58      58      56
##     IRI     KAZ     TUR     BLR     IRL     NGR     LTU     KEN     UZB
##      55      55      55      54      53      52      48      47      47
##     AZE     TUN     SLO     AUT     FIJ     QAT     ROU     SVK     GEO
##      42      41      39      37      37      37      34      32      30
##     MAR     BUL     NOR     EST     HON     FIN     IRQ     JAM     MGL
##      30      29      29      28      28      26      26      26      26
##     THA     TPE     CHI     ARM     ECU     ISR     DOM     TTO     BRN
##      26      26      25      24      23      22      21      21      20
##     LAT     ETH     BAH     INA     MAS     MNE     PER     GUA     HKG
##      19      18      17      17      17      17      17      15      15
##     MDA     UGA     PUR     KGZ     URU     ERI     PRK     CYP     VIE
##      14      14      13      12      12      11      11      10      10
##     BOT     SIN     UAE     ZIM     ANG     IOA     SEY     ANT     BAR
##       9       9       9       9       8       8       8       7       7
## (Other)
##     384
```

```r
subset.Female <- rio[rio$sex == 'female',]
#US has the most female athlets at 303.
summary(subset.Female$nationality)
```

```
##     USA     CHN     AUS     BRA     GER     CAN     FRA     GBR     JPN
##     303     251     220     216     204     191     176     169     167
##     ESP     ITA     RUS     NED     UKR     KOR     NZL     POL     SWE
##     146     144     142     139     118     106     105     100      91
##     ARG     COL     BLR     HUN     ROU     IND     RSA     KAZ     TUR
##      77      76      70      66      64      56      49      48      48
##     SUI     SRB     MEX     CZE     DEN     EGY     GRE     CUB     AUT
##      46      45      44      42      41      37      37      35      34
##     BEL     KEN     NOR     JAM     TPE     POR     FIN     THA     IRL
##      34      33      33      31      30      29      28      28      27
##     PUR     NGR     ZIM     ISR     VEN     SLO     HKG     UZB     BUL
```

```
##      27      26      26      25      25      24      23      23      21
##     ETH     PRK     TUN     CMR     CRO     LTU     MAR     SVK     ANG
##      20      20      20      19      19      19      19      19      18
##     EST     MNE     CHI     FIJ     MGL     SEN     SIN     ECU     MAS
##      18      18      17      17      17      16      16      15      15
##     AZE     BRN     BAH     LAT     VIE     PER     INA     TTO     ALG
##      14      14      13      13      13      12      11      11      10
##     GEO     GHA     IRI     MDA     ARM     DOM     CIV     KGZ     UGA
##      10       9       9       9       8       8       7       7       7
##     BOL     CYP     GUA     MRI     NAM     PAN     PHI     COK     CRC
##       6       6       6       6       6       6       6       5       5
## (Other)
##     250
```

```
#Challenge)
summary(rio$name)
```

```
##                       Ahmed Mohamed                         Ben Saxton
##                                   2                                   2
##                         Carli Lloyd                       Daniel Vargas
##                                   2                                   2
##                          David Graf                      Felipe Aguilar
##                                   2                                   2
##                      Gabriella Szucs                       Ivan Ivanov
##                                   2                                   2
##                        Ivan Zaytsev                       Jianan Wang
##                                   2                                   2
##                         Kevin Lopez                      Kuk Hyang Kim
##                                   2                                   2
##                         Ling Zhang                         Luis Lopez
##                                   2                                   2
##                       Mohamed Amer                        Paola Perez
##                                   2                                   2
##                            Qian Li                       Ryan Cochrane
##                                   2                                   2
##                          Yan Wang                           Zhen Wang
##                                   2                                   2
##                      Zsofia Kovacs                     A Jesus Garcia
##                                   2                                   1
##                       A Lam Shin                        Aaron Brown
##                                   1                                   1
##                        Aaron Cook                         Aaron Gate
##                                   1                                   1
##                       Aaron Royle                       Aaron Russell
##                                   1                                   1
##                      Aaron Younger                   Aauri Lorena Bokesa
##                                   1                                   1
##                     Ababel Yeshaneh                       Abadi Hadis
##                                   1                                   1
##                  Abbas Abubakar Abbas                        Abbas Qali
##                                   1                                   1
##                     Abbey D'Agostino                     Abbey Weitzeil
##                                   1                                   1
##                          Abbie Brown                      Abbos Rakhmonov
```

```
##                                 1                                 1
##                  Abbubaker Mobara                       Abby Erceg
##                                 1                                 1
##           Abd Elhalim Mohamed Abou                  Abdalaati Iguider
##                                 1                                 1
##                  Abdalelah Haroun                     Abdalla Targan
##                                 1                                 1
##                Abdel Aziz Mehelba                Abdelati El Guesse
##                                 1                                 1
##                Abdelaziz Merzougui           Abdelaziz Mohamed Ahmed
##                                 1                                 1
##                 Abdelghani Demmou              Abdelhafid Benchabla
##                                 1                                 1
##                Abdelhakim Amokrane                  Abdelkader Chadi
##                                 1                                 1
##                  Abdelkadir Salhi                Abdelkebir Ouaddar
##                                 1                                 1
##                Abdelkhalek Elbanna   Abdellatif Mohamed Ahmed Mohamed
##                                 1                                 1
##               Abdelmajid El Hissouf                Abdelmalik Lahoulou
##                                 1                                 1
## Abdelrahman Salah Orabi Abdelgawwad                Abdelraouf Benguit
##                                 1                                 1
##               Abderrahmane Benamadi             Abderrahmane Mansouri
##                                 1                                 1
##                Abderrahmane Meziane                   Abdi Hakin Ulad
##                                 1                                 1
##                       Abdi Nageeye               Abdi Waiss Mouhyadin
##                                 1                                 1
##           Abdoul Khadre Mbaye Niane                Abdoulkarim Fawziya
##                                 1                                 1
##                 Abdoullah Bamoussa        Abdoulrazak Issoufou Alfaga
##                                 1                                 1
##                      Abdul Khalili                         Abdul Omar
##                                 1                                 1
##                 Abdul Wahab Zahiri                Abdulaziz Alshatti
##                                 1                                 1
##              Abdulkadir Abdullayev            Abdullah Abkar Mohammed
##                                 1                                 1
##                 Abdullah Alrashidi                Abdullah Hel Baki
##                                 1                                 1
##                    Abdullahi Shehu                   Abdullo Tangriev
##                                 1                                 1
##               Abdulrahman Al Faihan               Abdulrashid Sadulaev
##                                 1                                 1
##                  Abdulrazzaq Murad               Abeku Gyekye Jackson
##                                 1                                 1
##                     Abhinav Bindra                   Abigail Johnston
##                                 1                                 1
##                         Abigel Joo                Ablaikhan Zhussupov
##                                 1                                 1
##          Abraham Kipchirchir Rotich            Abraham Naibei Cheroben
##                                 1                                 1
##                  Abraham Niyonkuru                       Abrar Osman
```

```
##                                   1                                   1
##             Abubaker Haydar Abdalla                     Achraf Kharroubi
##                                   1                                   1
##                         Adam Batirov                        Adam Cwalina
##                                   1                                   1
##                          Adam Decker                         Adam Dixon
##                                   1                                   1
##                          Adam Froese                           (Other)
##                                   1                                 11418
```

```
sum(summary(rio$name) > 1)
```

```
## [1] 22
```

```
#22 people share the same name

#2c)
sum(rio$gold)
```

```
## [1] 666
```

```
#[1] 666
sum(rio$silver)
```

```
## [1] 655
```

```
# [1] 655
sum(rio$bronze)
```

```
## [1] 704
```

```
# [1] 704
666+655+704
```

```
## [1] 2025
```

```
#[1] 2025
#Surpirsingly, there were different amounts of each medal awarded. I did my research and apparently thi
rio <- data.frame (rio, total = rio$gold + rio$silver + rio$bronze)
max(rio$total) #The athlete with most medals got 6 (Michael Phelps)
```

```
## [1] 6
```

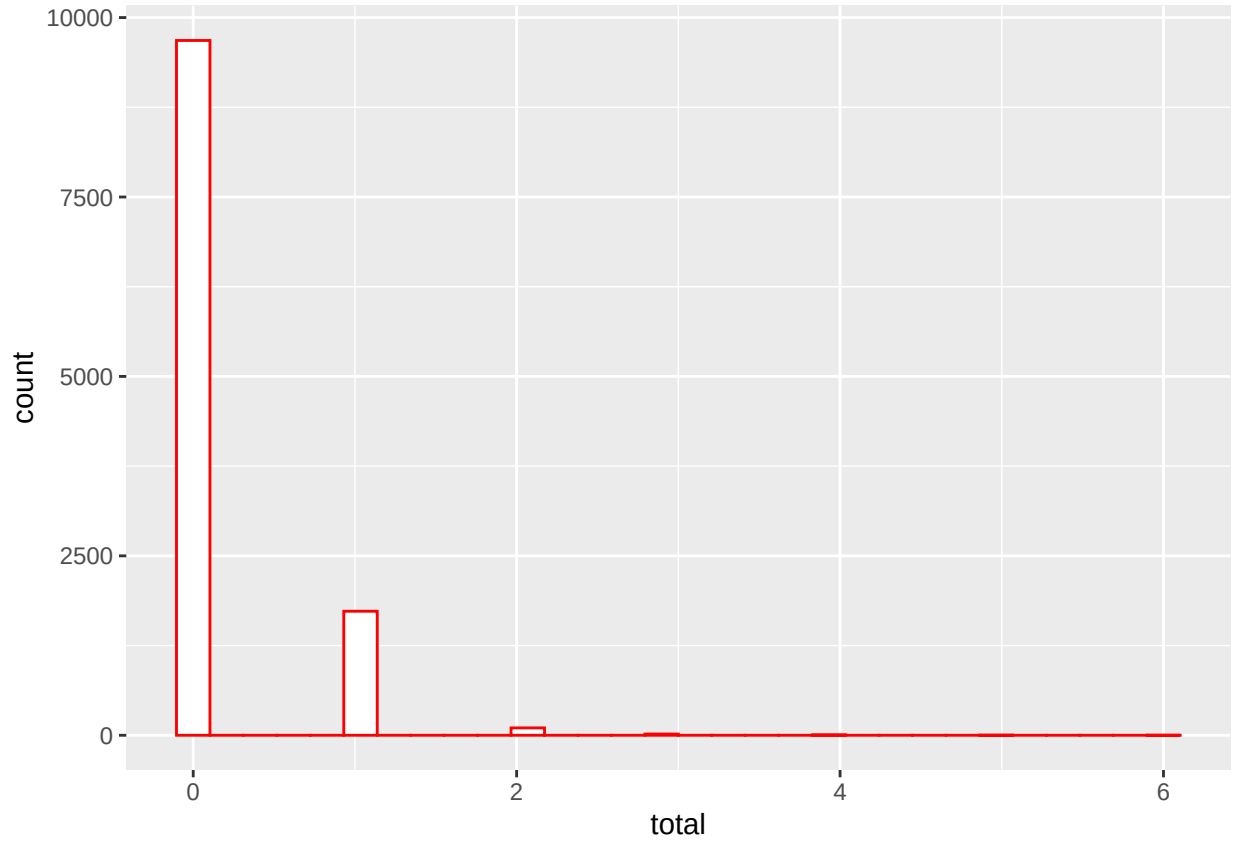
```
rio[rio$total == 6,]
```

```
##             id           name nationality  sex date_of_birth height weight
## 7402 491565031 Michael Phelps         USA male    1985-06-30   1.94     90
##         sport gold silver bronze year_of_birth total
## 7402 aquatics    5      1      0          1985     6
```

2. Segmented bar plots, proportions

```
#3a)
library("ggplot2")
ggplot(data=rio, mapping=aes(x=total))+geom_histogram(color = 'red', fill='white')
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
#b)
rio2 <- data.frame(rio, age = 2016 - rio$year_of_birth)
max(rio2$age)
```

```
## [1] 62
```

```
#[1] 62
rio[rio2$age == 62,]
```

```
##             id           name nationality    sex date_of_birth height
## 5300 271404469 Julie Brougham         NZL female    1954-05-20   1.57
## 7093 590552399     Mary Hanna         AUS female    1954-12-01   1.73
##      weight      sport gold silver bronze year_of_birth total
## 5300     48 equestrian    0      0      0          1954     0
## 7093     63 equestrian    0      0      0          1954     0
```

```r
#Julie Brougham from New Zealand and Mary Hanna from Australia are the oldest.
min(rio2$age) #min age is 14
```

```
## [1] 14
```

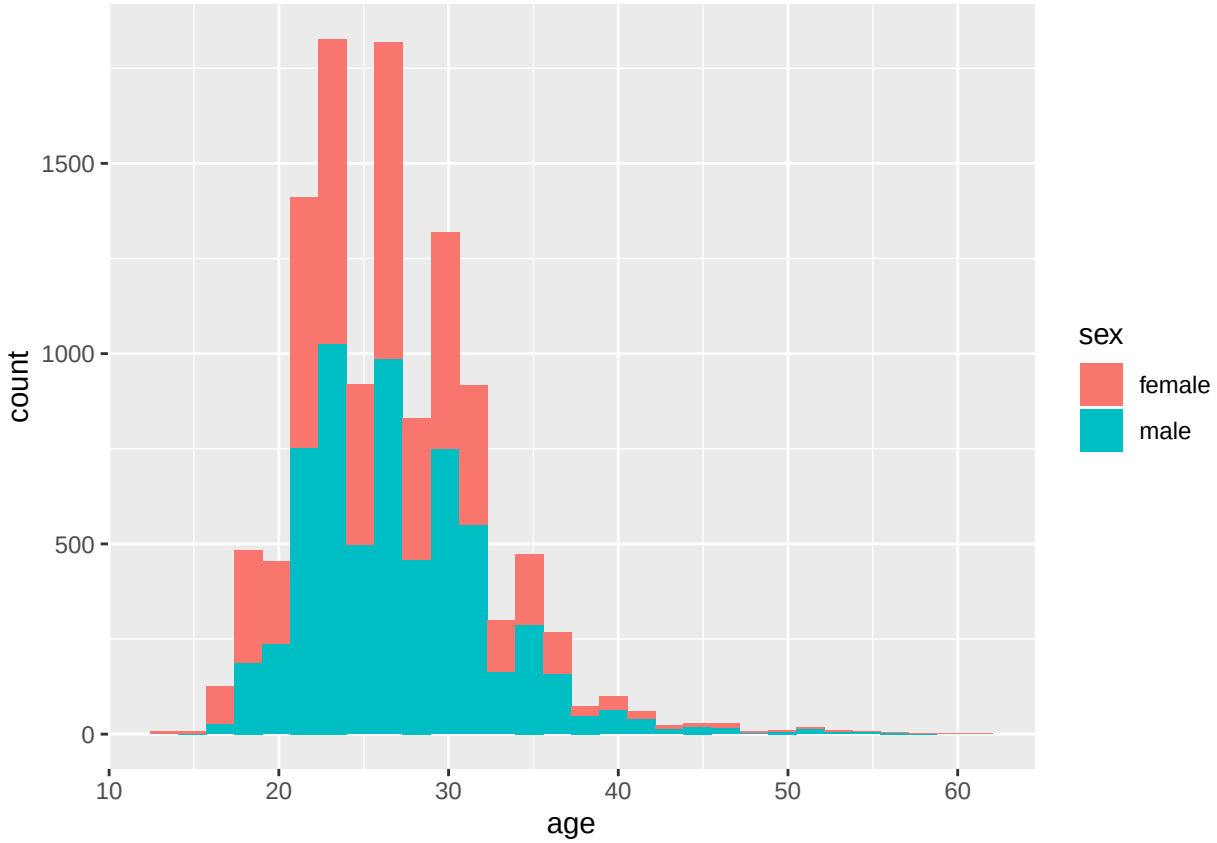
```r
rio[rio2$age == 14,] #there are 8 14 year olds,  7 female and 1 male, all participated in aquatics
```

```
##              id                 name nationality    sex date_of_birth
## 655   209671126     Ana Iulia Dascal         ROU female    2002-09-12
## 2432  380938305      Darya Semyonova         TKM female    2002-05-28
## 3306   91359398    Fatima Alkaramova         AZE female    2002-06-26
## 3599   32924852        Gaurika Singh         NEP female    2002-11-26
## 5577   55365531    Kaya Adwoa Forson         GHA female    2002-03-19
## 9919  112175885 Siri Arun Budcharern         LAO female    2002-01-12
## 10434 326914230          Thint Myaat         MYA   male    2002-04-14
## 11149 188592965            Yanhan Ai         CHN female    2002-02-07
##       height weight    sport gold silver bronze year_of_birth total
## 655     1.83     60 aquatics    0      0      0          2002     0
## 2432    1.70     50 aquatics    0      0      0          2002     0
## 3306    1.75     60 aquatics    0      0      0          2002     0
## 3599    1.55     45 aquatics    0      0      0          2002     0
## 5577      NA     NA aquatics    0      0      0          2002     0
## 9919    1.66     63 aquatics    0      0      0          2002     0
## 10434   1.60     52 aquatics    0      0      0          2002     0
## 11149   1.68     54 aquatics    0      0      0          2002     0
```

```r
ggplot(data=rio2, mapping=aes(x=age, fill=sex))+geom_histogram(bindwidth=10)
```

```
## Warning: Ignoring unknown parameters: bindwidth
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
rio3 <- rio2[rio2$gold>0,]
max(rio3$age)
```

```
## [1] 59
```

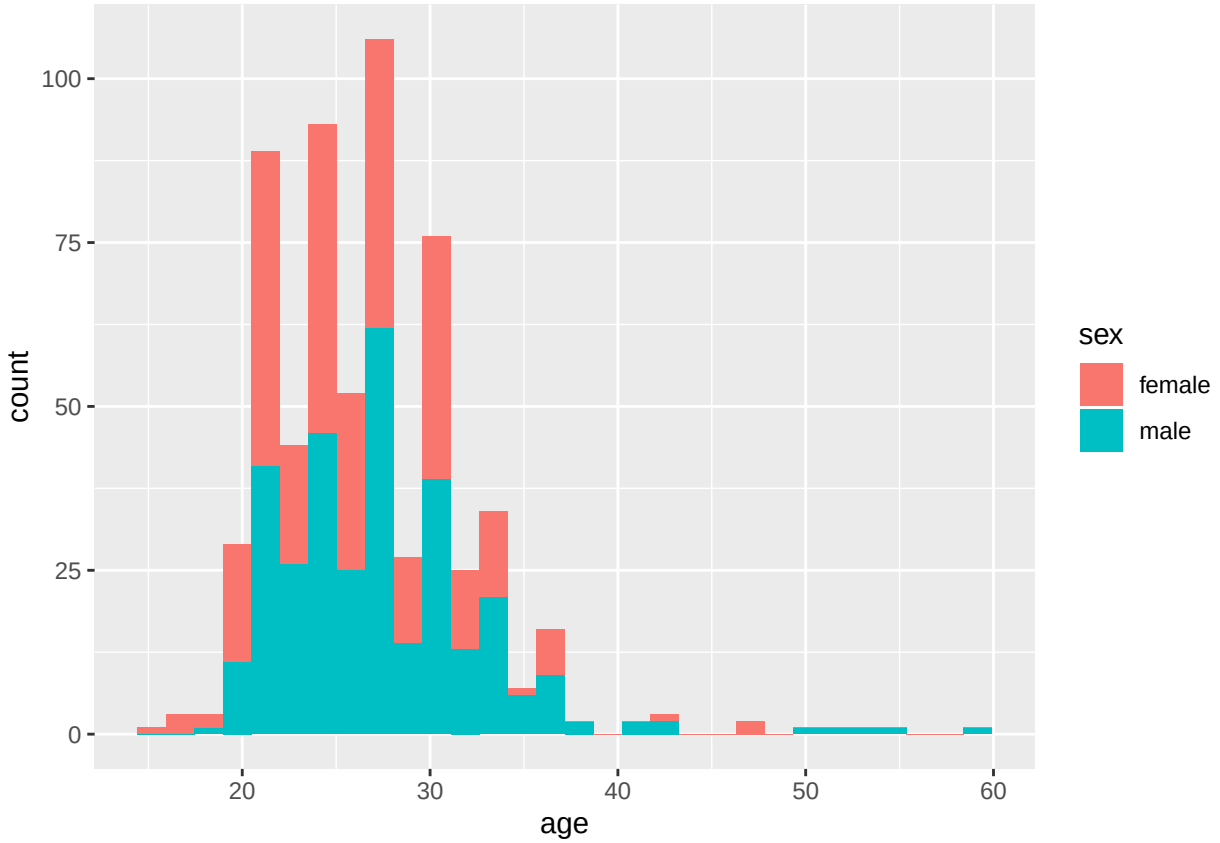
```
#[1] 59, The max age is 59, Nick Skelton from Great Britain
min(rio3$age)
```

```
## [1] 15
```

```
#[1] 15 Minimum age is 15, Qian Ren from China
ggplot(data=rio3, mapping=aes(x=age, fill=sex))+geom_histogram(bindwidth=10)
```

```
## Warning: Ignoring unknown parameters: bindwidth
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
#Looks like the athletes who won a medal tend to be younger
summary(rio3)
```

```
##        id                                name      nationality      sex
##  Min.   :  4878555   Abbey Weitzeil       :  1   USA    :119   female:294
##  1st Qu.:258965361   Abdulrashid Sadulaev:  1   GBR    : 59   male  :325
##  Median :507639455   Adam Peaty           :  1   RUS    : 49
##  Mean   :509007736   Agustin Mazzilli     :  1   GER    : 46
##  3rd Qu.:754494552   Ahmad Abughaush      :  1   CHN    : 42
##  Max.   :999578859   Aisen Chen           :  1   BRA    : 37
##                      (Other)              :613   (Other):267
##    date_of_birth     height          weight            sport
##  1988-09-29:  3   Min.   :1.450   Min.   : 40.00   aquatics : 99
##  1994-05-14:  3   1st Qu.:1.700   1st Qu.: 63.00   athletics: 60
##  1979-07-21:  2   Median :1.780   Median : 73.00   rowing   : 48
##  1980-08-01:  2   Mean   :1.792   Mean   : 74.87   football : 36
##  1982-05-06:  2   3rd Qu.:1.880   3rd Qu.: 85.00   hockey   : 34
##  1982-07-31:  2   Max.   :2.210   Max.   :157.00   handball : 29
##  (Other)   :605   NA's   :3       NA's   :17       (Other)  :313
##       gold            silver            bronze        year_of_birth
##  Min.   :1.000   Min.   :0.00000   Min.   :0.00000   Min.   :1957
##  1st Qu.:1.000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:1986
##  Median :1.000   Median :0.00000   Median :0.00000   Median :1990
##  Mean   :1.076   Mean   :0.06624   Mean   :0.05977   Mean   :1989
##  3rd Qu.:1.000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:1993
##  Max.   :5.000   Max.   :2.00000   Max.   :2.00000   Max.   :2001
```

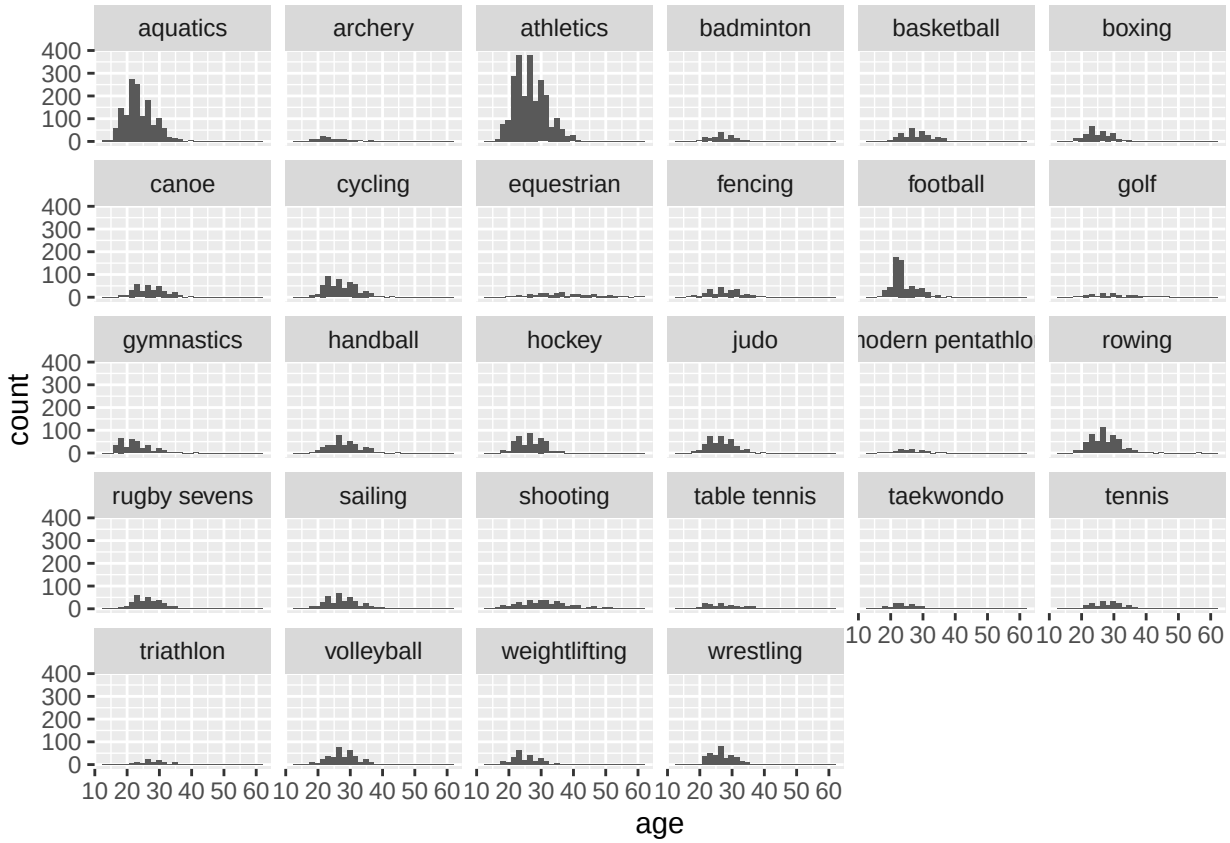
```
##
##      total            age
##  Min.   :1.000   Min.   :15.00
##  1st Qu.:1.000   1st Qu.:23.00
##  Median :1.000   Median :26.00
##  Mean   :1.202   Mean   :26.97
##  3rd Qu.:1.000   3rd Qu.:30.00
##  Max.   :6.000   Max.   :59.00
##
```

```
#Gold medalists are on average 26.97 years old
#I think age is not much of a factor
summary(rio)
```

```
##        id                          name        nationality       sex
##  Min.   :    18347   Ahmed Mohamed :    2   USA    : 567   female:5205
##  1st Qu.:245099667   Ben Saxton    :    2   BRA    : 485   male  :6333
##  Median :500201062   Carli Lloyd   :    2   GER    : 441
##  Mean   :499988509   Daniel Vargas :    2   AUS    : 431
##  3rd Qu.:753987424   David Graf    :    2   FRA    : 410
##  Max.   :999987786   Felipe Aguilar:    2   CHN    : 404
##                      (Other)       :11526   (Other):8800
##     date_of_birth       height          weight             sport
##  1988-03-05:    9   Min.   :1.210   Min.   : 31.00   athletics:2363
##  1990-12-20:    9   1st Qu.:1.690   1st Qu.: 60.00   aquatics :1445
##  1993-02-18:    9   Median :1.760   Median : 70.00   football : 611
##  1988-04-03:    8   Mean   :1.766   Mean   : 72.07   rowing   : 547
##  1988-04-29:    8   3rd Qu.:1.840   3rd Qu.: 81.00   cycling  : 525
##  1989-03-01:    8   Max.   :2.210   Max.   :170.00   hockey   : 432
##  (Other)   :11487   NA's   :330     NA's   :659      (Other)  :5615
##       gold             silver            bronze        year_of_birth
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :1954
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:1986
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :1990
##  Mean   :0.05772   Mean   :0.05677   Mean   :0.06102   Mean   :1989
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:1993
##  Max.   :5.00000   Max.   :2.00000   Max.   :2.00000   Max.   :2002
##
##      total
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.1755
##  3rd Qu.:0.0000
##  Max.   :6.0000
##
```

```
#I'm noticing that a disporportionate number of Americans win the gold medal.
#d)
ggplot(data=rio2, mapping = aes(x=age))+geom_histogram()+facet_wrap(.~sport)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
range(rio2[rio2$sport=='golf',]$age)
```

```
## [1] 18 47
```

```
#[1] 18 47 pretty big age range
range(rio2[rio2$sport=='gymnastics',]$age)
```

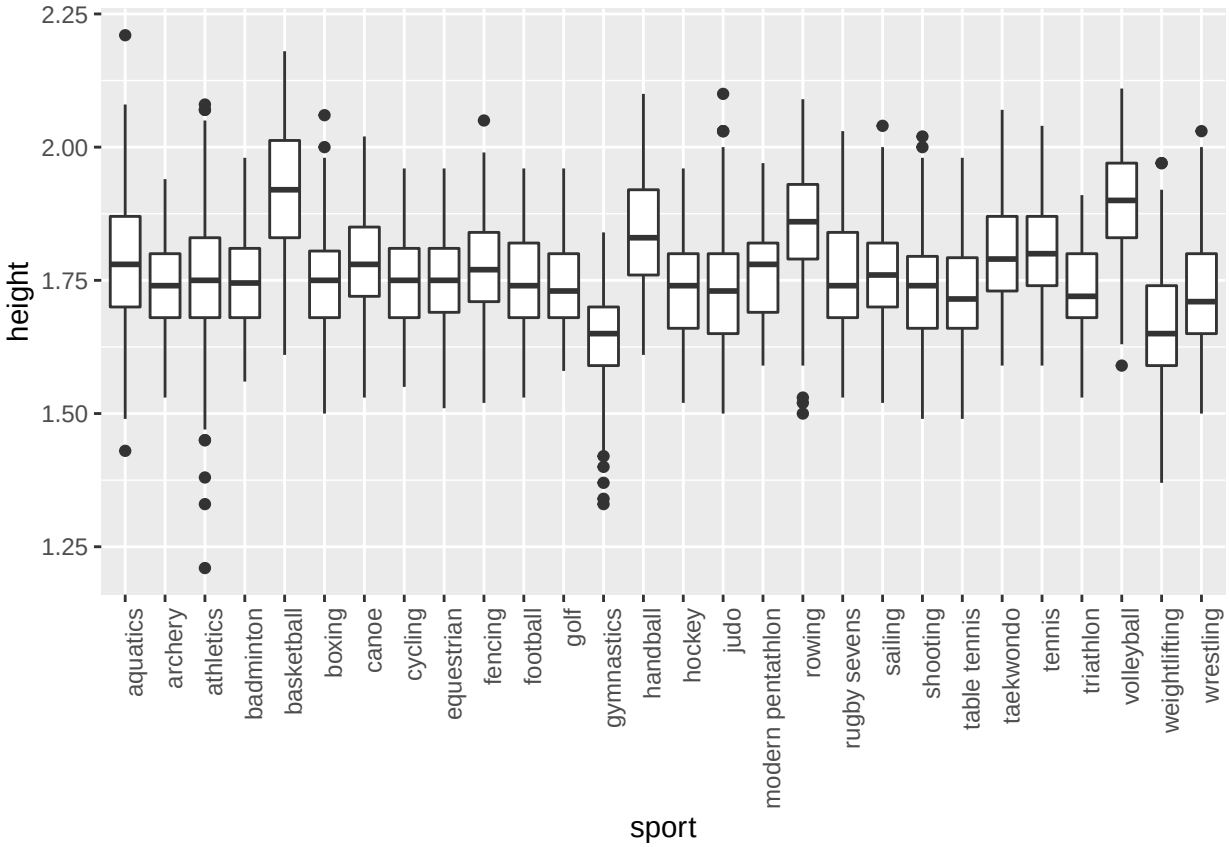
```
## [1] 16 41
```

```
#[1] 16 41, slightly narrower range than golf
range(rio2[rio2$sport=='rowing',]$age)
```

```
## [1] 18 57
```

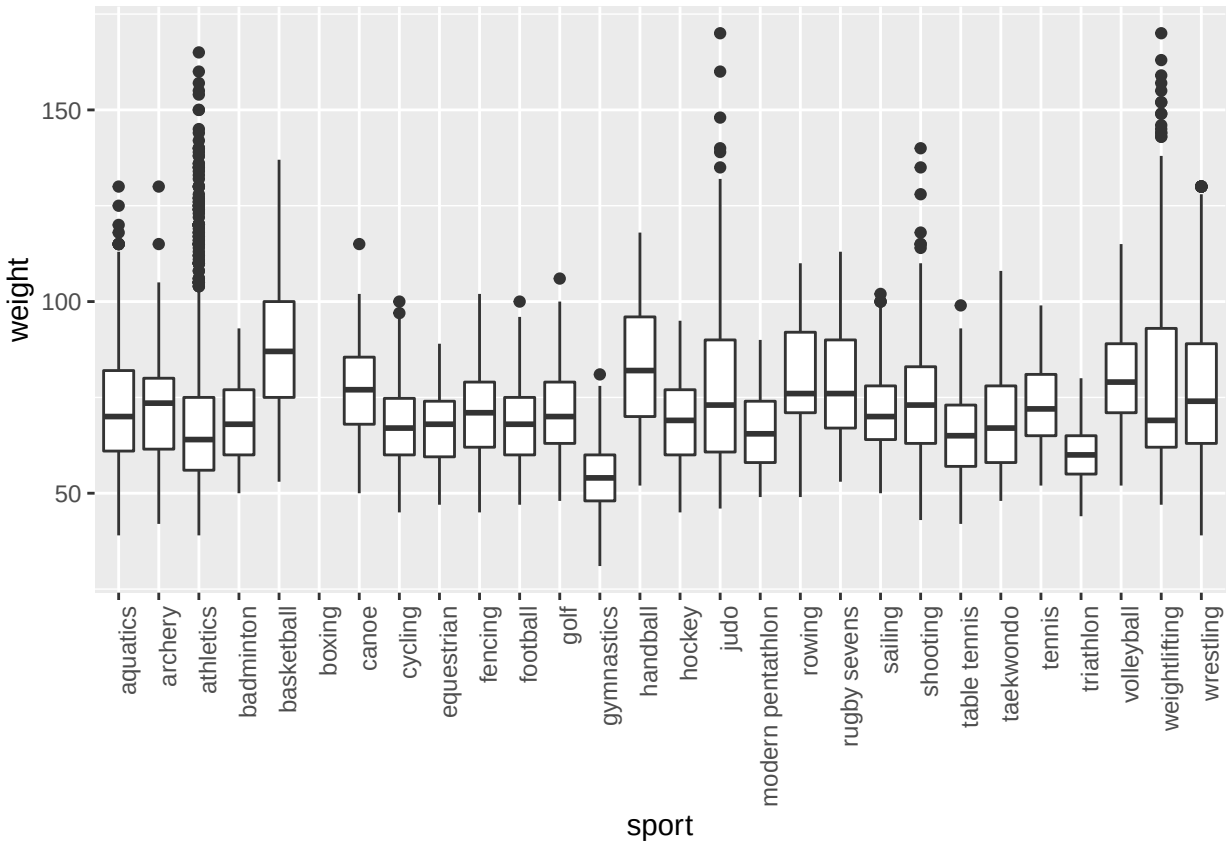
```
#[1] 18 57, more older people
#Makes sense since rowing isn't as physically taxing as gymnastics

#e)
ggplot(data=rio, mapping=aes(y=height, x=sport))+geom_boxplot() +theme(axis.text.x = element_text(angle
```

```
## Warning: Removed 330 rows containing non-finite values (stat_boxplot).
```



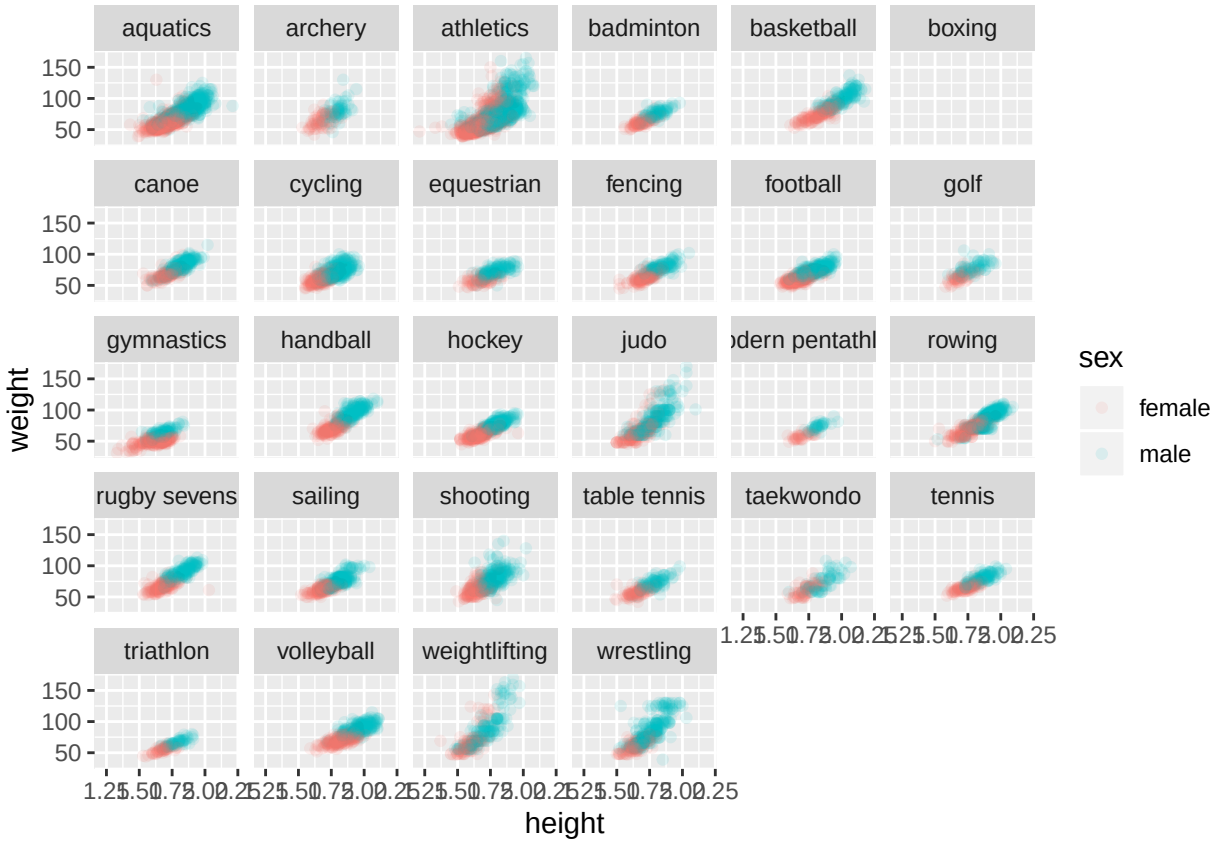

```
ggplot(data=rio, mapping=aes(y=weight, x=sport))+geom_boxplot() +theme(axis.text.x = element_text(angle
```

```
## Warning: Removed 659 rows containing non-finite values (stat_boxplot).
```

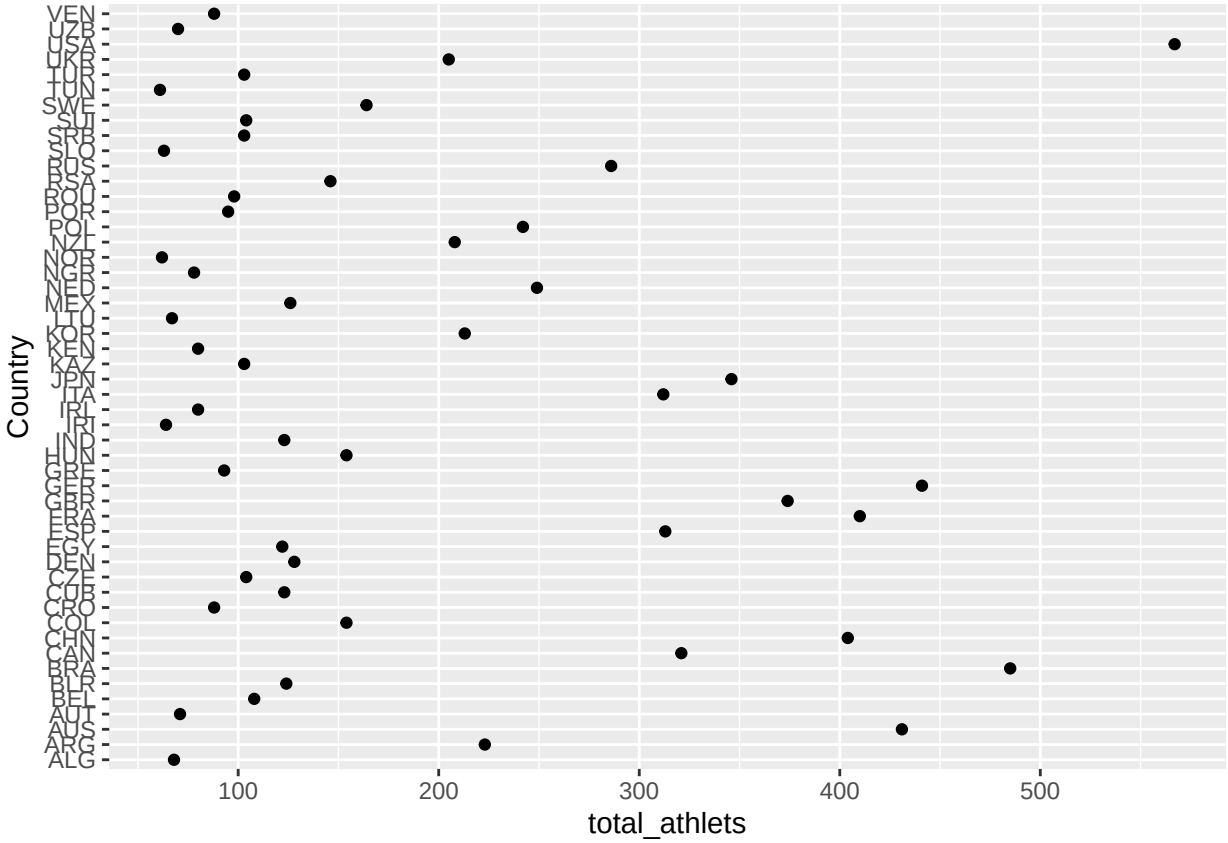



```
#f)
ggplot(data=rio, mapping=aes(x=height, y=weight, color=sex))+geom_point(alpha = 0.1) + facet_wrap(.~spo
```

```
## Warning: Removed 680 rows containing missing values (geom_point).
```




```
#Challenge)
counts <- summary(rio$nationality)
counts <- as.data.frame (counts)
Country <- row.names(counts)

counts <- data.frame("total_athletes"=counts,Country)

colnames(counts)[1] <- "total_athletes"

ggplot(data=counts[1:50,], mapping=aes(x=total_athletes, y=Country))+geom_point()
```



3. Segmented counts/proportions plots

‘‘‘

4. Knit to PDF

To be able to knit PDF file from Rmarkdown. Try run the following lines. You will see a warning message saying two error messages will be displayed. Acknowledge it. After two error messages and a long installation process, you should have the ability to generate PDF from Rmarkdown files. **You may want to do this after the class today, you will need the Rstudio to complete the group lab**

To check if you have install successfully. Run these lines in your console. You should see an output of `[1] "test.pdf"` and a `test.pdf` now in your folder.