STA130H1S – Fall 2022

Problem Set 3

() and STA130 Professors

Instructions

Complete the exercises in this .Rmd file and submit your .Rmd and .pdf output through Quercus on Thursday, September 29 by 5:00 p.m. ET.

Part 1: More Olympics Data

The code below loads the `VGAMdata` package (so you can access the data sets it contains) and the `tidyverse` package (so you can use the functions it contains) and glimpses the `oly12` data set, which you will use for this question. **Do not use the `olympics` data set from class to answer the prompts in this question**.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
#install.packages("VGAMdata")
library(VGAMdata)
```

```
## Loading required package: VGAM
## Loading required package: stats4
## Loading required package: splines
```

```
names(oly12) # convenient function to quickly glance at data set column names
```

```
##  [1] "Name"    "Country" "Age"     "Height"  "Weight"  "Sex"     "DOB"
##  [8] "PlaceOB" "Gold"    "Silver"  "Bronze"  "Total"   "Sport"   "Event"
```

```
glimpse(oly12)
```

```
## Rows: 10,384
## Columns: 14
## $ Name    <fct> Lamusi A, A G Kruger, Jamale Aarrass, Abdelhak Aatakni, Maria ~
## $ Country <fct> "People's Republic of China", "United States of America", "Fra~
## $ Age     <int> 23, 33, 30, 24, 26, 27, 30, 23, 27, 19, 37, 28, 28, 28, 22, 19~
## $ Height  <dbl> 1.70, 1.93, 1.87, NA, 1.78, 1.82, 1.82, 1.87, 1.90, 1.70, NA, ~
## $ Weight  <int> 60, 125, 76, NA, 85, 80, 73, 75, 80, NA, NA, NA, 60, 64, 62, N~
## $ Sex     <fct> M, M, M, M, F, M, F, M, M, M, M, M, F, F, M, F, M, M, M, M, F,~
```

```
## $ DOB     <date> 1989-02-06, NA, NA, 1988-09-02, NA, 1984-06-09, NA, 1989-03-0~
## $ PlaceOB <fct> "NEIMONGGOL (CHN)", "Sheldon (USA)", "BEZONS (FRA)", "AIN SEBA~
## $ Gold    <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ Silver  <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ Bronze  <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ Total   <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ Sport   <fct> "Judo", "Athletics", "Athletics", "Boxing", "Athletics", "Hand~
## $ Event   <fct> "Men's -60kg", "Men's Hammer Throw", "Men's 1500m", "Men's Lig~
```

Question 1: Practice with `filter()`

(a) In this week's class, we looked at data for each country which participated in the 2012 Olympics (e.g. size of each country's Olympic team, number of medals won, etc.), and there was one observation (i.e. one row) for each participating country. What does each row in the `oly12` dataset represent?

Each row in this instance represents the induvidual varibles i.e Name, Country,Age,Height,Weight,DOB, PlaceOB,Gold,Silver,Bronze,Total,Sport ,and Event

Hint: Type `?oly12` or `help(oly12)` in the console (on the bottom left corner) to view the help file for the `oly12` dataset in the Help tab (on the bottom right corner) of RStudio); or, just search for "oly12" in the Help tab. #Personal use code

```
# I checked for NA values which I thought would affect my code ,however upon asking a TA i realised tha

#na.omit(oly12) %>% filter(Gold==1) %>% print(4)
```

#Formating the pdf

(b) Determine the number of athletes who represented Canada (`Canada`) or the United States (`United States of America`) in the 2012 Olympic Games.

```
#oly12 %>% summarise(n=n())   check total number
oly12 %>% filter( 'Canada' == Country| 'United States of America' == Country) %>% summarise(n=n())
```

```
##     n
## 1 792
```

```
 oly12  %>%filter( 'Canada' == Country| 'United States of America' == Country) %>% summarise(n=n())
```

```
##     n
## 1 792
```

Hint: Apply the `filter()` function to the `Country` column of the `oly12` dataset

(c) Determine the number of athletes who competed in classical gymnastics (`Gymnastics - Artistic` and `Gymnastics - Rhythmic`) or classical pool sports (`Diving` and `Swimming`).

```
oly12 %>% filter( 'Canada' == Country| 'United States of America' == Country) %>%  summarise(n=n())
```

```
##     n
## 1 792
```

```
#levels(oly12$Sport)
```

Hint: You can see all the possible values for the `Sport` variable with `levels(oly12$Sport)`, and count the number of possible levels with `nlevels(oly12$Sport)`.

(d) Determine the number of athletes who competed in ANY gymnastic (`Gymnastics - Artistic`, `Gymnastics - Rhythmic`, `Trampoline`) or ANY pool sports (`Diving`, `Swimming`, `Synchronised Swimming`, and `Water Polo`)

```
oly12 %>% filter( 'Gymnastics - Artistic' == Sport| 'Gymnastics - Rhythmic' == Sport |'Trampoline' == Sp
```

```
##      n
## 1 1695
```

Hint: As indicated on stackoverflow, the `%in%` comparision operator could be useful here with `allGymnastics <- c("Gymnastics - Artistic", "Gymnastics - Rhythmic", "Trampoline")` and `allWaterPool <- c("Diving", "Swimming", "Synchronised Swimming", "Water Polo")` and `filter(Sport %in% allGymnastics | Sport %in% allWaterPool)`.

(e) Create the data subset `oly12_FemaleArtisticRhythmicGymnasts` which contains all female olympic athletes who competed in artistic gymnastics or rhythmic gymnastics.

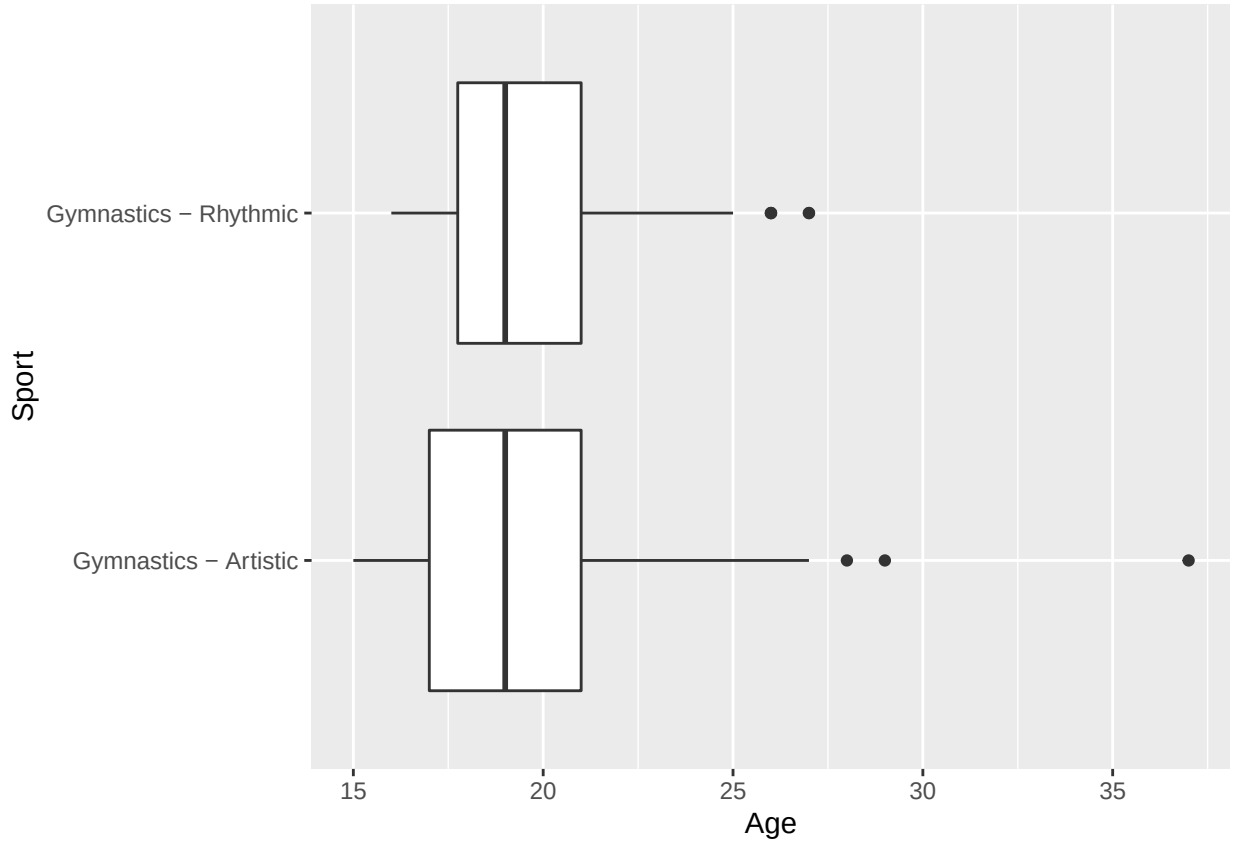
```
oly12 %>% filter( 'Gymnastics - Artistic' == Sport| 'Gymnastics - Rhythmic' == Sport |'Trampoline' == Sp
```

```
##      n
## 1 1695
```

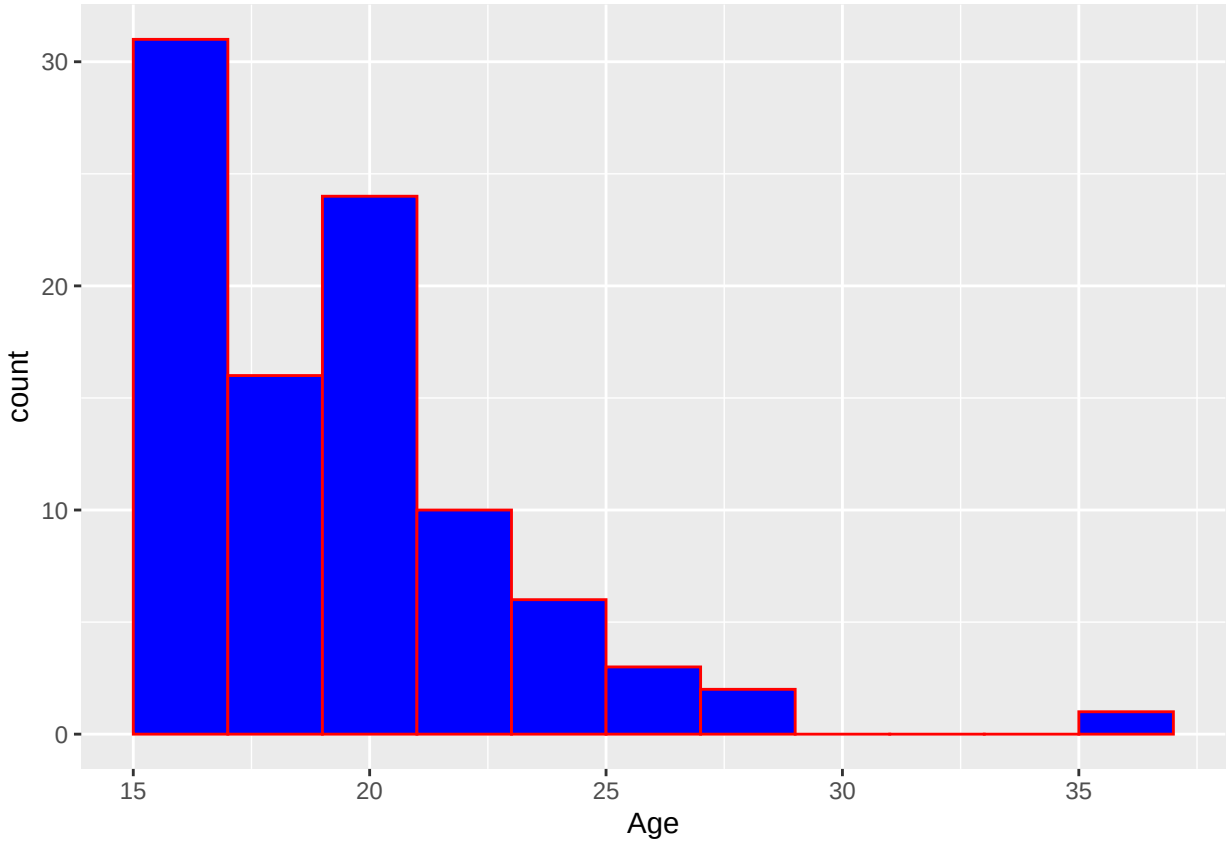
Hint: `names(oly12)` shows all the column names of the data set.

(f) Use `oly12_FemaleArtisticRhythmicGymnasts` and `ggplot2` to compare the age distribution of female olympic athletes competing in artistic gymnastics to the age distribution of female olympic athletes competing in rhythmic gymnastics using both boxplots and histrograms.

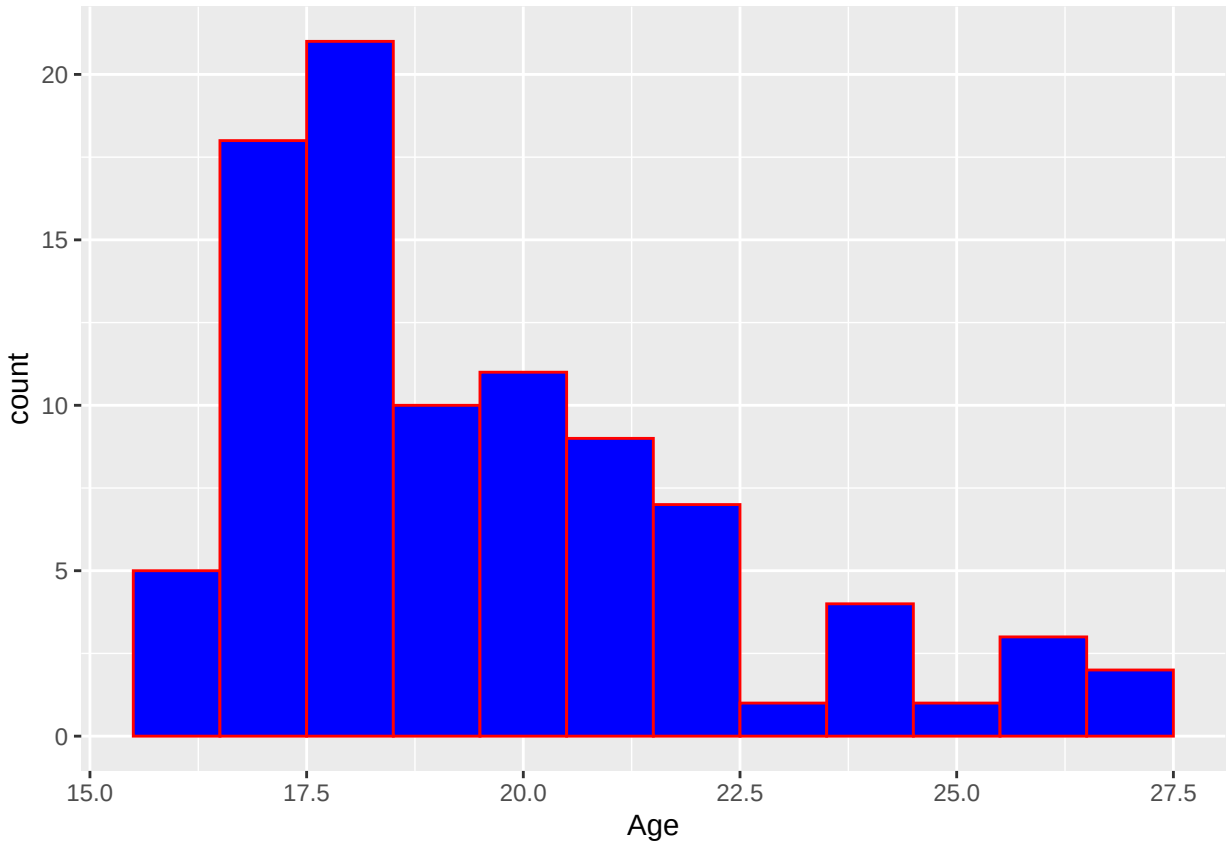
```
oly12 %>% filter(Sex=='F') %>% filter( 'Gymnastics - Artistic' == Sport| 'Gymnastics - Rhythmic' == Spo


oly12_FemaleArtisticRhythmicGymnasts  %>% ggplot(aes(x=Age,y=Sport))+geom_boxplot()
```




```
oly12_FemaleArtisticRhythmicGymnasts %>% filter( Sport=='Gymnastics - Artistic')%>% ggplot(aes(x=Age))+
```




```
oly12_FemaleArtisticRhythmicGymnasts %>% filter( Sport=='Gymnastics - Rhythmic') %>% ggplot(aes(x=Age))+
```




```
 #Rough Work
#oly12_FemaleArtisticRhythmicGymnasts %>% filter( Sport=='Gymnastics -  Rhythmic') %>% ggplot(aes(x=Age

#oly12_FemaleArtisticRhythmicGymnasts %>%  ggplot(aes(x=Sport,y=Age))+geom_histogram(stat="identity",bi



#oly12_FemaleArtisticRhythmicGymnasts%>% mutate(bool_return = case_when(
#Sport == 'Gymnastics - Artistic' ~ 1,
#Sport=='Gymnastics - Rhythmic' ~ 0
#), valueMy=bool_return) -> onlynew
 #onlynew%>% ggplot(aes(x=Age,y=bool_return))+geom_histogram(stat="identity")




# oly12_FemaleArtisticRhythmicGymnasts%>% mutate(bool_return = case_when(
#Sport == 'Gymnastics - Artistic' ~ 0,
#Sport=='Gymnastics - Rhythmic' ~ 1
#), valueMy=bool_return) -> onlynew
 #onlynew%>% ggplot(aes(x=Age))+geom_histogram(bins=5)+facet_wrap(~bool_return,ncol=1)
```

```
 #onlynew%>% filter(valueMy==1)%>% ggplot(aes(x=Age))+geom_histogram(bins=5)
#oly12_FemaleArtisticRhythmicGymnasts %>% mutate(bool_return = case_when(
#Sport == 'Gymnastics - Artistic' ~ 2,
#Sport == 'Gymnastics - Rhythmic' ~ 1,
#), valueMy=bool_return) -> onlynew
# onlynew%>% ggplot(aes(x=Age,y=bool_return))+geom_histogram(stat="identity")
```

Hint: don't forget `aes()` and to use + rather than %>%.

(g) Answer the following questions based on the plots you created in (d).

- Are the age distributions of female rhythmic gymnasts and female artistic gymnasts symmetrical or skewed?

Female artistic is right skewed and female rhythmic is symmetrical

- How do the medians, 25th percentiles, and 75th percentiles for ages of female rhythmic gymnasts and female artistic gymnasts compare?

In female rhythmic gymnast the 25th percentile is much higher than that of artistic gymnast while the 75th percentile are the same, The mediam is also the same. It is possible that artistic had more data to make the data look different

- Based only on the histogram and boxplots, predict whether the standard deviation of the ages is similar or different. Justify your answer in 1-2 sentences. *Come back to this Due to the difference in spread the standard deviation may be different. Gtmnastic- Artistic has a higher spread so it may have a higher standard deviation than gymnastic Artistic. The same observation can also be made via the histogram representation.*

Question 2: Practice with `summarise()`, `group_by()`, and `mutate()`

(a) Create a summary table of `oly12_FemaleArtisticRhythmicGymnasts` reporting the minimum (`min`), maximum (`min`), `mean`, `median`, and standard deviation (`sd`) of ages for female rhythmic gymnasts and female artistic gymnasts. Were you correct in your guess about the standard deviation in part (g) of the last question?

```
# Type your code here
oly12_FemaleArtisticRhythmicGymnasts %>% filter(Sport=='Gymnastics - Artistic') %>% summarise(mean_farg
```

```
##   mean_farg median_farg  farg_sd
## 1  19.73118          19 3.659828
```

```
oly12_FemaleArtisticRhythmicGymnasts %>% filter(Sport=='Gymnastics - Rhythmic') %>% summarise(mean_farg
```

```
##   mean_farg median_farg  farg_sd
## 1  19.48913          19 2.678751
```

My calim in part g was correct because both of the standard deviations are different as seen in the data above in the tibble

(b) Create a new variable called `total_medals` and create a new tibble called `oly12_OneMedalClub` that contains athletes who won exactly one medal at the 2012 olympics.

```
# Type your code here
oly12 %>% select(Name, Gold, Silver, Bronze) %>%
mutate(total_medals = Gold+Silver+Bronze) %>% filter(total_medals==1) -> oly12_OneMedalClub
oly12_OneMedalClub
```

```
##                              Name Gold Silver Bronze total_medals
## 1                   Jennifer Abel    0      0      1            1
## 2            Alaaeldin Abouelkassem    0      1      0            1
## 3               Chantal Achterberg    0      0      1            1
## 4                    Filip Adamski    1      0      0            1
## 5                 Rebecca Adlington    0      0      1            1
## 6                 Kseniia Afanaseva    0      1      0            1
## 7                      Mayra Aguiar    0      0      1            1
## 8                    Hannes Aigner    0      0      1            1
## 9                  Nasser Al-Attiya    0      0      1            1
## 10          Sara Algotsson Ostholt    0      1      0            1
## 11                    Yuri Alvear    0      0      1            1
## 12                      Kum Ae An    1      0      0            1
## 13                  Alyssa Anderson    1      0      0            1
## 14               Kristin Armstrong    1      0      0            1
## 15                    Judith Arndt    0      1      0            1
## 16                  Mariana Avitia    0      0      1            1
## 17              Alexander Kristoff    0      0      1            1
## 18                   Alina Dumitru    0      1      0            1
## 19                     Carina BAER    0      1      0            1
## 20                     Tim Baillie    1      0      0            1
## 21                  Angie Bainbridge    0      1      0            1
## 22                     Coralie Balmy    0      0      1            1
## 23                     Jacob Barsoe    0      0      1            1
## 24                  Danka Bartekova    0      0      1            1
## 25                     Chris Bartley    0      1      0            1
## 26                  Romano Battisti    0      1      0            1
## 27                     Gregory Bauge    0      1      0            1
## 28                  Elizabeth Beisel    0      1      0            1
## 29                Claudia Belderbos    0      0      1            1
## 30           Mireia Belmonte Garcia    0      1      0            1
## 31                Meaghan Benfeito    0      0      1            1
## 32                  Gabriel Bergen    0      1      0            1
## 33              Yanet Bermoy Acosta    0      1      0            1
## 34                    Alain Bernard    1      0      0            1
## 35                     Ole Bischof    0      1      0            1
## 36                   Sylwia Bogacka    0      1      0            1
## 37                     Hamish Bond    1      0      0            1
## 38                Charlotte Bonnet    0      0      1            1
## 39                     Edith Bosch    0      0      1            1
## 40                    David Boudia    0      0      1            1
## 41                    Carline Bouw    0      0      1            1
## 42                  Matthew Brittain    1      0      0            1
## 43                  Jeremiah Brown    0      1      0            1
## 44                    Kelci Bryant    0      1      0            1
## 45                  Karina Bryant    0      0      1            1
## 46                Ashley Brzozowicz    0      1      0            1
## 47              Diana Laura Bulimar    0      0      1            1
## 48                    Andrew Byrnes    0      1      0            1
## 49                    Brent Hayden    0      0      1            1
## 50                 Britta Heidemann    0      1      0            1
## 51                     Erin Cafaro    1      0      0            1
## 52            Ivan Cambar Rodriguez    0      0      1            1
## 53                    Alan Campbell    0      0      1            1
```

```
## 54                  Cate Campbell    1      0      0            1
## 55               Niccolo Campriani    0      1      0            1
## 56                        Yuan Cao    1      0      0            1
## 57              Corina Caprioriu      0      1      0            1
## 58                Peter Chambers      0      1      0            1
## 59               Richard Chambers     0      1      0            1
## 60                Germain Chardin     0      1      0            1
## 61             Diana Maria Chelaru    0      0      1            1
## 62                    Ruolin Chen     1      0      0            1
## 63                    Yibing Chen     1      0      0            1
## 64                      Ying Chen     0      1      0            1
## 65                     Ming Cheng     0      1      0            1
## 66               Zulfiya Chinshanlo    1      0      0            1
## 67                    Jun-Ho Cho      0      0      1            1
## 68                 Byungchul Choi     0      0      1            1
## 69                   Hyeonju Choi     1      0      0            1
## 70              Maialen Chourraut     0      0      1            1
## 71                    Tyler Clary     1      0      0            1
## 72            Roxana Daniela Cocos    0      1      0            1
## 73                    Nathan Cohen    1      0      0            1
## 74                  Kristina Cook     0      1      0            1
## 75                      Iztok Cop     0      0      1            1
## 76                Natalie Coughlin    0      0      1            1
## 77                     Lionel Cox     0      1      0            1
## 78                  Will Crothers     0      1      0            1
## 79                       Kim Crow     0      1      0            1
## 80                    Laszlo Cseh     0      0      1            1
## 81                    Caryn Davies    1      0      0            1
## 82                  Catalina Ponor    0      0      1            1
## 83                  Charlie Houchin   1      0      0            1
## 84               Michael D'almeida    0      1      0            1
## 85                    Hakan Dahlby     0      1      0            1
## 86                         Jun Dai     0      0      1            1
## 87                   Xiaoxiang Dai     0      0      1            1
## 88                  Sytske de Groot    0      0      1            1
## 89                 Annemiek de Haan    0      0      1            1
## 90                  Rajmond Debevec    0      0      1            1
## 91                    Lucie Decosse    1      0      0            1
## 92                     Natalie Dell    0      0      1            1
## 93                  Yana Dementieva    1      0      0            1
## 94                  Inna Deriglazova    0      1      0            1
## 95                        Feng Ding    0      0      1            1
## 96                        Ning Ding    0      1      0            1
## 97                        Dong Dong    1      0      0            1
## 98               Nataliya Dovgodko     1      0      0            1
## 99                     Conor Dwyer     1      0      0            1
## 100                  Douglas Csima     0      1      0            1
## 101                 Iuliia Efimova     0      0      1            1
## 102               Richard Egington     0      0      1            1
## 103                Irawan Eko Yuli     0      0      1            1
## 104                  Brady Ellison     0      1      0            1
## 105                   Rene Enders      0      0      1            1
## 106          Paola Espinosa Sanchez    0      1      0            1
## 107                  Tony Estanguet    1      0      0            1
```

```
## 108        Ophelie-Cyrielle Etienne    0      0      1            1
## 109                     Blair Evans    0      1      0            1
## 110                  Eskild Ebbesen    0      0      1            1
## 111                Eva Csernoviczki    0      0      1            1
## 112                     Yuting Fang    0      1      0            1
## 113                  Margaux Farrell    0      0      1            1
## 114                     James Feigen    0      1      0            1
## 115                     Tianwei Feng    0      0      1            1
## 116                         Zhe Feng    1      0      0            1
## 117                      Emilie Fer     1      0      0            1
## 118                   Sergei Fesikov    0      0      1            1
## 119 Oscar Albeiro Figueroa Mosquera    0      1      0            1
## 120                   Roseline Filion    0      0      1            1
## 121                   Joachim Fischer    0      0      1            1
## 122                   David Florence     0      1      0            1
## 123                       James Foad     0      0      1            1
## 124               Robert Forstemann      0      0      1            1
## 125              Karsten Forsterling     0      0      1            1
## 126                      Jessica Fox     0      1      0            1
## 127                 William Fox-Pitt     0      1      0            1
## 128                Zsuzsanna Francia     1      0      0            1
## 129                Michele Frangilli     1      0      0            1
## 130              Christopher Froome      0      0      1            1
## 131            Kamilla Gafurzianova      0      1      0            1
## 132                  Marco Galiazzo      1      0      0            1
## 133                  Arsen Galstyan      1      0      0            1
## 134              Ivan Garcia Navarro     0      1      0            1
## 135                    Gemma Gibbons     0      1      0            1
## 136                       Rob Gibson     0      1      0            1
## 137                     Fabien Gilot     1      0      0            1
## 138                  Christine Girard    0      0      1            1
## 139                      Helen Glover    1      0      0            1
## 140                  Priscilla Gneto     0      0      1            1
## 141                Celine Goberville     0      1      0            1
## 142                   Anders Golding     0      1      0            1
## 143                      Jinjie Gong     0      1      0            1
## 144                   Asley Gonzalez     0      1      0            1
## 145               Katherine Grainger     1      0      0            1
## 146                   Andrey Grechin     0      0      1            1
## 147               Anastasia Grishina     0      1      0            1
## 148                    Tim Grohmann      1      0      0            1
## 149                       Henk Grol     0      0      1            1
## 150                   Krista Guloien     0      1      0            1
## 151                    Daniel Gyurta     1      0      0            1
## 152                    Gevrise Emane     0      0      1            1
## 153                    Kosuke Hagino     0      0      1            1
## 154                  Juliette Haigh      0      0      1            1
## 155                  Brendan Hansen      0      0      1            1
## 156                   Janine Hanson      0      1      0            1
## 157                         Yun Hao     0      0      1            1
## 158                   Jessica Hardy     0      0      1            1
## 159                  Kayla Harrison     1      0      0            1
## 160                           Zi He     1      0      0            1
## 161                 Femke Heemskerk     0      1      0            1
```

```
## 162         Aliaksandra Herasimenia    0      1      0            1
## 163                  Emilie Heymans    0      0      1            1
## 164                     Phelan Hill    0      0      1            1
## 165                   Philip Hindes    1      0      0            1
## 166                  Hiroaki Hiraoka    0      1      0            1
## 167               Pavol Hochschorner    0      0      1            1
## 168               Peter Hochschorner    0      0      1            1
## 169                   Malcolm Howard    0      1      0            1
## 170                        Chris Hoy    1      0      0            1
## 171                   Hee Sook Jeon    0      0      1            1
## 172                   Ilias Iliadis    0      0      1            1
## 173                   Dong Hyun Im     0      0      1            1
## 174         Larisa Andreea Iordache    0      0      1            1
## 175                   Cristina Iovu    0      0      1            1
## 176                  Kristian Ipsen    0      0      1            1
## 177                     Mansur Isaev    1      0      0            1
## 178            Sandra Raluca Izbasa    0      0      1            1
## 179                   Danila Izotov    0      0      1            1
## 180                     Inge Dekker    0      1      0            1
## 181                Michael Jamieson    0      1      0            1
## 182                     Haiqi Jiang    0      0      1            1
## 183                   Liuyang Jiao    1      0      0            1
## 184                     Jongoh Jin    1      0      0            1
## 185                  Eric Johannesen    1      0      0            1
## 186                Abigail Johnston    0      1      0            1
## 187                   Cullen Jones    0      1      0            1
## 188                Morten Jorgensen    0      0      1            1
## 189                     Gil Ok Jung    0      0      1            1
## 190                     Jinsun Jung    0      0      1            1
## 191                   Yuliya Kalina    0      0      1            1
## 192                   Megan Kalmoe    0      0      1            1
## 193                   Jake Kaminski    0      1      0            1
## 194                      Miki Kanie    0      0      1            1
## 195                     Ryohei Kato    0      1      0            1
## 196                  Kaori Kawanaka    0      0      1            1
## 197                      Jason Kenny    1      0      0            1
## 198                Tagir Khaibulaev    1      0      0            1
## 199                   Olga Kharlan    0      0      1            1
## 200                     Bubmin Kim    0      0      1            1
## 201                     Jae-Bum Kim    1      0      0            1
## 202                     Jangmi Kim    1      0      0            1
## 203                     Jiyeon Kim    1      0      0            1
## 204                     Un Guk Kim    1      0      0            1
## 205                      Mary King    0      1      0            1
## 206                   Nienke Kingma    0      0      1            1
## 207                  Felipe Kitadai    0      0      1            1
## 208                   Ingrid Klimke    1      0      0            1
## 209                     Kara Kohler    0      0      1            1
## 210                  Nikolay Kovalev    0      0      1            1
## 211           Anastasiia Kozhenkova    1      0      0            1
## 212                 Andreas Kuffner    1      0      0            1
## 213                     Yolane Kukla    1      0      0            1
## 214                   Iryna Kulesha    0      0      1            1
## 215                     Vijay Kumar    0      1      0            1
```

```
## 216                Evgeny Kuznetsov    0      1      0            1
## 217                    Kate Hornsey    0      1      0            1
## 218                  Evgeny Lagunov    0      0      1            1
## 219                   Mylene Lazare    0      0      1            1
## 220                   Chad le Clos    1      0      0            1
## 221                   Sung Jin Lee    1      0      0            1
## 222                    Ugo Legrand    0      0      1            1
## 223                      Sheng Lei    1      0      0            1
## 224                Caitlin Leverenz    0      0      1            1
## 225                 Maximilian Levy    0      0      1            1
## 226                    Danell Leyva    0      0      1            1
## 227                    Jason Lezak    0      1      0            1
## 228                    Xiaoxia Li    1      0      0            1
## 229                    Xuanxu Li    0      0      1            1
## 230                    Xueying Li    1      0      0            1
## 231                      Yunqi Li    0      0      1            1
## 232           Ruben Limardo Gascon    1      0      0            1
## 233                   Qingfeng Lin    1      0      0            1
## 234                  Caroline Lind    1      0      0            1
## 235                Nikita Lobintsev    0      0      1            1
## 236                  Esther Lofgren    1      0      0            1
## 237                  Eleanor Logan    1      0      0            1
## 238          Constantine Louloudis    0      0      1            1
## 239                    Chunlong Lu    0      0      1            1
## 240                      Haojie Lu    0      1      0            1
## 241                    Xiaojun Lu    1      0      0            1
## 242                       Ying Lu    0      1      0            1
## 243                      Zhiwu Lu    0      0      1            1
## 244                    Yutong Luo    1      0      0            1
## 245                         Jin Ma    0      1      0            1
## 246                James Magnussen    0      1      0            1
## 247                 Gregory Mallet    0      1      0            1
## 248                   Marti Malloy    0      0      1            1
## 249                   Maiya Maneza    1      0      0            1
## 250                Mc Kayla Maroney    1      0      0            1
## 251                 Darcy Marquardt    0      1      0            1
## 252               Adrienne Martelli    0      0      1            1
## 253                 Michal Martikan    0      0      1            1
## 254                    Damir Martin    0      1      0            1
## 255        Razvan Constantin Martin    0      0      1            1
## 256                    Tony Martin    0      1      0            1
## 257                 Sergei Martynov    1      0      0            1
## 258               Natalie Mastracci    0      1      0            1
## 259                 Takeshi Matsuda    0      0      1            1
## 260                 Kaori Matsumoto    1      0      0            1
## 261                   Conlin Mccabe    0      1      0            1
## 262                Nicholas Mccrory    0      0      1            1
## 263                Kaarle Mcculloch    0      0      1            1
## 264                  Matthew Mclean    1      0      0            1
## 265                     James Mcrae    0      0      1            1
## 266                     Anna Meares    0      0      1            1
## 267                  Ruta Meilutyte    1      0      0            1
## 268                   Sarah Menezes    1      0      0            1
## 269                Florian Mennigen    1      0      0            1
```

```
## 270                Julia Michalska    0      0      1            1
## 271             Alexander Mikhaylin    0      1      0            1
## 272                   Hiromi Miyake    0      1      0            1
## 273          Alin George Moldoveanu    1      0      0            1
## 274               Daniele Molmenti     1      0      0            1
## 275              Christopher Morgan    0      0      1            1
## 276                 Andreanne Morin    0      1      0            1
## 277               Vladimir Morozov     0      0      1            1
## 278              Dorian Mortelette     0      1      0            1
## 279                   Vasily Mosin    0      0      1            1
## 280                 Lukas Mueller     1      0      0            1
## 281                   Eric Murray     1      0      0            1
## 282                Meghan Musnicki    1      0      0            1
## 283             Magdalena Fularczyk    0      0      1            1
## 284                   Mahe Drysdale    1      0      0            1
## 285                Masashi Ebinuma    0      0      1            1
## 286              Matthew Langridge    0      0      1            1
## 287             Tuvshinbayar Naidan    0      1      0            1
## 288                   Riki Nakaya    0      1      0            1
## 289                  Hyun Hee Nam    0      0      1            1
## 290                  Gagan Narang    0      0      1            1
## 291                   George Nash    0      0      1            1
## 292                  Sizwe Ndlovu    1      0      0            1
## 293                      Lia Neal    0      0      1            1
## 294                  Jade Neilsen    0      1      0            1
## 295                 Mauro Nespoli    1      0      0            1
## 296                 Marcel Nguyen    0      1      0            1
## 297               Andrew Nicholson    0      0      1            1
## 298                  Ivan Nifontov    0      0      1            1
## 299              Masashi Nishiyama    0      0      1            1
## 300                  Daniel Noonan    0      0      1            1
## 301                  Natsumi Hoshi    0      0      1            1
## 302                Diego Occhiuzzi    0      1      0            1
## 303                       Ha Na Oh    0      0      1            1
## 304                    Sam Oldham     0      0      1            1
## 305                   Yun Chol Om     1      0      0            1
## 306                  Britta Oppelt    0      1      0            1
## 307           Alejandra Orozco Loza    0      1      0            1
## 308                   Idalys Ortiz    1      0      0            1
## 309              Dimitrij Ovtcharov    0      0      1            1
## 310                  Jonathan Paget    0      0      1            1
## 311                    Kylie Palmer    0      1      0            1
## 312                  Alex Partridge    0      0      1            1
## 313                    Maria Paseka    0      1      0            1
## 314                  Automne Pavia     0      0      1            1
## 315             Christinna Pedersen    0      0      1            1
## 316                  Lauren Perdue     1      0      0            1
## 317                  Thiago Pereira    0      1      0            1
## 318                  Dimitri Peters    0      0      1            1
## 319                   Zara Phillips    0      1      0            1
## 320                Bartosz Piasecki    0      1      0            1
## 321                 Caroline Powell    0      0      1            1
## 322                  Brooke Pratley    0      1      0            1
## 323                     Brian Price    0      1      0            1
```

```
## 324                   Leuris Pupo    1      0      0            1
## 325                 Daniel Purvis    0      0      1            1
## 326                       Kai Qin    1      0      0            1
## 327                   Tom Ransley    0      0      1            1
## 328            Maximilian Reinelt    1      0      0            1
## 329      Roline Repelaer van Driel    0      0      1            1
## 330                Kimberly Rhode    1      0      0            1
## 331              Jonelle Richards    0      0      1            1
## 332                 Julia Richter    0      1      0            1
## 333                  Jong Sim Rim    1      0      0            1
## 334                   Teddy Riner    1      0      0            1
## 335                  Taylor Ritzel    1      0      0            1
## 336                     Aida Roman    0      1      0            1
## 337                     Kyla Ross    1      0      0            1
## 338                     Marc Ryan    0      0      1            1
## 339                Chun Hwa Ryang    0      0      1            1
## 340                   Ren Hayakawa    0      0      1            1
## 341              Richard Hounslow    0      1      0            1
## 342             Rosalba Forciniti    0      0      1            1
## 343                     David Sain    0      1      0            1
## 344          Nyam-Ochir Sainjargal    0      0      1            1
## 345               Ilaria Salvatori    1      0      0            1
## 346         German Sanchez Sanchez    0      1      0            1
## 347                Alessio Sartori    0      1      0            1
## 348                  William Satch    0      0      1            1
## 349                   Martin Sauer    1      0      0            1
## 350                  Mohamed Sbihi    0      0      1            1
## 351              Anne Schellekens    0      0      1            1
## 352                Richard Schmidt    1      0      0            1
## 353                 Lauritz Schoof    1      0      0            1
## 354                   Dirk Schrade    1      0      0            1
## 355            Hinkelien Schreuder    0      1      0            1
## 356                   Karl Schulze    1      0      0            1
## 357                  Rebecca Scown    0      0      1            1
## 358                    Greg Searle    0      0      1            1
## 359                  Jesse Sergent    0      0      1            1
## 360                  Aida Shanaeva    0      1      0            1
## 361          Lasha Shavdatuashvili    1      0      0            1
## 362            Maryna Shkermankova    0      0      1            1
## 363                   Rafael Silva    0      0      1            1
## 364                Martin Sinkovic    0      1      0            1
## 365                Valent Sinkovic    0      1      0            1
## 366                   Kevin Sireau    0      1      0            1
## 367                Pimsiri Sirikaew    0      1      0            1
## 368                     John Smith    1      0      0            1
## 369                    Louis Smith    0      0      1            1
## 370                   Dae-Nam Song    1      0      0            1
## 371                      Luka Spik    0      0      1            1
## 372             Christian Sprenger    0      1      0            1
## 373               Heather Stanning    1      0      0            1
## 374                  Etienne Stott    1      0      0            1
## 375                  Mika Sugimoto    0      1      0            1
## 376                Joseph Sullivan    1      0      0            1
## 377                     Yujie Sun    0      0      1            1
```

```
## 378                  Ondrej Synek    0      1      0            1
## 379                 Shu-Ching Hsu    0      1      0            1
## 380            Svetlana Podobedova    1      0      0            1
## 381                   Sarah Tait    0      1      0            1
## 382                Kazuhito Tanaka    0      1      0            1
## 383                  Yusuke Tanaka    0      1      0            1
## 384                        Yi Tang    0      0      1            1
## 385             Kateryna Tarasenko    1      0      0            1
## 386                 Davis Tarwater    1      0      0            1
## 387               Sideris Tasiadis    0      1      0            1
## 388                  Ryo Tateishi    0      0      1            1
## 389               Audrey Tcheumeo    0      0      1            1
## 390                  Aya Terakawa    0      0      1            1
## 391                 Luca Tesconi    0      1      0            1
## 392              Annekatrin Thiele    0      1      0            1
## 393                 Kerstin Thiele    0      1      0            1
## 394                   Nick Thoman    0      1      0            1
## 395               Kristian Thomas    0      0      1            1
## 396                 James Thompson    1      0      0            1
## 397          Lesley Thompson-Willie    0      1      0            1
## 398                  Peter Thomsen    1      0      0            1
## 399                     Mark Todd    0      0      1            1
## 400                Andreas Toelzer    0      0      1            1
## 401                 Libby Trickett    1      0      0            1
## 402             Svetlana Tsarukaeva    0      1      0            1
## 403              Takaharu Furukawa    0      1      0            1
## 404                   Troy Dumais    0      0      1            1
## 405                 Miklos Ungvari    0      1      0            1
## 406             Rigoberto Uran Uran    0      1      0            1
## 407                 Dmitry Ushakov    0      1      0            1
## 408          Antoine Valois-Fortier    0      0      1            1
## 409           Cameron van der Burgh    1      0      0            1
## 410              Charline van Snick    0      0      1            1
## 411                Peter Vanderkaay    0      0      1            1
## 412               Jacobine Veenhoven    0      0      1            1
## 413                Marleen Veldhuis    0      1      0            1
## 414                   Sofya Velikaya    0      1      0            1
## 415                Rachelle Viinberg    0      1      0            1
## 416               Alexandr Vinokurov    1      0      0            1
## 417                   Kristina Vogel    1      0      0            1
## 418                    Marianne Vos    1      0      0            1
## 419                Shannon Vreeland    1      0      0            1
## 420                Valentin Hristov    0      0      1            1
## 421               Vavrinec Hradilek    0      1      0            1
## 422                Vincent Hancock    1      0      0            1
## 423                       Hao Wang    1      0      0            1
## 424                       Hao Wang    0      1      0            1
## 425                   Mingjuan Wang    1      0      0            1
## 426                    Anna Watkins    1      0      0            1
## 427                       Ning Wei    0      1      0            1
## 428                    Amanda Weir    0      0      1            1
## 429                    Miriam Welte    1      0      0            1
## 430                  Phillipp Wende    1      0      0            1
## 431                    Mary Whipple    1      0      0            1
```

```
## 432                    Max Whitlock    0      0      1            1
## 433                   Jordyn Wieber    1      0      0            1
## 434                 Bradley Wiggins    1      0      0            1
## 435                   Kristof Wilke    1      0      0            1
## 436                Lauren Wilkinson    0      1      0            1
## 437                    Rob Williams    0      1      0            1
## 438                   Nicola Wilson    0      1      0            1
## 439     Peter Robert Russell Wilson    1      0      0            1
## 440                  Kasper Winther    0      0      1            1
## 441                     Jingbiao Wu    0      1      0            1
## 442                       Minxia Wu    1      0      0            1
## 443                     Jacob Wukie    0      1      0            1
## 444                         Chen Xu    0      1      0            1
## 445                         Jing Xu    0      1      0            1
## 446                         Lili Xu    0      1      0            1
## 447                   Koji Yamamuro    0      1      0            1
## 448                      Siling Yi    1      0      0            1
## 449                         Dan Yu    0      0      1            1
## 450             Natalya Zabolotnaya    0      1      0            1
## 451                   Ilya Zakharov    0      1      0            1
## 452                 Chenglong Zhang    1      0      0            1
## 453                      Jike Zhang    1      0      0            1
## 454                       Nan Zhang    1      0      0            1
## 455                  Andrija Zlatic    0      0      1            1
## 456                    Urska Zolnir    1      0      0            1
## 457                         Kai Zou    1      0      0            1
```

(c) Uncomment the code below and run the glimpse of the data created in part (c).

```
glimpse(oly12_OneMedalClub)
```

```
## Rows: 457
## Columns: 5
## $ Name         <fct> Jennifer Abel, Alaaeldin Abouelkassem, Chantal Achterberg~
## $ Gold         <int> 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, ~
## $ Silver       <int> 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, ~
## $ Bronze       <int> 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, ~
## $ total_medals <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
```

Question 3: Practice with `select()`, `arrange()`, `desc()`, and `filter()`

(b) Find the `Name` and `Age` of the 6 oldest athletes who competed in the 2012 Olympics.

```
oly12 %>% select(Name,Age)%>% arrange(desc(Age)) -> newArrangedTibble
head(newArrangedTibble, 6)
```

```
##                Name Age
## 1   Hiroshi Hoketsu  71
## 2 Afanasijs Kuzmins  65
## 3        Ian Millar  65
## 4    Carl Bouckaert  58
## 5  Andrei Kavalenka  57
## 6        Mary Hanna  57
```

(b) Find the `Name`, `Age` and `Sport` of the 6 youngest female athletes who competed in the 2012 Olympics.

```
oly12 %>% select(Name,Age)%>% arrange(Age) -> newArrangedTibble
head(newArrangedTibble, 6)
```

```
##                       Name Age
## 1              Adzo Kpossi  13
## 2         Aurelie Fanchette  14
## 3                 Suji Kim  14
## 4 Nafissatou Moussa Adamou  14
## 5  Lea Melissa Moutoussamy  14
## 6                Yuhan Qiu  14
```

(c) Find the `Name`, `Age`, `Sport`, and `Event` for the 6 youngest and 6 oldest competitors who won gold medals at the 2012 olympics. [This can be run as two pieces of code rather than one piece of combined code].

```
oly12 %>% select(Name,Age,Sport,Event)%>% arrange(Age) -> newArrangedTibble
head(newArrangedTibble, 6)
```

```
##                       Name Age    Sport                          Event
## 1              Adzo Kpossi  13 Swimming          Women's 50m Freestyle
## 2         Aurelie Fanchette  14 Swimming         Women's 200m Freestyle
## 3                 Suji Kim  14   Diving           Women's 10m Platform
## 4 Nafissatou Moussa Adamou  14 Swimming          Women's 50m Freestyle
## 5  Lea Melissa Moutoussamy  14  Fencing       Women's Individual Sabre
## 6                Yuhan Qiu  14 Swimming Women's 4x100m Freestyle Relay
```

```
oly12 %>% select(Name,Age,Sport,Event)%>% arrange(desc(Age)) -> newArrangedTibble
head(newArrangedTibble, 6)
```

```
##                Name Age      Sport                                      Event
## 1   Hiroshi Hoketsu  71 Equestrian                 Individual Dressage, WHISPER
## 2 Afanasijs Kuzmins  65   Shooting                  Men's 25m Rapid Fire Pistol
## 3        Ian Millar  65 Equestrian Individual Jumping, Team Jumping, STAR POWER
## 4    Carl Bouckaert  58 Equestrian Individual Eventing, Team Eventing, CYRANO Z
## 5  Andrei Kavalenka  57   Shooting                                   Men's Trap
## 6        Mary Hanna  57 Equestrian Individual Dressage, Team Dressage, SANCETTE
```

Question 4: The Data Consultant

You have just been hired by a consultancy company. Congratulations! They are doing a report on each Olympics for the past 10 years. Given your recent experience in STA130, you ask to be responsible for the 2012 summary. Write a short report to your boss on information that can be gleaned about the ages of the athletes across sports. As it turns out, you happen to know that your new boss' favourite sports are badminton and weightlifting, so addressing these sports specifically might be an easy way to capture their attention; but, other features athletes' ages which can be learned from your plots and tables will of course be appreciated, too. The more interesting the better!

Question Constraints This is a quick report for your boss, so use full sentences and communicate in a clear and professional manner. Grammar isn't the main focus of the assessment, but don't use slang or emojis.

- ***Avoid Analysis Paralysis***: this is envisioned as a 30 minute exercise, so you don't have time to exhaustively explore every aspect of the data set.

- ***Avoid Writer's Block***: this is envisioned as a 200-400 word exercise, so quickly find something you can communicate and write about.

(a) Watch this 7-minute video introduction to hedging.

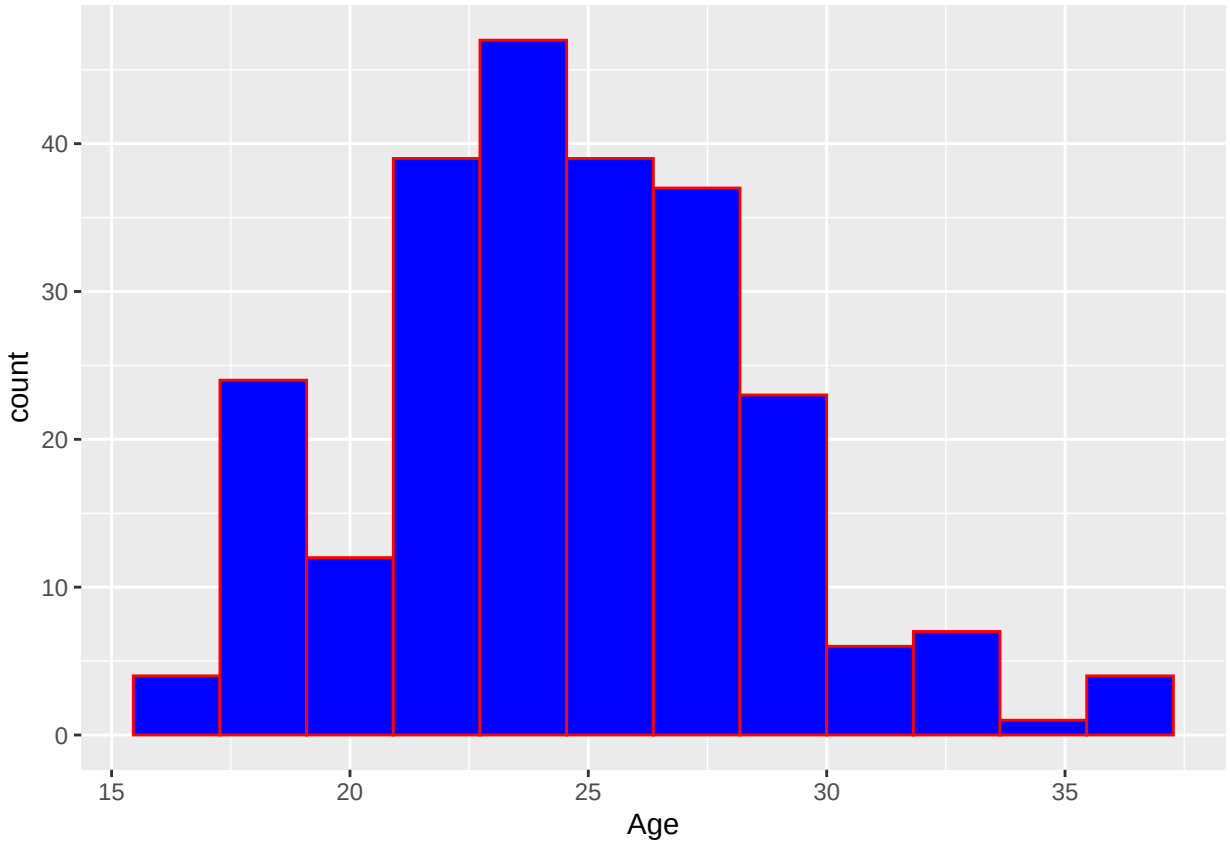
Hedging is helpful whenever you can't say something is 100% one way or another, as is often the case. In statistics, hedging should always be used with respect to the limitations of data and the strength and generalizability of the conclusions.

(b) Provide a small introduction of one or two sentences to draw your reader in and then explain what you'll be discussing. Be definitive about what your data is, and use *hedging* to caveat the limitations of the data.

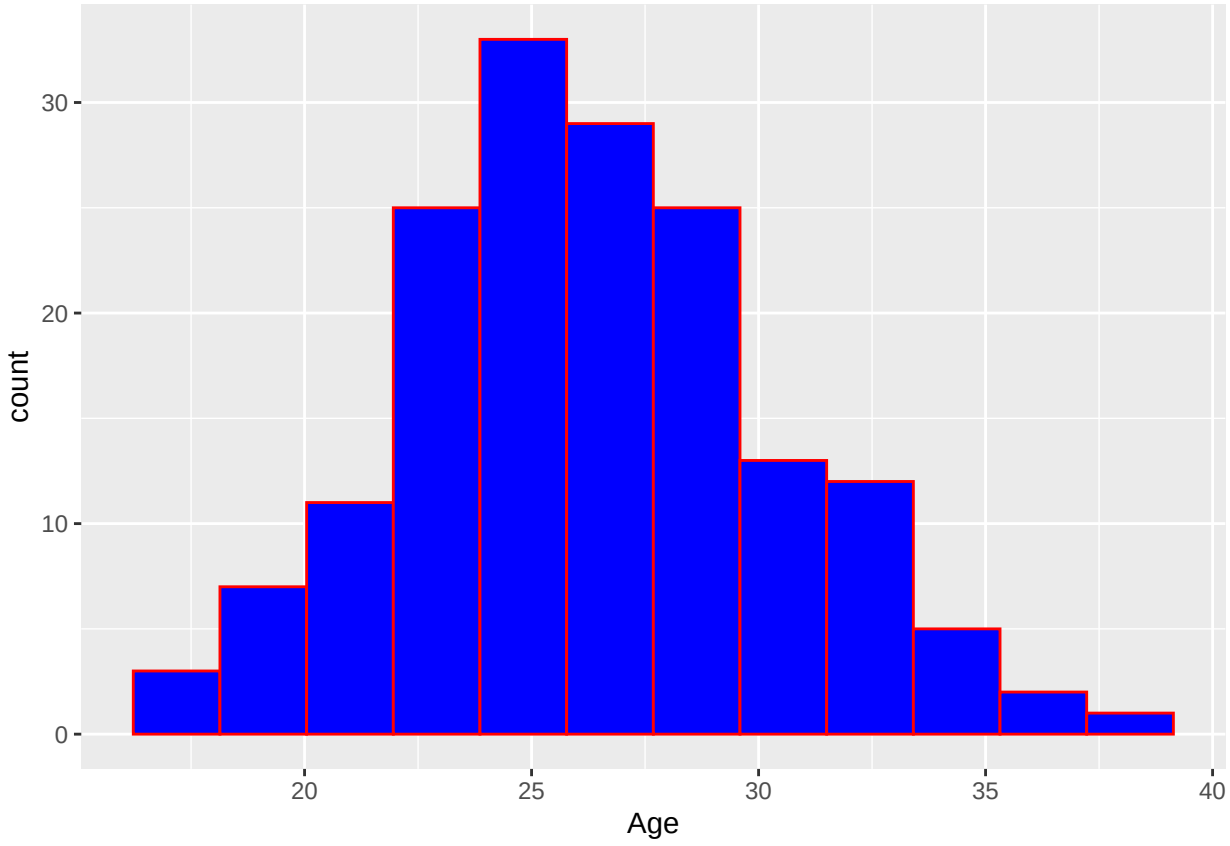
Answer: The data collected has been in regards to a number of athletes and their ages.The relationships may represent a key relationship ,however it should be noted that other factors were ignored while the two were compared. This was because a direct relationship was the area of interest.

(c) Provide one or two clearly titled and labeled figures addressing interesting features of athletes' ages.

```
oly12 %>% filter( Sport=="Weightlifting") %>% ggplot(aes(x=Age))+geom_histogram(bins=12,color="red",fil
```




```
oly12 %>% filter( Sport=="Badminton") %>% ggplot(aes(x=Age))+geom_histogram(bins=12,color="red",fill="bl
```



WRITE HERE

(d) Provide one or two clearly labeled summary tables addressing interesting features of athletes' ages.

```
oly12 %>% filter(Sport=="Weightlifting") %>% summarise(mean_farg=mean(Age),median_farg=median(Age),farg
```

```
##   mean_farg median_farg  farg_sd
## 1  24.59671          24 4.064804
```

```
oly12 %>% filter(Sport=="Badminton") %>% summarise(mean_farg=mean(Age),median_farg=median(Age),farg_sd=
```

```
##   mean_farg median_farg  farg_sd
## 1  26.15663          26 4.123787
```

(e) Watch this 8-minute video introduction to plagiarism.

You don't need to cite any outside references for your report to your boss, but you will be referring to your own created figures and tables. We'll use this as an excuse to get started early thinking about this important topic, and also use it as an exercise to start getting into the right referencing habits. It's easy and natural and makes your writing better (not mention avoids potential serious academic integrity violations...)

(f) Describe the interesting features of athletes' ages that you've found, referencing the figures and summary tables created in (c) and (d) just above. Use at least two of the vocabulary words listed below; but, your boss isn't a statistician, so make sure to clearly define and explain the vocabulary you use.

Answer- After filtering and therefore cleaning the data the data and grouping data based on sport (Weightlifting and Badminton.) ,by removing columns, from weightlifting athletes and badminton athletes some interesting relations between age. From the histogram we may see a skew for weightlifters we see it is right skewed while badminton players have a histogram which is uniform. This may indicate that the location of the majority of players we see that the weightlifters are usually are younger.The mean value seen in the summary table could also justify our suspicions because the mean for badminton players is 2 less than those for weightlifters. After the sorting occured on the backend to find the 50th percentile it has also shown that the median is also 2 less for weightlifting thatn badminton. Another run fact that can also be seen is how the standard deviation for badminton is higher than that of weightlifting which implies that the variabliilty is higher so the interval for age is wider for badminton so more diverse ages could play as athletes.

(f) Finish with a conclusion to remind your boss of the key take home points from your summary about the athletes' ages. Be definitive about what your findings are, but use *hedging* to caveat the limitations of the conclusion more generally.

This concludes that according to the data the weightlifting may require younger athletes and this may mean weightlifting may require a person to be in the healthiest/fittest section of life because it may require you to be at your peak.

Vocabulary

- Cleaning data
- Tidy data
- Handling missing values (NAs)
- Removing a column
- Extracting a subset of variables
- Filtering a tibble based on a condition (e.g. based on the values in one or more of the variables/columns)
- Sorting data based on the values of a variable
- Defining new variables
- Renaming the variables
- Producing new data frames
- Grouping categories
- Creating summary tables

You may also find these vocabulary words from last week useful with your writing this week

- location/center (mean, median, mode) and scale/spread (range, IQR, var, sd)
  - note: interpreting center and spread relative to each other can be helpful
- shape (symmetric, left-skewed, right-skewed, unimodal, bimodal, multimodal, uniform)
- outliers/extreme values
  - note: this can be related to the tails of a distribution (heavy-tailed, thin-tailed)
- frequency (most, least, pattern tendencies)

Part 2: OPTIONAL but Recommended

You may complete these questions for practice if you wish. ***You are not required to complete these questions as they ARE NOT included as part of your mark.***

Question 5: Amazon Books

The code below reads in data about books sold on Amazon.
- Note that the height (`Height`), width (`Width`), and thickness (`Thick`) of books in this data frame are measured in inches.

```
library(tidyverse) # Load the tidyverse package so it is available to use
books <- read.csv("amazonbooks.csv")
```

(a) What is the name of the book(s) with the smallest number of pages in this sample of books, and how many pages does it have?

```
# Type your code here
```

(b) Create a summary table which reports the total number of books written by each author and the mean and variance of the number of pages per book for each author, for the books represented in this sample of books.

```
# Type your code here
```

(c) Modify your code from (b) so to create a new summary table which contains only information for authors who wrote more than 2 books, and sort them in decreasing order of number of books written.

```
# Type your code here
```

Part 3: OPTIONAL for Additional Practice

You may complete these questions for practice if you wish. ***You are not required to complete these questions as they ARE NOT included as part of your mark.***

Question 6: Titanic Data

At the time it departed from England in April 1912, the RMS Titanic was the largest ship in the world. In the night of April 14th to April 15th, the Titanic struck an iceberg and sank approximately 600km south of Newfoundland (a province in eastern Canada). Many people perished in this accident. The code below loads data about the passengers who were on board the Titanic at the time of the accident.

```
titanic <- read_csv("titanic.csv")
```

```
## Rows: 2208 Columns: 14
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (12): Name, Survived, Boarded, Class, MWC, Adut_or_Chld, Sex, Ticket_No,...
## dbl  (2): Age, Paid
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
glimpse(titanic)
```

```
## Rows: 2,208
## Columns: 14
## $ Name         <chr> "ABBING, Mr Anthony", "ABBOTT, Mr Ernest Owen", "ABBOTT, ~
## $ Survived     <chr> "Dead", "Dead", "Dead", "Dead", "Alive", "Alive", "Alive"~
```

```
## $ Boarded      <chr> "Southampton", "Southampton", "Southampton", "Southampton~
## $ Class        <chr> "3", "Crew", "3", "3", "3", "3", "3", "2", "2", "3", "3",~
## $ MWC          <chr> "Man", "Man", "Child", "Man", "Woman", "Woman", "Man", "M~
## $ Age          <dbl> 42.00, 21.00, 14.00, 16.00, 39.00, 16.00, 25.00, 30.00, 2~
## $ Adut_or_Chld <chr> "Adult", "Adult", "Child", "Adult", "Adult", "Adult", "Ad~
## $ Sex          <chr> "Male", "Male", "Male", "Male", "Female", "Female", "Male~
## $ Paid         <dbl> 7.550000, NA, 20.250000, 20.250000, 20.250000, 7.650000, ~
## $ Ticket_No    <chr> "5547", NA, "CA2673", "CA2673", "CA2673", "348125", "3481~
## $ Boat_or_Body <chr> NA, NA, NA, "[190]", "A", "16", "A", NA, "10", "15", "C",~
## $ Job          <chr> "Blacksmith", "Lounge Pantry Steward", "Scholar", "Jewell~
## $ Class_Dept   <chr> "3rd Class Passenger", "Victualling Crew", "3rd Class Pas~
## $ Class_Full   <chr> "3", "V", "3", "3", "3", "3", "3", "2", "2", "3", "3", "E~
```

(a) Often, before you start working with a dataset you need to clean it.

- The variable `Adut_or_Chld` indicates which passengers were adults and which were children. Use the `rename()` function to change the name of this variable to `Adult_or_Child`. The variable `MWC` records whether the passenger was a man, woman or child. Use the `rename()` function to change the name of this variable to `Man_Woman_or_Child` to make this clear.

```
# Type your code here
```

Hint: Unless the transformed tibble is saved into a new object or overwrites the original tibble, like `oly12 <- oly12 %>% rename(Place_of_birth = PlaceOB)`, the changes won't be permanent.

- Since many of their values are missing or unclear, modify the `titanic` data frame by removing the following variables: `Ticket_No`, `Boat_or_Body`, `CLass_Dept`, `Class_Full`.

```
# Type your code here
```

(b) Create a summary table reporting the number of passengers on the Titanic (n), the number of passengers who survied (n_surv), and the proportion of passengers who survived (prop_surv).

```
# Type your code here
```

(c) Calculate the proportion of deaths for the following groups of passengers.

- For men, women, and children:

```
# Type your code here
```

- For passengers aged between 25-40 years of age:

```
# Type your code here
```

- For men, women, and children among the passengers who paid more than 50 British pounds for their tickets:

```
# Type your code here
```

- Write several sentences interpreting the summary tables created for the three groups above.

REPLACE THIS TEXT WITH YOUR ANSWER

(d) What was the most common job among passengers of the Titanic? Write 1-2 sentences explaining your answer.

```
# Type your code here
```

REPLACE THIS TEXT WITH YOUR ANSWER

(e) Plot the age distribution for passengers with the job "General Labourer", and describe this distribution in 1-2 sentences.

```
# Type your code here
```

REPLACE THIS TEXT WITH YOUR ANSWER

(f) Were any of the general labourers on the titanic women? If so, how many?

```
# Type your code here
```

REPLACE THIS TEXT WITH YOUR ANSWER

(g) What are the names of the passengers with the top 4 most expensive tickets? Did these passengers survive the accident?

```
# Type your code here
```

REPLACE THIS TEXT WITH YOUR ANSWER

(h) In this question, you will compare the distribution of ticket prices for survivors and non-survivors of the Titanic using both visualizations and summary tables.

- Construct two histograms to visualize the distribution of ticket prices for survivors and non-survivors (i.e. one histogram for survivors and one for non-survivors). Write 2-3 sentences comparing the two distributions based on these plots.

```
# Type your code here
```

REPLACE THIS TEXT WITH YOUR ANSWER

- Construct a pair of boxplots (in the same figure) to visualize the distribution of ticket prices for survivors and non-survivors. Write 2-3 sentences comparing the two distributions based on these plots.

```
# Type your code here
```

REPLACE THIS TEXT WITH YOUR ANSWER

- Construct a summary table with the minimum, first quartile, median, mean, third quartile, and maximum ticket price for survivors and non-survivors.

```
#### Example code to demo quantile() function and is.na ####
x <- c(1,2,3,4,5,6,NA,10)
quantile(x, probs = 0.25, na.rm=TRUE); # Calculate the first quartile (25% quantile), and tell R to exc
```

```
## 25%
## 2.5
```

```
quantile(x, probs = 0.75, na.rm=TRUE); # Calculate the third quartile (75% quantile), and tell R to exc
```

```
## 75%
## 5.5
```

```r
# If there are missing values in the vector you're working with (or in one of the columns of a tibble),
mean(x)
```

```
## [1] NA
```

```r
mean(x, na.rm=TRUE)
```

```
## [1] 4.428571
```

```r
median(x)
```

```
## [1] NA
```

```r
median(x, na.rm=TRUE)
```

```
## [1] 4
```

- Write 2-3 sentences comparing the two distributions based on this summary table.

REPLACE THIS TEXT WITH YOUR ANSWER

- Comment on the strengths and weaknesses of each of the visualizations and summary table constructed above.

REPLACE THIS TEXT WITH YOUR ANSWER