

#Final Project

Question - Can type of Sport be predicted based on a set of variables?

```
require(randomForest)

## Loading required package: randomForest
## randomForest 4.6-14
## Type rfNews() to see new features/changes/bug fixes.
require(tidyverse)

## Loading required package: tidyverse
## -- Attaching packages ----- tidyverse 1.3.1 --
## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.3      v dplyr  1.0.7
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   2.0.0      v forcats 0.5.1
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::combine() masks randomForest::combine()
## x dplyr::filter()  masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## x ggplot2::margin() masks randomForest::margin()
require(reshape2)

## Loading required package: reshape2
##
## Attaching package: 'reshape2'
## The following object is masked from 'package:tidyr':
##
##     smiths
require(scales)

## Loading required package: scales
##
## Attaching package: 'scales'
## The following object is masked from 'package:purrr':
##
##     discard
## The following object is masked from 'package:readr':
##
##     col_factor
require(ggplot2)
require(moments)

## Loading required package: moments
```

```
# Loading dataset
```

```
setwd("/Users/rajindersingh/Documents")
```

```
olympics <- read.csv("athlete_events.csv")
```

```
str(olympics)
```

```
## 'data.frame': 271116 obs. of 15 variables:
## $ ID : int 1 2 3 4 5 5 5 5 5 ...
## $ Name : chr "A Dijiang" "A Lamusi" "Gunnar Nielsen Aaby" "Edgar Lindenau Aabye" ...
## $ Sex : chr "M" "M" "M" "M" ...
## $ Age : int 24 23 24 34 21 21 25 25 27 27 ...
## $ Height: int 180 170 NA NA 185 185 185 185 185 185 ...
## $ Weight: num 80 60 NA NA 82 82 82 82 82 82 ...
## $ Team : chr "China" "China" "Denmark" "Denmark/Sweden" ...
## $ NOC : chr "CHN" "CHN" "DEN" "DEN" ...
## $ Games : chr "1992 Summer" "2012 Summer" "1920 Summer" "1900 Summer" ...
## $ Year : int 1992 2012 1920 1900 1988 1988 1992 1992 1994 1994 ...
## $ Season: chr "Summer" "Summer" "Summer" "Summer" ...
## $ City : chr "Barcelona" "London" "Antwerpen" "Paris" ...
## $ Sport : chr "Basketball" "Judo" "Football" "Tug-Of-War" ...
## $ Event : chr "Basketball Men's Basketball" "Judo Men's Extra-Lightweight" "Football Men's Football" ...
## $ Medal : chr NA NA NA "Gold" ...
```

```
summary(olympics)
```

```
##      ID      Name      Sex      Age
## Min.   : 1      Length:271116      Length:271116      Min.   :10.00
## 1st Qu.:34643      Class :character      Class :character      1st Qu.:21.00
## Median :68205      Mode  :character      Mode  :character      Median :24.00
## Mean   :68249
## 3rd Qu.:102097
## Max.   :135571
##                                     NA's :9474
##      Height      Weight      Team      NOC
## Min.   :127.0      Min.   : 25.0      Length:271116      Length:271116
## 1st Qu.:168.0      1st Qu.: 60.0      Class :character      Class :character
## Median :175.0      Median : 70.0      Mode  :character      Mode  :character
## Mean   :175.3      Mean   : 70.7
## 3rd Qu.:183.0      3rd Qu.: 79.0
## Max.   :226.0      Max.   :214.0
## NA's   :60171      NA's   :62875
##      Games      Year      Season      City
## Length:271116      Min.   :1896      Length:271116      Length:271116
## Class :character      1st Qu.:1960      Class :character      Class :character
## Mode  :character      Median :1988      Mode  :character      Mode  :character
##                                     Mean   :1978
##                                     3rd Qu.:2002
##                                     Max.   :2016
##      Sport      Event      Medal
## Length:271116      Length:271116      Length:271116
## Class :character      Class :character      Class :character
## Mode  :character      Mode  :character      Mode  :character
##
```

```
##
##
##
```

```
head(olympics)
```

```
##      ID              Name Sex Age Height Weight      Team NOC
## 1  1          A Dijiang   M  24    180     80      China CHN
## 2  2          A Lamusi   M  23    170     60      China CHN
## 3  3      Gunnar Nielsen Aaby M  24     NA     NA      Denmark DEN
## 4  4      Edgar Lindenau Aabye M  34     NA     NA Denmark/Sweden DEN
## 5  5 Christine Jacoba Aaftink F  21    185     82 Netherlands NED
## 6  5 Christine Jacoba Aaftink F  21    185     82 Netherlands NED
##      Games Year Season      City      Sport
## 1 1992 Summer 1992 Summer Barcelona Basketball
## 2 2012 Summer 2012 Summer   London       Judo
## 3 1920 Summer 1920 Summer Antwerpen   Football
## 4 1900 Summer 1900 Summer   Paris    Tug-Of-War
## 5 1988 Winter 1988 Winter  Calgary Speed Skating
## 6 1988 Winter 1988 Winter  Calgary Speed Skating
##      Event Medal
## 1      Basketball Men's Basketball <NA>
## 2      Judo Men's Extra-Lightweight <NA>
## 3      Football Men's Football <NA>
## 4      Tug-Of-War Men's Tug-Of-War Gold
## 5      Speed Skating Women's 500 metres <NA>
## 6      Speed Skating Women's 1,000 metres <NA>
```

Omit 'NA'

```
any(is.na(olympics))
```

```
## [1] TRUE
```

```
olympics <- na.omit(olympics)
any(is.na(olympics))
```

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

```
str(olympics)
```

```
## 'data.frame':   30181 obs. of  15 variables:
## $ ID      : int  16 17 17 17 17 17 20 20 20 20 ...
## $ Name    : chr  "Juhamatti Tapio Aaltonen" "Paavo Johannes Aaltonen" "Paavo Johannes Aaltonen" "Paavo
## $ Sex      : chr  "M" "M" "M" "M" ...
## $ Age      : int  28 28 28 28 28 32 20 20 22 22 ...
## $ Height   : int  184 175 175 175 175 175 176 176 176 176 ...
## $ Weight   : num  85 64 64 64 64 64 85 85 85 85 ...
## $ Team     : chr  "Finland" "Finland" "Finland" "Finland" ...
## $ NOC      : chr  "FIN" "FIN" "FIN" "FIN" ...
## $ Games    : chr  "2014 Winter" "1948 Summer" "1948 Summer" "1948 Summer" ...
## $ Year     : int  2014 1948 1948 1948 1948 1952 1992 1992 1994 1994 ...
## $ Season   : chr  "Winter" "Summer" "Summer" "Summer" ...
## $ City     : chr  "Sochi" "London" "London" "London" ...
## $ Sport    : chr  "Ice Hockey" "Gymnastics" "Gymnastics" "Gymnastics" ...
```

```
## $ Event : chr "Ice Hockey Men's Ice Hockey" "Gymnastics Men's Individual All-Around" "Gymnastics M
## $ Medal : chr "Bronze" "Bronze" "Gold" "Gold" ...
## - attr(*, "na.action")= 'omit' Named int [1:240935] 1 2 3 4 5 6 7 8 9 10 ...
## ..- attr(*, "names")= chr [1:240935] "1" "2" "3" "4" ...
```

```
summary(olympics)
```

```
##           ID           Name           Sex           Age
## Min.      :    16   Length:30181   Length:30181   Min.      :13.00
## 1st Qu.: 37494   Class :character   Class :character   1st Qu.:22.00
## Median : 69771   Mode  :character   Mode  :character   Median :25.00
## Mean      : 70226                                     Mean      :25.43
## 3rd Qu.:104111                                     3rd Qu.:28.00
## Max.      :135563                                     Max.      :66.00
##           Height           Weight           Team           NOC
## Min.      :136.0   Min.      : 28.00   Length:30181   Length:30181
## 1st Qu.:170.0   1st Qu.: 63.00   Class :character   Class :character
## Median :178.0   Median : 73.00   Mode  :character   Mode  :character
## Mean      :177.6   Mean      : 73.75
## 3rd Qu.:185.0   3rd Qu.: 83.00
## Max.      :223.0   Max.      :182.00
##           Games           Year           Season           City
## Length:30181   Min.      :1896   Length:30181   Length:30181
## Class :character   1st Qu.:1976   Class :character   Class :character
## Mode  :character   Median :1992   Mode  :character   Mode  :character
##                                     Mean      :1988
##                                     3rd Qu.:2006
##                                     Max.      :2016
##           Sport           Event           Medal
## Length:30181   Length:30181   Length:30181
## Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character
##
##
##
```

```
head(olympics)
```

```
##           ID           Name Sex Age Height Weight   Team NOC           Games
## 41 16 Juhamatti Tapio Aaltonen M 28 184 85 Finland FIN 2014 Winter
## 42 17 Paavo Johannes Aaltonen M 28 175 64 Finland FIN 1948 Summer
## 43 17 Paavo Johannes Aaltonen M 28 175 64 Finland FIN 1948 Summer
## 45 17 Paavo Johannes Aaltonen M 28 175 64 Finland FIN 1948 Summer
## 49 17 Paavo Johannes Aaltonen M 28 175 64 Finland FIN 1948 Summer
## 51 17 Paavo Johannes Aaltonen M 32 175 64 Finland FIN 1952 Summer
##           Year Season   City      Sport           Event
## 41 2014 Winter   Sochi Ice Hockey   Ice Hockey Men's Ice Hockey
## 42 1948 Summer   London Gymnastics Gymnastics Men's Individual All-Around
## 43 1948 Summer   London Gymnastics   Gymnastics Men's Team All-Around
## 45 1948 Summer   London Gymnastics   Gymnastics Men's Horse Vault
## 49 1948 Summer   London Gymnastics   Gymnastics Men's Pommel Horse
## 51 1952 Summer   Helsinki Gymnastics   Gymnastics Men's Team All-Around
##           Medal
## 41 Bronze
## 42 Bronze
```

```
## 43 Gold
## 45 Gold
## 49 Gold
## 51 Bronze
```

Table of category proportions for Sex

```
prop.table(table(olympics$Sex))
```

```
##
##           F           M
## 0.342931 0.657069
```

Number of unique Sport

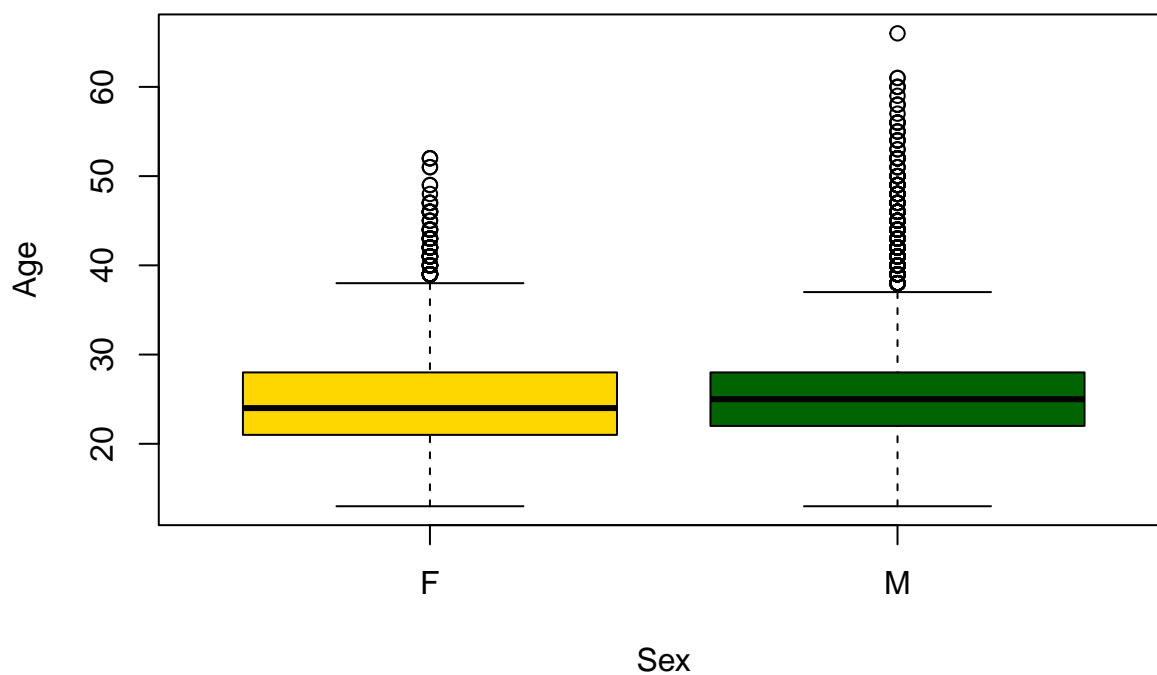
```
unique(olympics$Sport)
```

```
## [1] "Ice Hockey"           "Gymnastics"
## [3] "Alpine Skiing"        "Rowing"
## [5] "Football"            "Fencing"
## [7] "Taekwondo"           "Athletics"
## [9] "Canoeing"            "Handball"
## [11] "Water Polo"          "Wrestling"
## [13] "Sailing"              "Cycling"
## [15] "Hockey"               "Figure Skating"
## [17] "Softball"            "Swimming"
## [19] "Boxing"               "Basketball"
## [21] "Nordic Combined"     "Diving"
## [23] "Baseball"            "Volleyball"
## [25] "Speed Skating"       "Cross Country Skiing"
## [27] "Bobsleigh"           "Modern Pentathlon"
## [29] "Curling"              "Judo"
## [31] "Rugby Sevens"        "Tennis"
## [33] "Rhythmic Gymnastics" "Weightlifting"
## [35] "Equestrianism"       "Badminton"
## [37] "Beach Volleyball"    "Ski Jumping"
## [39] "Shooting"            "Short Track Speed Skating"
## [41] "Biathlon"            "Synchronized Swimming"
## [43] "Freestyle Skiing"    "Triathlon"
## [45] "Luge"                 "Table Tennis"
## [47] "Snowboarding"        "Skeleton"
## [49] "Rugby"                "Archery"
## [51] "Tug-Of-War"          "Trampolining"
## [53] "Lacrosse"            "Golf"
## [55] "Art Competitions"
```

Comparing Men & Women dataset for Age, Height & Weight

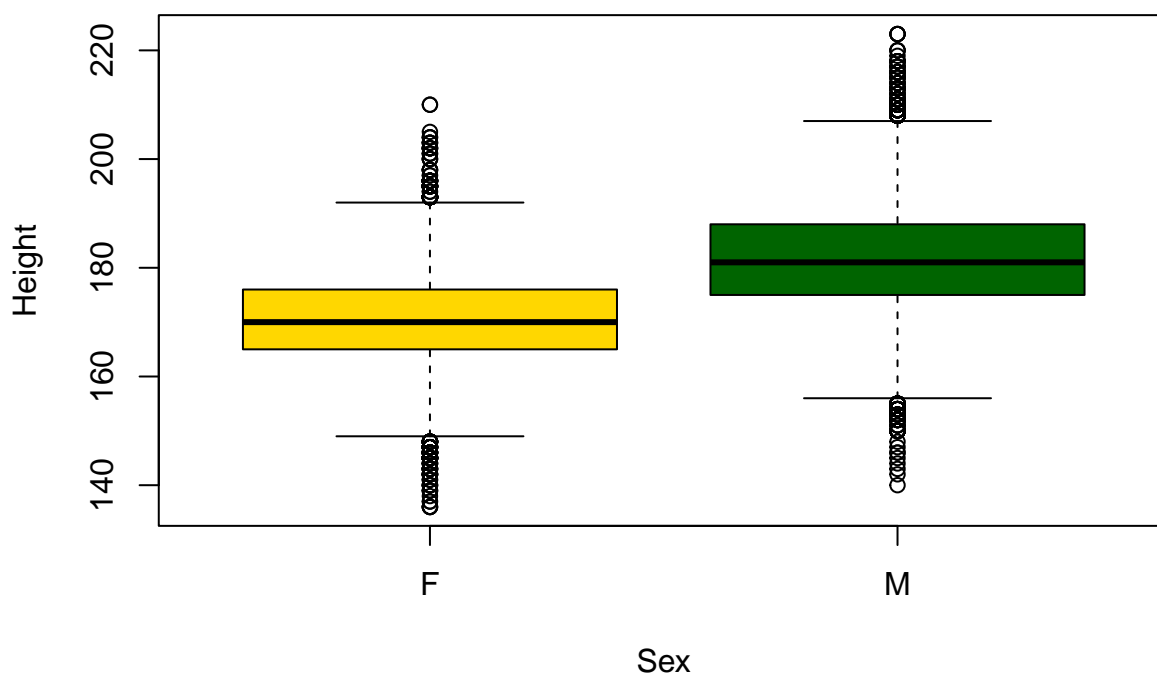
```
boxplot(Age~Sex, data=olympics, col=c("gold","darkgreen")),
        main = "Age of Men & Women", xlab="Sex",ylab="Age")
```

Age of Men & Women



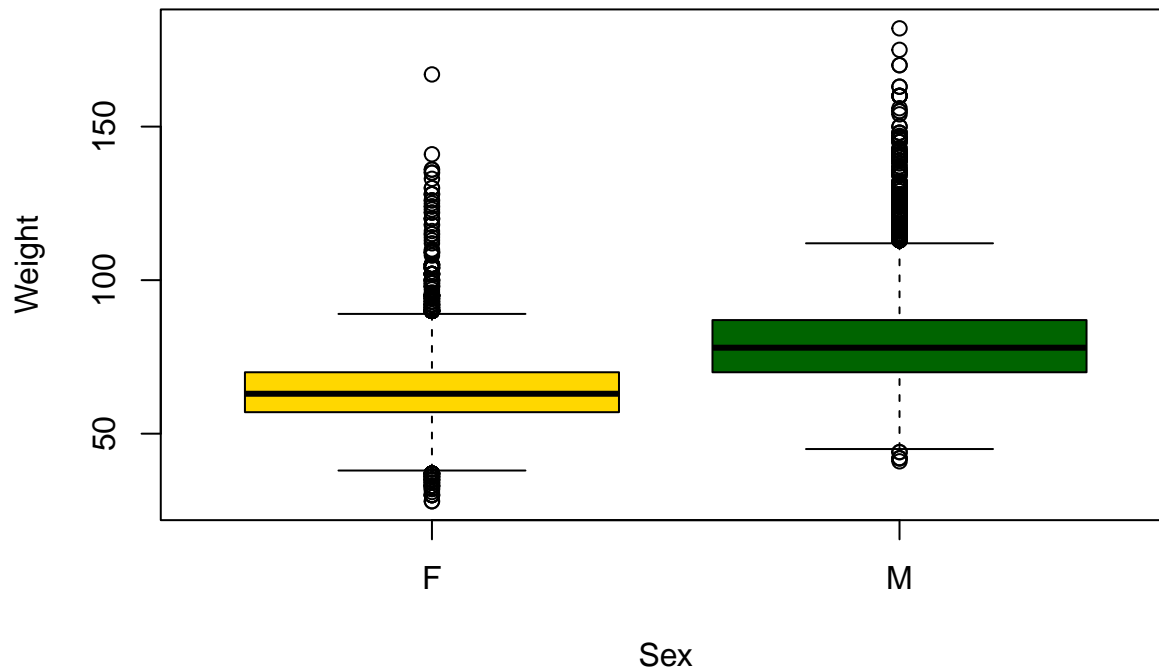
```
boxplot(Age~Sex, data=olympics, col=(c("gold","darkgreen")),
        main = "Height of Men & Women", xlab="Sex",ylab="Height")
```

Height of Men & Women



```
boxplot(Height~Sex, data=olympics, col=(c("gold","darkgreen")),
        main = "Weight of Male & Female", xlab="Sex",ylab="Weight")
```

Weight of Male & Female



mean of Men's and Women's age is almost similar, but the mean for # Men's height and weight is higher than women.

Men Athlete_Events - randomForest analysis

```
results <- olympics
str(results)
```

```
## 'data.frame': 30181 obs. of 15 variables:
## $ ID : int 16 17 17 17 17 17 20 20 20 20 ...
## $ Name : chr "Juhamatti Tapio Aaltonen" "Paavo Johannes Aaltonen" "Paavo Johannes Aaltonen" "Paavo Johannes Aaltonen" ...
## $ Sex : chr "M" "M" "M" "M" ...
## $ Age : int 28 28 28 28 28 32 20 20 22 22 ...
## $ Height: int 184 175 175 175 175 175 176 176 176 176 ...
## $ Weight: num 85 64 64 64 64 64 85 85 85 85 ...
## $ Team : chr "Finland" "Finland" "Finland" "Finland" ...
## $ NOC : chr "FIN" "FIN" "FIN" "FIN" ...
## $ Games: chr "2014 Winter" "1948 Summer" "1948 Summer" "1948 Summer" ...
## $ Year : int 2014 1948 1948 1948 1948 1952 1992 1992 1994 1994 ...
## $ Season: chr "Winter" "Summer" "Summer" "Summer" ...
## $ City : chr "Sochi" "London" "London" "London" ...
## $ Sport : chr "Ice Hockey" "Gymnastics" "Gymnastics" "Gymnastics" ...
## $ Event : chr "Ice Hockey Men's Ice Hockey" "Gymnastics Men's Individual All-Around" "Gymnastics Men's Individual All-Around" ...
## $ Medal : chr "Bronze" "Bronze" "Gold" "Gold" ...
## - attr(*, "na.action")= 'omit' Named int [1:240935] 1 2 3 4 5 6 7 8 9 10 ...
## ..- attr(*, "names")= chr [1:240935] "1" "2" "3" "4" ...
```

```
length(which(results$Sex == "M"))
```

```
## [1] 19831
```

```
results.men <- results[results$Sex == "M", ]
str(results.men)
```

```
## 'data.frame': 19831 obs. of 15 variables:
## $ ID : int 16 17 17 17 17 17 20 20 20 20 ...
## $ Name : chr "Juhamatti Tapio Aaltonen" "Paavo Johannes Aaltonen" "Paavo Johannes Aaltonen" "Paavo
## $ Sex : chr "M" "M" "M" "M" ...
## $ Age : int 28 28 28 28 28 32 20 20 22 22 ...
## $ Height: int 184 175 175 175 175 175 176 176 176 176 ...
## $ Weight: num 85 64 64 64 64 64 85 85 85 85 ...
## $ Team : chr "Finland" "Finland" "Finland" "Finland" ...
## $ NOC : chr "FIN" "FIN" "FIN" "FIN" ...
## $ Games : chr "2014 Winter" "1948 Summer" "1948 Summer" "1948 Summer" ...
## $ Year : int 2014 1948 1948 1948 1948 1952 1992 1992 1994 1994 ...
## $ Season: chr "Winter" "Summer" "Summer" "Summer" ...
## $ City : chr "Sochi" "London" "London" "London" ...
## $ Sport : chr "Ice Hockey" "Gymnastics" "Gymnastics" "Gymnastics" ...
## $ Event : chr "Ice Hockey Men's Ice Hockey" "Gymnastics Men's Individual All-Around" "Gymnastics M
## $ Medal : chr "Bronze" "Bronze" "Gold" "Gold" ...
## - attr(*, "na.action")= 'omit' Named int [1:240935] 1 2 3 4 5 6 7 8 9 10 ...
## ..- attr(*, "names")= chr [1:240935] "1" "2" "3" "4" ...
```

```
results.men$Sport <- as.factor(results.men$Sport)
str(results.men)
```

```
## 'data.frame': 19831 obs. of 15 variables:
## $ ID : int 16 17 17 17 17 17 20 20 20 20 ...
## $ Name : chr "Juhamatti Tapio Aaltonen" "Paavo Johannes Aaltonen" "Paavo Johannes Aaltonen" "Paavo
## $ Sex : chr "M" "M" "M" "M" ...
## $ Age : int 28 28 28 28 28 32 20 20 22 22 ...
## $ Height: int 184 175 175 175 175 175 176 176 176 176 ...
## $ Weight: num 85 64 64 64 64 64 85 85 85 85 ...
## $ Team : chr "Finland" "Finland" "Finland" "Finland" ...
## $ NOC : chr "FIN" "FIN" "FIN" "FIN" ...
## $ Games : chr "2014 Winter" "1948 Summer" "1948 Summer" "1948 Summer" ...
## $ Year : int 2014 1948 1948 1948 1948 1952 1992 1992 1994 1994 ...
## $ Season: chr "Winter" "Summer" "Summer" "Summer" ...
## $ City : chr "Sochi" "London" "London" "London" ...
## $ Sport : Factor w/ 52 levels "Alpine Skiing",...: 26 23 23 23 23 23 1 1 1 1 ...
## $ Event : chr "Ice Hockey Men's Ice Hockey" "Gymnastics Men's Individual All-Around" "Gymnastics M
## $ Medal : chr "Bronze" "Bronze" "Gold" "Gold" ...
## - attr(*, "na.action")= 'omit' Named int [1:240935] 1 2 3 4 5 6 7 8 9 10 ...
## ..- attr(*, "names")= chr [1:240935] "1" "2" "3" "4" ...
```

```
results.men.rf <- randomForest(results.men[, -13], results.men[, 13])
results.men.rf
```

```
##
## Call:
## randomForest(x = results.men[, -13], y = results.men[, 13])
##           Type of random forest: classification
##           Number of trees: 500
## No. of variables tried at each split: 3
##
##           OOB estimate of error rate: 2.72%
```



```

## Confusion matrix:
##
## Alpine Skiing      174      0      0      0
## Archery            0     106      0      0
## Art Competitions   0      0      0      2
## Athletics          0      0      0    2409
## Badminton         0      0      0      0
## Baseball          0      0      0      0
## Basketball         0      0      0      0
## Beach Volleyball   0      0      0      0
## Biathlon           0      0      0      0
## Bobsleigh          0      0      0      0
## Boxing             0      0      0      0
## Canoeing           0      0      0      0
## Cross Country Skiing 0      0      0      0
## Curling            0      0      0      0
## Cycling            0      0      0      0
## Diving             0      0      0      0
## Equestrianism      0      0      0      0
## Fencing            0      0      0      0
## Figure Skating     0      0      0      0
## Football           0      0      0      0
## Freestyle Skiing   0      0      0      0
## Golf               0      0      0      0
## Gymnastics         0      0      0      0
## Handball           0      0      0      0
## Hockey             0      0      0      0
## Ice Hockey         0      0      0      0
## Judo               0      0      0      0
## Lacrosse           0      0      0      0
## Luge               0      0      0      0
## Modern Pentathlon  0      0      0      0
## Nordic Combined    0      0      0      0
## Rowing             0      0      0      0
## Rugby             0      0      0      0
## Rugby Sevens      0      0      0      0
## Sailing            0      0      0      0
## Shooting           0      0      0      0
## Short Track Speed Skating 0      0      0      0
## Skeleton           0      0      0      0
## Ski Jumping        0      0      0      0
## Snowboarding       0      0      0      0
## Speed Skating      0      0      0      0
## Swimming           0      0      0      0
## Table Tennis       0      0      0      0
## Taekwondo          0      0      0      0
## Tennis             0      0      0      0
## Trampoline         0      0      0      0
## Triathlon          0      0      0      0
## Tug-Of-War         0      0      0      0
## Volleyball         0      0      0      0
## Water Polo         0      0      0      0
## Weightlifting       0      0      0      0
## Wrestling          0      0      0      0

```

##	Badminton	Baseball	Basketball	Beach Volleyball
## Alpine Skiing	0	0	0	0
## Archery	0	0	0	0
## Art Competitions	0	0	0	0
## Athletics	0	0	0	0
## Badminton	71	4	1	0
## Baseball	0	332	1	0
## Basketball	0	5	600	0
## Beach Volleyball	0	0	6	18
## Biathlon	0	0	0	0
## Bobsleigh	0	0	0	0
## Boxing	0	1	7	0
## Canoeing	0	0	2	0
## Cross Country Skiing	0	0	0	0
## Curling	0	0	0	0
## Cycling	0	1	1	0
## Diving	0	0	0	0
## Equestrianism	0	0	0	0
## Fencing	0	0	0	0
## Figure Skating	0	0	0	0
## Football	0	0	1	0
## Freestyle Skiing	0	0	0	0
## Golf	0	0	0	0
## Gymnastics	0	0	0	0
## Handball	0	0	1	0
## Hockey	0	0	0	0
## Ice Hockey	0	0	0	0
## Judo	0	0	0	0
## Lacrosse	0	0	0	0
## Luge	0	0	0	0
## Modern Pentathlon	0	0	0	0
## Nordic Combined	0	0	0	0
## Rowing	0	0	0	0
## Rugby	0	0	0	0
## Rugby Sevens	0	0	0	0
## Sailing	0	0	0	0
## Shooting	0	0	0	0
## Short Track Speed Skating	0	0	0	0
## Skeleton	0	0	0	0
## Ski Jumping	0	0	0	0
## Snowboarding	0	0	0	0
## Speed Skating	0	0	0	0
## Swimming	0	0	0	0
## Table Tennis	0	0	0	0
## Taekwondo	0	0	0	0
## Tennis	0	0	0	0
## Trampolining	0	0	0	0
## Triathlon	0	0	0	0
## Tug-Of-War	0	0	0	0
## Volleyball	0	0	0	0
## Water Polo	0	0	0	0
## Weightlifting	0	0	0	0
## Wrestling	0	0	0	0
##	Biathlon	Bobsleigh	Boxing	Canoeing

## Alpine Skiing	0	0	0	0
## Archery	0	0	0	0
## Art Competitions	0	0	0	0
## Athletics	0	0	0	0
## Badminton	0	0	4	1
## Baseball	0	0	0	0
## Basketball	0	0	4	1
## Beach Volleyball	0	0	3	6
## Biathlon	239	1	0	0
## Bobsleigh	1	258	0	0
## Boxing	0	0	656	18
## Canoeing	0	0	11	757
## Cross Country Skiing	0	0	0	0
## Curling	0	1	0	0
## Cycling	0	0	1	2
## Diving	0	0	0	0
## Equestrianism	0	0	0	0
## Fencing	0	0	0	2
## Figure Skating	0	0	0	0
## Football	0	0	0	1
## Freestyle Skiing	0	0	0	0
## Golf	0	0	0	0
## Gymnastics	0	0	0	0
## Handball	0	0	0	0
## Hockey	0	0	0	0
## Ice Hockey	0	0	0	0
## Judo	0	0	0	0
## Lacrosse	0	0	0	0
## Luge	0	0	0	0
## Modern Pentathlon	0	0	0	0
## Nordic Combined	0	0	0	0
## Rowing	0	0	1	0
## Rugby	0	0	0	0
## Rugby Sevens	0	0	0	0
## Sailing	0	0	0	0
## Shooting	0	0	0	0
## Short Track Speed Skating	0	0	0	0
## Skeleton	0	0	0	0
## Ski Jumping	0	0	0	0
## Snowboarding	0	0	0	0
## Speed Skating	0	0	0	0
## Swimming	0	0	0	0
## Table Tennis	0	0	0	0
## Taekwondo	0	0	0	0
## Tennis	0	0	0	0
## Trampoline	0	0	0	0
## Triathlon	0	0	0	0
## Tug-Of-War	0	0	0	0
## Volleyball	0	0	0	0
## Water Polo	0	0	0	0
## Weightlifting	0	0	0	0
## Wrestling	0	0	0	0
##				
	Cross Country Skiing	Curling	Cycling	Diving
## Alpine Skiing		0	0	0

## Archery	0	0	0	0
## Art Competitions	0	0	0	0
## Athletics	0	0	0	0
## Badminton	0	0	0	0
## Baseball	0	0	0	0
## Basketball	0	0	0	0
## Beach Volleyball	0	0	0	0
## Biathlon	1	0	0	0
## Bobsleigh	0	0	0	0
## Boxing	0	0	0	0
## Canoeing	0	0	0	0
## Cross Country Skiing	370	0	0	0
## Curling	1	62	0	0
## Cycling	0	0	651	2
## Diving	0	0	15	134
## Equestrianism	0	0	2	2
## Fencing	0	0	1	1
## Figure Skating	1	3	0	0
## Football	0	0	0	0
## Freestyle Skiing	0	2	0	0
## Golf	0	0	0	0
## Gymnastics	0	0	0	0
## Handball	0	0	0	0
## Hockey	0	0	0	0
## Ice Hockey	0	0	0	0
## Judo	0	0	0	0
## Lacrosse	0	0	0	0
## Luge	0	0	0	0
## Modern Pentathlon	0	0	0	0
## Nordic Combined	0	0	0	0
## Rowing	0	0	0	0
## Rugby	0	0	0	0
## Rugby Sevens	0	0	0	0
## Sailing	0	0	0	0
## Shooting	0	0	0	0
## Short Track Speed Skating	0	0	0	0
## Skeleton	0	0	0	0
## Ski Jumping	0	0	0	0
## Snowboarding	0	0	0	0
## Speed Skating	0	0	0	0
## Swimming	0	0	0	0
## Table Tennis	0	0	0	0
## Taekwondo	0	0	0	0
## Tennis	0	0	0	0
## Trampolining	0	0	0	0
## Triathlon	0	0	0	0
## Tug-Of-War	0	0	0	0
## Volleyball	0	0	0	0
## Water Polo	0	0	0	0
## Weightlifting	0	0	0	0
## Wrestling	0	0	0	0
##				
	Equestrianism	Fencing	Figure Skating	Football
## Alpine Skiing	0	0	0	0
## Archery	0	0	0	0

## Art Competitions	0	0	0	0		
## Athletics	0	0	0	0		
## Badminton	0	0	0	0		
## Baseball	0	0	0	0		
## Basketball	0	0	0	0		
## Beach Volleyball	0	0	0	0		
## Biathlon	0	0	0	0		
## Bobsleigh	0	0	0	0		
## Boxing	0	0	0	0		
## Canoeing	0	0	0	0		
## Cross Country Skiing	0	0	0	0		
## Curling	0	0	1	0		
## Cycling	3	5	0	0		
## Diving	3	1	0	0		
## Equestrianism	455	9	0	2		
## Fencing	1	768	0	2		
## Figure Skating	0	1	117	0		
## Football	1	6	0	763		
## Freestyle Skiing	0	0	11	0		
## Golf	1	0	0	0		
## Gymnastics	0	0	0	3		
## Handball	0	0	0	0		
## Hockey	0	2	0	0		
## Ice Hockey	0	0	0	0		
## Judo	0	1	0	1		
## Lacrosse	0	0	0	0		
## Luge	0	0	0	0		
## Modern Pentathlon	0	0	0	0		
## Nordic Combined	0	0	0	0		
## Rowing	0	1	0	0		
## Rugby	0	0	0	0		
## Rugby Sevens	0	0	0	0		
## Sailing	0	0	0	0		
## Shooting	0	0	0	0		
## Short Track Speed Skating	0	0	0	0		
## Skeleton	0	0	0	0		
## Ski Jumping	0	0	0	0		
## Snowboarding	0	0	0	0		
## Speed Skating	0	0	0	0		
## Swimming	0	0	0	0		
## Table Tennis	0	0	0	0		
## Taekwondo	0	0	0	0		
## Tennis	0	0	0	0		
## Trampoline	0	0	0	0		
## Triathlon	0	0	0	0		
## Tug-Of-War	0	0	0	0		
## Volleyball	0	0	0	0		
## Water Polo	0	0	0	0		
## Weightlifting	0	0	0	0		
## Wrestling	0	0	0	0		
##						
## Alpine Skiing		Freestyle Skiing	Golf	Gymnastics	Handball	Hockey
## Archery		0	0	0	0	0
## Art Competitions		0	0	0	0	0

## Athletics	0	0	0	0	0
## Badminton	0	0	0	0	0
## Baseball	0	0	0	0	0
## Basketball	0	0	0	0	0
## Beach Volleyball	0	0	0	0	0
## Biathlon	0	0	0	0	0
## Bobsleigh	0	0	0	0	0
## Boxing	0	0	0	2	0
## Canoeing	0	0	0	0	0
## Cross Country Skiing	0	0	0	0	0
## Curling	0	0	0	0	0
## Cycling	0	0	2	0	0
## Diving	0	0	4	0	0
## Equestrianism	0	0	0	0	1
## Fencing	0	0	1	7	0
## Figure Skating	1	0	0	0	0
## Football	0	0	5	5	0
## Freestyle Skiing	35	0	0	0	0
## Golf	0	0	0	0	0
## Gymnastics	0	0	643	0	4
## Handball	0	0	3	509	2
## Hockey	0	0	0	3	708
## Ice Hockey	0	0	0	0	0
## Judo	0	0	7	8	7
## Lacrosse	0	0	0	0	0
## Luge	0	0	0	0	0
## Modern Pentathlon	0	0	2	3	3
## Nordic Combined	0	0	0	0	0
## Rowing	0	0	1	1	0
## Rugby	0	0	0	0	0
## Rugby Sevens	0	0	0	0	0
## Sailing	0	0	2	0	2
## Shooting	0	0	2	0	1
## Short Track Speed Skating	0	0	0	0	0
## Skeleton	0	0	0	0	0
## Ski Jumping	0	0	0	0	0
## Snowboarding	0	0	0	0	0
## Speed Skating	0	0	0	0	0
## Swimming	0	0	0	0	0
## Table Tennis	0	0	1	0	0
## Taekwondo	0	0	0	0	0
## Tennis	0	0	0	0	0
## Trampolining	0	0	0	0	0
## Triathlon	0	0	1	0	0
## Tug-Of-War	0	0	0	0	0
## Volleyball	0	0	0	0	0
## Water Polo	0	0	0	0	0
## Weightlifting	0	0	0	0	0
## Wrestling	0	0	0	0	0
##					
	Ice Hockey	Judo	Lacrosse	Luge	Modern Pentathlon
## Alpine Skiing	0	0	0	0	0
## Archery	0	0	0	0	0
## Art Competitions	0	0	0	0	0
## Athletics	0	0	0	0	0

## Badminton	0	0	0	0	0
## Baseball	0	0	0	0	0
## Basketball	0	0	0	0	0
## Beach Volleyball	0	0	0	0	0
## Biathlon	0	0	0	0	0
## Bobsleigh	0	0	0	0	0
## Boxing	0	0	0	0	0
## Canoeing	0	0	0	0	0
## Cross Country Skiing	1	0	0	0	0
## Curling	0	0	0	0	0
## Cycling	0	0	0	0	0
## Diving	0	0	0	0	0
## Equestrianism	0	0	0	0	0
## Fencing	0	0	0	0	0
## Figure Skating	0	0	0	0	0
## Football	0	0	0	0	0
## Freestyle Skiing	2	0	0	0	0
## Golf	0	0	0	0	0
## Gymnastics	0	0	0	0	0
## Handball	0	0	0	0	0
## Hockey	0	0	0	0	0
## Ice Hockey	1001	0	0	0	0
## Judo	0	307	0	0	2
## Lacrosse	0	0	2	0	0
## Luge	2	0	0	129	0
## Modern Pentathlon	0	4	0	0	108
## Nordic Combined	1	0	0	3	0
## Rowing	0	1	0	0	1
## Rugby	0	0	0	0	0
## Rugby Sevens	0	0	0	0	0
## Sailing	0	2	0	0	0
## Shooting	0	2	0	0	0
## Short Track Speed Skating	1	0	0	0	0
## Skeleton	2	0	0	0	0
## Ski Jumping	1	0	0	1	0
## Snowboarding	1	0	0	1	0
## Speed Skating	1	0	0	0	0
## Swimming	0	0	0	0	0
## Table Tennis	0	0	0	0	0
## Taekwondo	0	0	0	0	0
## Tennis	0	0	0	0	0
## Trampoline	0	0	0	0	0
## Triathlon	0	0	0	0	0
## Tug-Of-War	0	0	0	0	0
## Volleyball	0	0	0	0	0
## Water Polo	0	0	0	0	0
## Weightlifting	0	0	0	0	0
## Wrestling	0	0	0	0	0
##					
	Nordic Combined	Rowing	Rugby	Rugby Sevens	Sailing
## Alpine Skiing	0	0	0	0	0
## Archery	0	0	0	0	0
## Art Competitions	0	0	0	0	0
## Athletics	0	0	0	0	0
## Badminton	0	0	0	0	0

## Baseball	0	0	0	0	0
## Basketball	0	0	0	0	0
## Beach Volleyball	0	0	0	0	0
## Biathlon	0	0	0	0	0
## Bobsleigh	0	0	0	0	0
## Boxing	0	1	0	0	0
## Canoeing	0	0	0	0	0
## Cross Country Skiing	0	0	0	0	0
## Curling	0	0	0	0	0
## Cycling	0	1	0	0	0
## Diving	0	0	0	0	0
## Equestrianism	0	0	0	0	0
## Fencing	0	0	0	0	0
## Figure Skating	0	0	0	0	0
## Football	0	0	0	0	0
## Freestyle Skiing	1	0	0	0	0
## Golf	0	1	0	0	0
## Gymnastics	0	3	0	0	0
## Handball	0	1	0	0	0
## Hockey	0	1	0	0	0
## Ice Hockey	0	0	0	0	0
## Judo	0	3	0	0	0
## Lacrosse	0	0	0	0	0
## Luge	2	0	0	0	0
## Modern Pentathlon	0	14	0	0	2
## Nordic Combined	132	0	0	0	0
## Rowing	0	1387	0	0	3
## Rugby	0	2	28	0	0
## Rugby Sevens	0	1	0	36	0
## Sailing	0	15	2	2	519
## Shooting	0	6	0	4	25
## Short Track Speed Skating	0	0	0	0	0
## Skeleton	0	0	0	0	0
## Ski Jumping	0	0	0	0	0
## Snowboarding	0	0	0	0	0
## Speed Skating	0	0	0	0	0
## Swimming	0	2	0	0	0
## Table Tennis	0	0	0	0	0
## Taekwondo	0	0	0	0	1
## Tennis	0	1	2	0	1
## Trampoline	0	0	0	0	0
## Triathlon	0	0	0	0	1
## Tug-Of-War	0	0	0	0	0
## Volleyball	0	0	0	0	0
## Water Polo	0	0	0	0	0
## Weightlifting	0	0	0	0	0
## Wrestling	0	0	0	0	0
##					
	Shooting	Short Track	Speed Skating	Skeleton	
## Alpine Skiing	0			0	0
## Archery	0			0	0
## Art Competitions	0			0	0
## Athletics	0			0	0
## Badminton	0			0	0
## Baseball	0			0	0

## Basketball	0	0	0	
## Beach Volleyball	0	0	0	
## Biathlon	0	0	0	
## Bobsleigh	0	0	0	
## Boxing	0	0	0	
## Canoeing	0	0	0	
## Cross Country Skiing	0	0	0	
## Curling	0	0	0	
## Cycling	0	0	0	
## Diving	0	0	0	
## Equestrianism	0	0	0	
## Fencing	0	0	0	
## Figure Skating	0	1	0	
## Football	0	0	0	
## Freestyle Skiing	0	0	0	
## Golf	0	0	0	
## Gymnastics	0	0	0	
## Handball	0	0	0	
## Hockey	0	0	0	
## Ice Hockey	0	0	0	
## Judo	0	0	0	
## Lacrosse	0	0	0	
## Luge	0	0	0	
## Modern Pentathlon	2	0	0	
## Nordic Combined	0	1	0	
## Rowing	0	0	0	
## Rugby	0	0	0	
## Rugby Sevens	1	0	0	
## Sailing	7	0	0	
## Shooting	352	0	0	
## Short Track Speed Skating	0	135	0	
## Skeleton	0	1	5	
## Ski Jumping	0	0	0	
## Snowboarding	0	1	0	
## Speed Skating	0	2	0	
## Swimming	0	0	0	
## Table Tennis	2	0	0	
## Taekwondo	0	0	0	
## Tennis	2	0	0	
## Trampoline	2	0	0	
## Triathlon	0	0	0	
## Tug-Of-War	0	0	0	
## Volleyball	1	0	0	
## Water Polo	0	0	0	
## Weightlifting	1	0	0	
## Wrestling	0	0	0	
##				
	Ski Jumping	Snowboarding	Speed Skating	Swimming
## Alpine Skiing	0	0	0	0
## Archery	0	0	0	0
## Art Competitions	0	0	0	0
## Athletics	0	0	0	0
## Badminton	0	0	0	0
## Baseball	0	0	0	0
## Basketball	0	0	0	0

## Beach Volleyball	0	0	0	0	
## Biathlon	0	0	0	0	
## Bobsleigh	0	0	0	0	
## Boxing	0	0	0	0	
## Canoeing	0	0	0	0	
## Cross Country Skiing	0	0	0	0	
## Curling	0	0	0	0	
## Cycling	0	0	0	0	
## Diving	0	0	0	0	
## Equestrianism	0	0	0	0	
## Fencing	0	0	0	0	
## Figure Skating	0	0	0	0	
## Football	0	0	0	0	
## Freestyle Skiing	0	0	0	0	
## Golf	0	0	0	0	
## Gymnastics	0	0	0	0	
## Handball	0	0	0	0	
## Hockey	0	0	0	0	
## Ice Hockey	0	0	0	0	
## Judo	0	0	0	0	
## Lacrosse	0	0	0	0	
## Luge	0	0	0	0	
## Modern Pentathlon	0	0	0	0	
## Nordic Combined	0	0	0	0	
## Rowing	0	0	0	0	
## Rugby	0	0	0	0	
## Rugby Sevens	0	0	0	0	
## Sailing	0	0	0	1	
## Shooting	0	0	0	2	
## Short Track Speed Skating	1	1	0	0	
## Skeleton	1	3	0	0	
## Ski Jumping	166	0	4	0	
## Snowboarding	6	27	8	0	
## Speed Skating	1	0	248	0	
## Swimming	0	0	0	1311	
## Table Tennis	0	0	0	7	
## Taekwondo	0	0	0	12	
## Tennis	0	0	0	10	
## Trampolineing	0	0	0	2	
## Triathlon	0	0	0	0	
## Tug-Of-War	0	0	0	0	
## Volleyball	0	0	0	3	
## Water Polo	0	0	0	2	
## Weightlifting	0	0	0	4	
## Wrestling	0	0	0	0	
##					
	Table Tennis	Taekwondo	Tennis	Trampolineing	Triathlon
## Alpine Skiing	0	0	0	0	0
## Archery	0	0	0	0	0
## Art Competitions	0	0	0	0	0
## Athletics	0	0	0	0	0
## Badminton	0	0	0	0	0
## Baseball	0	0	0	0	0
## Basketball	0	0	0	0	0
## Beach Volleyball	0	0	0	0	0

## Biathlon	0	0	0	0	0
## Bobsleigh	0	0	0	0	0
## Boxing	0	0	0	0	0
## Canoeing	0	0	0	0	0
## Cross Country Skiing	0	0	0	0	0
## Curling	0	0	0	0	0
## Cycling	0	0	0	0	0
## Diving	0	0	0	0	0
## Equestrianism	0	0	0	0	0
## Fencing	0	0	0	0	0
## Figure Skating	0	0	0	0	0
## Football	0	0	0	0	0
## Freestyle Skiing	0	0	0	0	0
## Golf	0	0	0	0	0
## Gymnastics	0	0	0	0	0
## Handball	0	0	0	0	0
## Hockey	0	0	0	0	0
## Ice Hockey	0	0	0	0	0
## Judo	0	0	0	0	0
## Lacrosse	0	0	0	0	0
## Luge	0	0	0	0	0
## Modern Pentathlon	0	0	0	0	0
## Nordic Combined	0	0	0	0	0
## Rowing	0	0	0	0	0
## Rugby	0	0	0	0	0
## Rugby Sevens	0	0	0	0	0
## Sailing	0	0	0	0	0
## Shooting	2	0	0	0	0
## Short Track Speed Skating	0	0	0	0	0
## Skeleton	0	0	0	0	0
## Ski Jumping	0	0	0	0	0
## Snowboarding	0	0	0	0	0
## Speed Skating	0	0	0	0	0
## Swimming	0	0	0	0	0
## Table Tennis	71	0	0	0	0
## Taekwondo	0	48	4	0	0
## Tennis	1	0	61	0	1
## Trampoline	0	3	0	6	0
## Triathlon	4	2	1	0	6
## Tug-Of-War	0	0	0	0	0
## Volleyball	0	0	1	0	0
## Water Polo	0	0	0	0	0
## Weightlifting	0	0	0	0	0
## Wrestling	0	0	0	0	0
##					
## Alpine Skiing	0	0	0	0	
## Archery	0	0	0	0	
## Art Competitions	0	0	0	0	
## Athletics	0	0	0	0	
## Badminton	0	0	0	0	
## Baseball	0	0	0	0	
## Basketball	0	0	0	0	
## Beach Volleyball	0	0	0	0	
## Biathlon	0	0	0	0	

## Bobsleigh	0	0	0	0
## Boxing	0	0	0	0
## Canoeing	0	0	0	0
## Cross Country Skiing	0	0	0	0
## Curling	0	0	0	0
## Cycling	0	0	0	0
## Diving	0	0	0	0
## Equestrianism	0	0	0	0
## Fencing	0	0	0	0
## Figure Skating	0	0	0	0
## Football	0	0	1	0
## Freestyle Skiing	0	0	0	0
## Golf	0	0	0	0
## Gymnastics	0	0	0	0
## Handball	0	0	0	0
## Hockey	0	0	0	0
## Ice Hockey	0	0	0	0
## Judo	0	0	0	1
## Lacrosse	0	0	0	0
## Luge	0	0	0	0
## Modern Pentathlon	0	0	0	0
## Nordic Combined	0	0	0	0
## Rowing	0	0	0	0
## Rugby	0	0	0	0
## Rugby Sevens	0	0	0	0
## Sailing	0	0	0	0
## Shooting	0	0	0	1
## Short Track Speed Skating	0	0	0	0
## Skeleton	0	0	0	0
## Ski Jumping	0	0	0	0
## Snowboarding	0	0	0	0
## Speed Skating	0	0	0	0
## Swimming	0	3	2	1
## Table Tennis	0	0	2	0
## Taekwondo	0	6	0	1
## Tennis	0	8	2	0
## Trampoline	0	0	0	1
## Triathlon	0	0	0	0
## Tug-Of-War	4	1	1	2
## Volleyball	0	484	0	0
## Water Polo	0	0	570	1
## Weightlifting	0	1	3	415
## Wrestling	0	0	0	1
##	Wrestling class.error			
## Alpine Skiing	0	0.000000000		
## Archery	0	0.000000000		
## Art Competitions	0	1.000000000		
## Athletics	0	0.000000000		
## Badminton	0	0.123456790		
## Baseball	0	0.003003003		
## Basketball	0	0.016393443		
## Beach Volleyball	0	0.454545455		
## Biathlon	0	0.008298755		
## Bobsleigh	0	0.003861004		

## Boxing	0 0.042335766
## Canoeing	0 0.016883117
## Cross Country Skiing	0 0.002695418
## Curling	0 0.046153846
## Cycling	0 0.026905830
## Diving	0 0.146496815
## Equestrianism	0 0.033970276
## Fencing	0 0.019157088
## Figure Skating	0 0.056451613
## Football	0 0.025542784
## Freestyle Skiing	0 0.313725490
## Golf	0 1.000000000
## Gymnastics	0 0.015313936
## Handball	0 0.013565891
## Hockey	0 0.008403361
## Ice Hockey	0 0.000000000
## Judo	1 0.091715976
## Lacrosse	0 0.000000000
## Luge	0 0.030075188
## Modern Pentathlon	0 0.217391304
## Nordic Combined	0 0.036496350
## Rowing	0 0.006446991
## Rugby	0 0.066666667
## Rugby Sevens	0 0.052631579
## Sailing	0 0.059782609
## Shooting	0 0.113350126
## Short Track Speed Skating	0 0.021739130
## Skeleton	0 0.583333333
## Ski Jumping	0 0.034883721
## Snowboarding	0 0.386363636
## Speed Skating	0 0.015873016
## Swimming	0 0.006065201
## Table Tennis	0 0.144578313
## Taekwondo	0 0.333333333
## Tennis	0 0.314606742
## Trampolining	1 0.600000000
## Triathlon	0 0.600000000
## Tug-Of-War	0 0.500000000
## Volleyball	0 0.010224949
## Water Polo	0 0.005235602
## Weightlifting	3 0.028103044
## Wrestling	898 0.001112347

importance(results.men.rf)

##	MeanDecreaseGini
## ID	754.9679
## Name	767.9187
## Sex	0.0000
## Age	862.2896
## Height	1261.1880
## Weight	1422.6836
## Team	1083.8689
## NOC	1068.5179
## Games	652.7355

```
## Year      622.0779
## Season    686.9855
## City      449.1892
## Event     8821.9892
## Medal     321.9505
```

Remove Event variable since similar to Sport variable

```
results1.men <- results.men[, -14]
str(results1.men)
```

```
## 'data.frame':  19831 obs. of  14 variables:
## $ ID      : int  16 17 17 17 17 17 20 20 20 20 ...
## $ Name    : chr  "Juhamatti Tapio Aaltonen" "Paavo Johannes Aaltonen" "Paavo Johannes Aaltonen" "Paavo Johannes Aaltonen" ...
## $ Sex     : chr  "M" "M" "M" "M" ...
## $ Age     : int  28 28 28 28 28 32 20 20 22 22 ...
## $ Height  : int  184 175 175 175 175 175 176 176 176 176 ...
## $ Weight  : num  85 64 64 64 64 64 85 85 85 85 ...
## $ Team    : chr  "Finland" "Finland" "Finland" "Finland" ...
## $ NOC     : chr  "FIN" "FIN" "FIN" "FIN" ...
## $ Games   : chr  "2014 Winter" "1948 Summer" "1948 Summer" "1948 Summer" ...
## $ Year    : int  2014 1948 1948 1948 1948 1952 1992 1992 1994 1994 ...
## $ Season  : chr  "Winter" "Summer" "Summer" "Summer" ...
## $ City    : chr  "Sochi" "London" "London" "London" ...
## $ Sport   : Factor w/ 52 levels "Alpine Skiing",...: 26 23 23 23 23 23 1 1 1 1 ...
## $ Medal   : chr  "Bronze" "Bronze" "Gold" "Gold" ...
```

Athlete__Events - randomForest analysis (Event variable removed)

```
results1.men.rf <- randomForest(results1.men[, -13], results1.men[, 13])
results1.men.rf
```

```
##
## Call:
## randomForest(x = results1.men[, -13], y = results1.men[, 13])
##           Type of random forest: classification
##           Number of trees: 500
## No. of variables tried at each split: 3
##
## OOB estimate of  error rate: 33.36%
## Confusion matrix:
##           Alpine Skiing Archery Art Competitions Athletics
## Alpine Skiing          117         0             0          0
## Archery                 0         32             0         14
## Art Competitions        0         0             0          1
## Athletics               0         0             0        1811
## Badminton               0         0             0          7
## Baseball                0         0             0          7
## Basketball              0         0             0         56
## Beach Volleyball        0         0             0          1
## Biathlon                0         0             0          0
## Bobsleigh               1         0             0          0
## Boxing                  0         2             0        112
```

## Canoeing	0	0	0	39
## Cross Country Skiing	3	0	0	0
## Curling	2	0	0	0
## Cycling	0	1	0	78
## Diving	0	0	0	19
## Equestrianism	0	0	0	40
## Fencing	0	1	0	76
## Figure Skating	4	0	0	0
## Football	0	0	0	35
## Freestyle Skiing	3	0	0	0
## Golf	0	0	0	0
## Gymnastics	0	0	0	23
## Handball	0	0	0	10
## Hockey	0	0	0	16
## Ice Hockey	0	0	0	17
## Judo	0	4	0	46
## Lacrosse	0	0	0	2
## Luge	10	0	0	0
## Modern Pentathlon	0	0	0	30
## Nordic Combined	4	0	0	0
## Rowing	0	0	0	92
## Rugby	0	0	0	9
## Rugby Sevens	0	0	0	0
## Sailing	0	0	0	56
## Shooting	0	2	0	35
## Short Track Speed Skating	1	0	0	0
## Skeleton	0	0	0	0
## Ski Jumping	4	0	0	0
## Snowboarding	4	0	0	0
## Speed Skating	3	0	0	0
## Swimming	0	2	0	75
## Table Tennis	0	1	0	4
## Taekwondo	0	0	0	21
## Tennis	0	0	0	10
## Trampolining	0	0	0	1
## Triathlon	0	0	0	5
## Tug-Of-War	0	0	0	5
## Volleyball	0	0	0	15
## Water Polo	0	0	0	27
## Weightlifting	0	0	0	28
## Wrestling	0	1	0	93
##				
	Badminton	Baseball	Basketball	Beach Volleyball
## Alpine Skiing	0	0	0	0
## Archery	0	9	0	0
## Art Competitions	0	0	0	0
## Athletics	1	16	20	0
## Badminton	26	4	0	0
## Baseball	1	291	2	0
## Basketball	0	1	423	0
## Beach Volleyball	0	0	2	11
## Biathlon	0	0	0	0
## Bobsleigh	0	0	0	0
## Boxing	2	13	3	0
## Canoeing	0	4	5	1

## Cross Country Skiing	0	0	0	0
## Curling	0	0	0	0
## Cycling	0	7	0	0
## Diving	0	0	0	0
## Equestrianism	0	0	0	0
## Fencing	0	6	0	0
## Figure Skating	0	0	0	0
## Football	0	0	5	0
## Freestyle Skiing	0	0	0	0
## Golf	0	0	0	0
## Gymnastics	0	1	0	0
## Handball	0	0	14	0
## Hockey	1	3	0	0
## Ice Hockey	0	0	0	0
## Judo	0	9	1	0
## Lacrosse	0	0	0	0
## Luge	0	0	0	0
## Modern Pentathlon	1	1	0	0
## Nordic Combined	0	0	0	0
## Rowing	0	2	20	0
## Rugby	0	0	0	0
## Rugby Sevens	0	0	0	0
## Sailing	0	3	4	0
## Shooting	5	5	1	0
## Short Track Speed Skating	0	0	0	0
## Skeleton	0	0	0	0
## Ski Jumping	0	0	0	0
## Snowboarding	0	0	0	0
## Speed Skating	0	0	0	0
## Swimming	0	5	8	0
## Table Tennis	0	0	1	0
## Taekwondo	1	2	1	0
## Tennis	0	0	2	0
## Trampolining	0	0	0	0
## Triathlon	0	0	0	0
## Tug-Of-War	0	0	0	0
## Volleyball	0	0	18	0
## Water Polo	0	1	8	0
## Weightlifting	0	0	1	0
## Wrestling	1	11	4	0
##				
	Biathlon	Bobsleigh	Boxing	Canoeing
## Alpine Skiing	7	6	0	0
## Archery	0	0	4	4
## Art Competitions	0	0	0	0
## Athletics	0	0	44	21
## Badminton	0	0	3	2
## Baseball	0	0	1	0
## Basketball	0	0	3	2
## Beach Volleyball	0	0	0	2
## Biathlon	179	0	0	0
## Bobsleigh	1	234	0	0
## Boxing	0	0	250	16
## Canoeing	0	0	7	457
## Cross Country Skiing	16	0	0	0

## Curling	2	1	0	0	
## Cycling	0	0	12	20	
## Diving	0	0	13	1	
## Equestrianism	0	0	2	6	
## Fencing	0	0	9	25	
## Figure Skating	14	2	0	0	
## Football	0	0	2	10	
## Freestyle Skiing	4	1	0	0	
## Golf	0	0	0	1	
## Gymnastics	0	0	12	2	
## Handball	0	0	3	19	
## Hockey	0	0	4	19	
## Ice Hockey	1	2	0	0	
## Judo	0	0	10	20	
## Lacrosse	0	0	0	0	
## Luge	8	26	0	0	
## Modern Pentathlon	0	0	2	3	
## Nordic Combined	11	0	0	0	
## Rowing	0	0	32	21	
## Rugby	0	0	0	0	
## Rugby Sevens	0	0	0	1	
## Sailing	0	0	8	11	
## Shooting	0	0	3	19	
## Short Track Speed Skating	1	0	0	0	
## Skeleton	2	0	0	0	
## Ski Jumping	4	0	0	0	
## Snowboarding	3	0	0	0	
## Speed Skating	4	1	0	0	
## Swimming	0	0	8	17	
## Table Tennis	0	0	3	1	
## Taekwondo	0	0	4	1	
## Tennis	0	0	0	7	
## Trampoline	0	0	1	0	
## Triathlon	0	0	0	0	
## Tug-Of-War	0	0	0	0	
## Volleyball	0	0	1	5	
## Water Polo	0	0	1	17	
## Weightlifting	0	0	3	9	
## Wrestling	0	0	34	35	
##					
		Cross Country Skiing	Curling	Cycling	Diving
## Alpine Skiing		8	1	0	0
## Archery		0	0	1	1
## Art Competitions		0	0	0	0
## Athletics		0	0	19	2
## Badminton		0	0	4	0
## Baseball		0	0	0	0
## Basketball		0	0	0	0
## Beach Volleyball		0	0	0	0
## Biathlon		27	0	0	0
## Bobsleigh		3	0	0	0
## Boxing		0	0	20	1
## Canoeing		0	0	32	0
## Cross Country Skiing		323	0	0	0
## Curling		12	17	0	0

## Cycling	0	0	312	1
## Diving	0	0	5	60
## Equestrianism	0	0	8	0
## Fencing	0	0	20	2
## Figure Skating	17	2	0	0
## Football	0	0	9	0
## Freestyle Skiing	2	0	0	0
## Golf	0	0	0	0
## Gymnastics	0	0	1	7
## Handball	0	0	7	0
## Hockey	0	0	16	0
## Ice Hockey	6	1	0	0
## Judo	0	0	8	0
## Lacrosse	0	0	0	0
## Luge	3	0	0	0
## Modern Pentathlon	0	0	3	0
## Nordic Combined	10	0	0	0
## Rowing	0	0	27	1
## Rugby	0	0	0	0
## Rugby Sevens	0	0	1	0
## Sailing	0	0	31	1
## Shooting	0	0	8	3
## Short Track Speed Skating	1	0	0	0
## Skeleton	2	1	0	0
## Ski Jumping	10	0	0	0
## Snowboarding	4	1	0	0
## Speed Skating	12	1	0	0
## Swimming	0	0	8	1
## Table Tennis	0	0	4	1
## Taekwondo	0	0	2	0
## Tennis	0	0	2	1
## Trampoline	0	0	0	3
## Triathlon	0	0	4	0
## Tug-Of-War	0	0	0	0
## Volleyball	0	0	4	0
## Water Polo	0	0	10	0
## Weightlifting	0	0	4	3
## Wrestling	0	0	8	3
##				
	Equestrianism	Fencing	Figure Skating	Football
## Alpine Skiing	0	0	1	0
## Archery	2	5	0	3
## Art Competitions	0	0	0	0
## Athletics	7	46	0	43
## Badminton	0	2	0	4
## Baseball	0	0	0	1
## Basketball	0	5	0	5
## Beach Volleyball	1	1	0	1
## Biathlon	0	0	2	0
## Bobsleigh	0	0	0	0
## Boxing	2	25	0	30
## Canoeing	2	32	0	25
## Cross Country Skiing	0	0	1	0
## Curling	0	0	0	0
## Cycling	7	36	0	38

## Diving	1	3	0	3	
## Equestrianism	330	16	0	9	
## Fencing	17	466	0	18	
## Figure Skating	0	0	46	0	
## Football	4	26	0	595	
## Freestyle Skiing	0	0	1	0	
## Golf	0	0	0	0	
## Gymnastics	1	5	0	8	
## Handball	0	9	0	4	
## Hockey	4	8	0	13	
## Ice Hockey	0	0	0	0	
## Judo	0	9	0	16	
## Lacrosse	0	0	0	0	
## Luge	0	0	3	0	
## Modern Pentathlon	2	16	0	5	
## Nordic Combined	0	0	1	0	
## Rowing	5	18	0	10	
## Rugby	0	0	0	0	
## Rugby Sevens	0	0	0	1	
## Sailing	39	23	0	17	
## Shooting	25	16	0	20	
## Short Track Speed Skating	0	0	0	0	
## Skeleton	0	0	1	0	
## Ski Jumping	0	0	2	0	
## Snowboarding	0	0	0	0	
## Speed Skating	0	0	1	0	
## Swimming	0	11	0	14	
## Table Tennis	0	1	0	3	
## Taekwondo	0	1	0	5	
## Tennis	0	4	0	2	
## Trampolining	0	0	0	2	
## Triathlon	0	1	0	1	
## Tug-Of-War	0	0	0	0	
## Volleyball	1	8	0	10	
## Water Polo	1	25	0	10	
## Weightlifting	0	2	0	6	
## Wrestling	4	14	0	22	
##	Freestyle Skiing	Golf	Gymnastics	Handball	Hockey
## Alpine Skiing	2	0	0	0	0
## Archery	0	0	0	0	1
## Art Competitions	0	0	0	0	0
## Athletics	0	0	14	9	25
## Badminton	0	0	3	0	4
## Baseball	0	0	0	0	1
## Basketball	0	0	0	1	2
## Beach Volleyball	0	0	0	0	1
## Biathlon	1	0	0	0	0
## Bobsleigh	1	0	0	0	0
## Boxing	0	0	26	11	11
## Canoeing	0	0	5	14	21
## Cross Country Skiing	0	0	0	0	0
## Curling	1	0	0	0	0
## Cycling	0	0	7	5	34
## Diving	0	0	21	0	6

## Equestrianism	0	0	3	3	13
## Fencing	0	0	3	13	15
## Figure Skating	0	0	0	0	0
## Football	0	0	14	7	7
## Freestyle Skiing	10	0	0	0	0
## Golf	0	0	0	0	0
## Gymnastics	0	0	556	0	2
## Handball	0	0	1	372	5
## Hockey	0	0	1	5	568
## Ice Hockey	0	0	0	0	0
## Judo	0	0	24	7	11
## Lacrosse	0	0	0	0	0
## Luge	0	0	0	0	0
## Modern Pentathlon	0	0	1	0	2
## Nordic Combined	0	0	0	0	0
## Rowing	0	0	2	17	7
## Rugby	0	0	0	0	0
## Rugby Sevens	0	0	0	2	0
## Sailing	0	0	4	8	41
## Shooting	0	0	12	6	7
## Short Track Speed Skating	0	0	0	0	0
## Skeleton	0	0	0	0	0
## Ski Jumping	0	0	0	0	0
## Snowboarding	3	0	0	0	0
## Speed Skating	0	0	0	0	0
## Swimming	0	0	2	5	14
## Table Tennis	0	0	3	0	0
## Taekwondo	0	0	0	2	3
## Tennis	0	0	0	0	9
## Trampoline	0	0	1	0	2
## Triathlon	0	0	1	0	1
## Tug-Of-War	0	0	0	0	0
## Volleyball	0	0	0	13	1
## Water Polo	0	0	0	20	5
## Weightlifting	0	0	28	1	3
## Wrestling	0	0	72	10	7
##					
	Ice Hockey	Judo	Lacrosse	Luge	Modern Pentathlon
## Alpine Skiing	14	0	0	5	0
## Archery	0	1	0	0	0
## Art Competitions	0	0	0	0	0
## Athletics	0	13	0	0	3
## Badminton	0	3	0	0	0
## Baseball	0	2	0	0	0
## Basketball	0	0	0	0	0
## Beach Volleyball	0	0	0	0	0
## Biathlon	17	0	0	0	0
## Bobsleigh	11	0	0	8	0
## Boxing	0	6	0	0	0
## Canoeing	0	1	0	0	0
## Cross Country Skiing	14	0	0	0	0
## Curling	21	0	0	1	0
## Cycling	0	3	0	0	0
## Diving	0	1	0	0	0
## Equestrianism	0	0	0	0	0

## Fencing	0	3	0	0	4	
## Figure Skating	18	0	0	0	0	
## Football	0	1	0	0	0	
## Freestyle Skiing	8	0	0	1	0	
## Golf	0	0	0	0	0	
## Gymnastics	0	1	0	0	0	
## Handball	0	3	0	0	0	
## Hockey	0	0	0	0	0	
## Ice Hockey	971	0	0	0	0	
## Judo	0	68	0	0	0	
## Lacrosse	0	0	0	0	0	
## Luge	8	0	0	70	0	
## Modern Pentathlon	0	0	0	0	59	
## Nordic Combined	4	0	0	0	0	
## Rowing	0	0	0	0	0	
## Rugby	0	0	0	0	0	
## Rugby Sevens	0	0	0	0	0	
## Sailing	0	1	0	0	0	
## Shooting	0	2	0	0	1	
## Short Track Speed Skating	2	0	0	1	0	
## Skeleton	3	0	0	0	0	
## Ski Jumping	2	0	0	1	0	
## Snowboarding	9	0	0	1	0	
## Speed Skating	30	0	0	0	0	
## Swimming	0	1	0	0	0	
## Table Tennis	0	1	0	0	0	
## Taekwondo	0	1	0	0	0	
## Tennis	0	0	0	0	0	
## Trampoline	0	0	0	0	0	
## Triathlon	0	0	0	0	0	
## Tug-Of-War	0	0	0	0	0	
## Volleyball	0	0	0	0	0	
## Water Polo	0	2	0	0	0	
## Weightlifting	0	4	0	0	0	
## Wrestling	0	16	0	0	0	
##		Nordic Combined	Rowing	Rugby	Rugby Sevens	Sailing
## Alpine Skiing		1	0	0	0	0
## Archery		0	2	0	1	4
## Art Competitions		0	0	1	0	0
## Athletics		0	84	1	0	16
## Badminton		0	3	0	0	1
## Baseball		0	6	0	0	1
## Basketball		0	26	0	0	1
## Beach Volleyball		0	6	0	0	0
## Biathlon		0	0	0	0	0
## Bobsleigh		0	0	0	0	0
## Boxing		0	37	0	0	3
## Canoeing		0	59	0	1	4
## Cross Country Skiing		7	0	0	0	0
## Curling		1	0	0	0	0
## Cycling		0	27	0	2	17
## Diving		0	1	0	0	1
## Equestrianism		0	12	0	0	11
## Fencing		0	34	0	0	4

## Figure Skating	4	0	0	0	0
## Football	0	17	0	0	3
## Freestyle Skiing	2	0	0	0	0
## Golf	0	0	0	0	0
## Gymnastics	0	1	0	0	0
## Handball	0	26	0	1	2
## Hockey	0	24	0	0	4
## Ice Hockey	1	0	0	0	0
## Judo	0	17	0	0	5
## Lacrosse	0	0	0	0	0
## Luge	0	0	0	0	0
## Modern Pentathlon	0	2	0	0	1
## Nordic Combined	91	0	0	0	0
## Rowing	0	1009	0	1	17
## Rugby	0	0	21	0	0
## Rugby Sevens	0	5	0	25	0
## Sailing	0	56	0	0	190
## Shooting	0	18	0	2	22
## Short Track Speed Skating	3	0	0	0	0
## Skeleton	0	0	0	0	0
## Ski Jumping	4	0	0	0	0
## Snowboarding	3	0	0	0	0
## Speed Skating	0	0	0	0	0
## Swimming	0	33	0	1	0
## Table Tennis	0	1	0	0	2
## Taekwondo	0	4	0	0	0
## Tennis	0	3	0	0	1
## Trampoline	0	0	0	0	0
## Triathlon	0	0	0	0	0
## Tug-Of-War	0	0	0	0	0
## Volleyball	0	31	0	0	1
## Water Polo	0	42	0	0	5
## Weightlifting	0	0	1	0	0
## Wrestling	0	29	0	0	9
##					
	Shooting	Short Track	Speed Skating	Skeleton	
## Alpine Skiing	0			2	0
## Archery	4			0	0
## Art Competitions	0			0	0
## Athletics	2			0	0
## Badminton	4			0	0
## Baseball	2			0	0
## Basketball	0			0	0
## Beach Volleyball	0			0	0
## Biathlon	0			1	1
## Bobsleigh	0			0	0
## Boxing	1			0	0
## Canoeing	6			0	0
## Cross Country Skiing	0			0	0
## Curling	0			3	0
## Cycling	5			0	0
## Diving	1			0	0
## Equestrianism	6			0	0
## Fencing	13			0	0
## Figure Skating	0			9	0

## Football	1	0	0
## Freestyle Skiing	0	12	0
## Golf	0	0	0
## Gymnastics	2	0	0
## Handball	4	0	0
## Hockey	1	0	0
## Ice Hockey	0	2	0
## Judo	3	0	0
## Lacrosse	0	0	0
## Luge	0	2	0
## Modern Pentathlon	3	0	0
## Nordic Combined	0	1	0
## Rowing	2	0	0
## Rugby	0	0	0
## Rugby Sevens	0	0	0
## Sailing	15	0	0
## Shooting	132	0	0
## Short Track Speed Skating	0	124	0
## Skeleton	0	1	1
## Ski Jumping	0	0	0
## Snowboarding	0	5	0
## Speed Skating	0	7	0
## Swimming	0	0	0
## Table Tennis	1	0	0
## Taekwondo	0	0	0
## Tennis	1	0	0
## Trampoline	1	0	0
## Triathlon	0	0	0
## Tug-Of-War	0	0	0
## Volleyball	0	0	0
## Water Polo	5	0	0
## Weightlifting	5	0	0
## Wrestling	5	0	0
##			
	Ski Jumping	Snowboarding	Speed Skating
## Alpine Skiing	4	2	4
## Archery	0	0	0
## Art Competitions	0	0	0
## Athletics	0	0	0
## Badminton	0	0	0
## Baseball	0	0	0
## Basketball	0	0	0
## Beach Volleyball	0	0	0
## Biathlon	6	0	7
## Bobsleigh	0	0	0
## Boxing	0	0	0
## Canoeing	0	0	0
## Cross Country Skiing	5	0	2
## Curling	0	0	4
## Cycling	0	0	0
## Diving	0	0	0
## Equestrianism	0	0	0
## Fencing	0	0	0
## Figure Skating	4	0	3
## Football	0	0	0

## Freestyle Skiing	1	3	3	0	
## Golf	0	0	0	1	
## Gymnastics	0	0	0	2	
## Handball	0	0	0	7	
## Hockey	0	0	0	19	
## Ice Hockey	0	0	0	0	
## Judo	0	0	0	7	
## Lacrosse	0	0	0	0	
## Luge	1	0	2	0	
## Modern Pentathlon	0	0	0	3	
## Nordic Combined	14	0	1	0	
## Rowing	0	0	0	48	
## Rugby	0	0	0	0	
## Rugby Sevens	0	0	0	0	
## Sailing	0	0	0	12	
## Shooting	0	0	0	8	
## Short Track Speed Skating	0	0	5	0	
## Skeleton	0	0	1	0	
## Ski Jumping	145	0	0	0	
## Snowboarding	1	10	0	0	
## Speed Skating	1	0	192	0	
## Swimming	0	0	0	1094	
## Table Tennis	0	0	0	1	
## Taekwondo	0	0	0	6	
## Tennis	0	0	0	13	
## Trampolining	0	0	0	0	
## Triathlon	0	0	0	0	
## Tug-Of-War	0	0	0	0	
## Volleyball	0	0	0	23	
## Water Polo	0	0	0	17	
## Weightlifting	0	0	0	3	
## Wrestling	0	0	0	10	
##	Table Tennis	Taekwondo	Tennis	Trampolining	Triathlon
## Alpine Skiing	0	0	0	0	0
## Archery	1	0	0	0	0
## Art Competitions	0	0	0	0	0
## Athletics	0	0	1	0	0
## Badminton	2	0	0	0	0
## Baseball	0	0	0	0	0
## Basketball	0	0	0	0	0
## Beach Volleyball	0	0	1	0	0
## Biathlon	0	0	0	0	0
## Bobsleigh	0	0	0	0	0
## Boxing	0	0	0	1	0
## Canoeing	0	0	1	0	0
## Cross Country Skiing	0	0	0	0	0
## Curling	0	0	0	0	0
## Cycling	0	0	2	0	0
## Diving	1	0	0	0	0
## Equestrianism	0	0	0	0	0
## Fencing	0	0	0	0	0
## Figure Skating	0	0	0	0	0
## Football	1	1	1	0	0
## Freestyle Skiing	0	0	0	0	0

## Golf	0	0	0	0	0
## Gymnastics	1	0	0	0	0
## Handball	0	0	0	0	0
## Hockey	0	0	0	0	0
## Ice Hockey	0	0	0	0	0
## Judo	0	0	0	0	0
## Lacrosse	0	0	0	0	0
## Luge	0	0	0	0	0
## Modern Pentathlon	0	0	0	0	0
## Nordic Combined	0	0	0	0	0
## Rowing	0	0	0	0	0
## Rugby	0	0	0	0	0
## Rugby Sevens	0	0	0	0	0
## Sailing	0	0	0	0	0
## Shooting	2	0	0	0	0
## Short Track Speed Skating	0	0	0	0	0
## Skeleton	0	0	0	0	0
## Ski Jumping	0	0	0	0	0
## Snowboarding	0	0	0	0	0
## Speed Skating	0	0	0	0	0
## Swimming	0	1	0	0	0
## Table Tennis	49	0	0	0	0
## Taekwondo	0	12	0	0	0
## Tennis	0	0	32	0	0
## Trampoline	0	0	0	3	0
## Triathlon	0	0	0	0	2
## Tug-Of-War	0	0	0	0	0
## Volleyball	0	1	0	0	0
## Water Polo	0	0	0	0	0
## Weightlifting	1	0	0	0	0
## Wrestling	0	0	0	0	0
##					
		Tug-Of-War	Volleyball	Water Polo	Weightlifting
## Alpine Skiing	0	0	0	0	0
## Archery	0	0	1	0	0
## Art Competitions	0	0	0	0	0
## Athletics	0	10	17	5	0
## Badminton	0	0	1	0	0
## Baseball	0	0	0	0	0
## Basketball	0	16	15	0	0
## Beach Volleyball	0	5	0	0	0
## Biathlon	0	0	0	0	0
## Bobsleigh	0	0	0	0	0
## Boxing	0	7	8	6	0
## Canoeing	0	4	14	5	0
## Cross Country Skiing	0	0	0	0	0
## Curling	0	0	0	0	0
## Cycling	0	2	7	5	0
## Diving	0	0	0	2	0
## Equestrianism	0	0	0	0	0
## Fencing	0	4	19	0	0
## Figure Skating	0	0	0	0	0
## Football	0	8	6	0	0
## Freestyle Skiing	0	0	0	0	0
## Golf	0	0	0	0	0

## Gymnastics	0	0	0	4
## Handball	0	10	14	2
## Hockey	0	0	2	3
## Ice Hockey	0	0	0	0
## Judo	0	1	4	19
## Lacrosse	0	0	0	0
## Luge	0	0	0	0
## Modern Pentathlon	0	2	1	0
## Nordic Combined	0	0	0	0
## Rowing	0	21	27	1
## Rugby	0	0	0	0
## Rugby Sevens	0	0	1	1
## Sailing	0	1	9	6
## Shooting	0	1	6	16
## Short Track Speed Skating	0	0	0	0
## Skeleton	0	0	0	0
## Ski Jumping	0	0	0	0
## Snowboarding	0	0	0	0
## Speed Skating	0	0	0	0
## Swimming	0	13	5	1
## Table Tennis	0	0	0	1
## Taekwondo	0	5	0	0
## Tennis	0	2	0	0
## Trampoline	0	0	0	1
## Triathlon	0	0	0	0
## Tug-Of-War	0	0	0	2
## Volleyball	0	351	6	0
## Water Polo	0	8	359	1
## Weightlifting	0	0	1	265
## Wrestling	1	2	19	34
##	Wrestling class.error			
## Alpine Skiing	0	0.32758621		
## Archery	5	0.69811321		
## Art Competitions	0	1.00000000		
## Athletics	42	0.24823578		
## Badminton	1	0.67901235		
## Baseball	2	0.12612613		
## Basketball	1	0.30655738		
## Beach Volleyball	0	0.66666667		
## Biathlon	0	0.25726141		
## Bobsleigh	0	0.09652510		
## Boxing	56	0.63503650		
## Canoeing	14	0.40649351		
## Cross Country Skiing	0	0.12938005		
## Curling	0	0.73846154		
## Cycling	7	0.53363229		
## Diving	8	0.61783439		
## Equestrianism	6	0.29936306		
## Fencing	7	0.40485313		
## Figure Skating	1	0.62903226		
## Football	10	0.24010217		
## Freestyle Skiing	0	0.80392157		
## Golf	0	1.00000000		
## Gymnastics	24	0.14854518		

```
## Handball 3 0.27906977
## Hockey 3 0.20448179
## Ice Hockey 0 0.02997003
## Judo 49 0.79881657
## Lacrosse 0 1.00000000
## Luge 0 0.47368421
## Modern Pentathlon 1 0.57246377
## Nordic Combined 0 0.33576642
## Rowing 16 0.27722063
## Rugby 0 0.30000000
## Rugby Sevens 1 0.34210526
## Sailing 16 0.65579710
## Shooting 20 0.66750630
## Short Track Speed Skating 0 0.10144928
## Skeleton 0 0.91666667
## Ski Jumping 0 0.15697674
## Snowboarding 0 0.77272727
## Speed Skating 0 0.23809524
## Swimming 0 0.17058378
## Table Tennis 5 0.40963855
## Taekwondo 1 0.83333333
## Tennis 0 0.64044944
## Trampoline 0 0.80000000
## Triathlon 0 0.86666667
## Tug-Of-War 1 1.00000000
## Volleyball 0 0.28220859
## Water Polo 9 0.37347295
## Weightlifting 59 0.37939110
## Wrestling 455 0.49388209
```

```
importance(results1.men.rf)
```

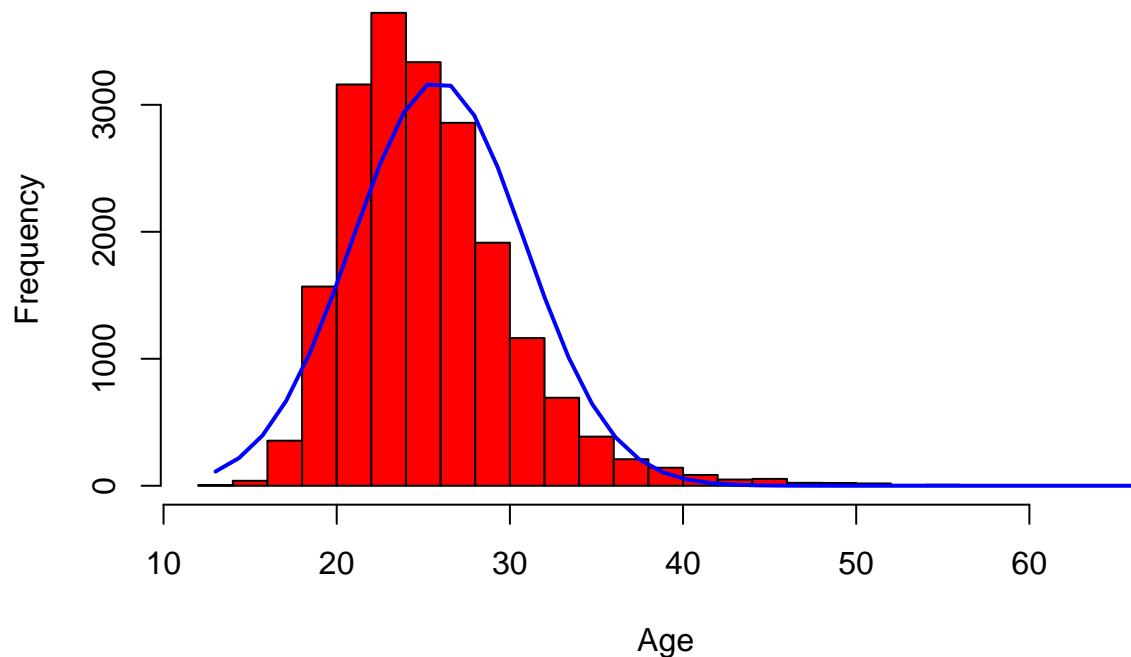
```
##      MeanDecreaseGini
## ID      1884.8159
## Name    1918.7664
## Sex      0.0000
## Age     1654.6780
## Height   2172.0614
## Weight   2425.1049
## Team     1840.8065
## NOC      1809.6330
## Games    1189.6189
## Year     1153.9786
## Season    597.2503
## City     879.1901
## Medal    780.3868
```

Men - histogram with normal curve of age

```
h<-hist(results1.men$Age, breaks=20, col="red", xlab="Age",
        main="Histogram with Normal Curve of Men's Age")
xfit<-seq(min(results1.men$Age),max(results1.men$Age),length=40)
yfit<-dnorm(xfit,mean=mean(results1.men$Age),sd=sd(results1.men$Age))
yfit <- yfit*diff(h$mids[1:2])*length(results1.men$Age)
```

```
lines(xfit, yfit, col="blue", lwd=2)
```

Histogram with Normal Curve of Men's Age



Men -

mean and standard deviation of age

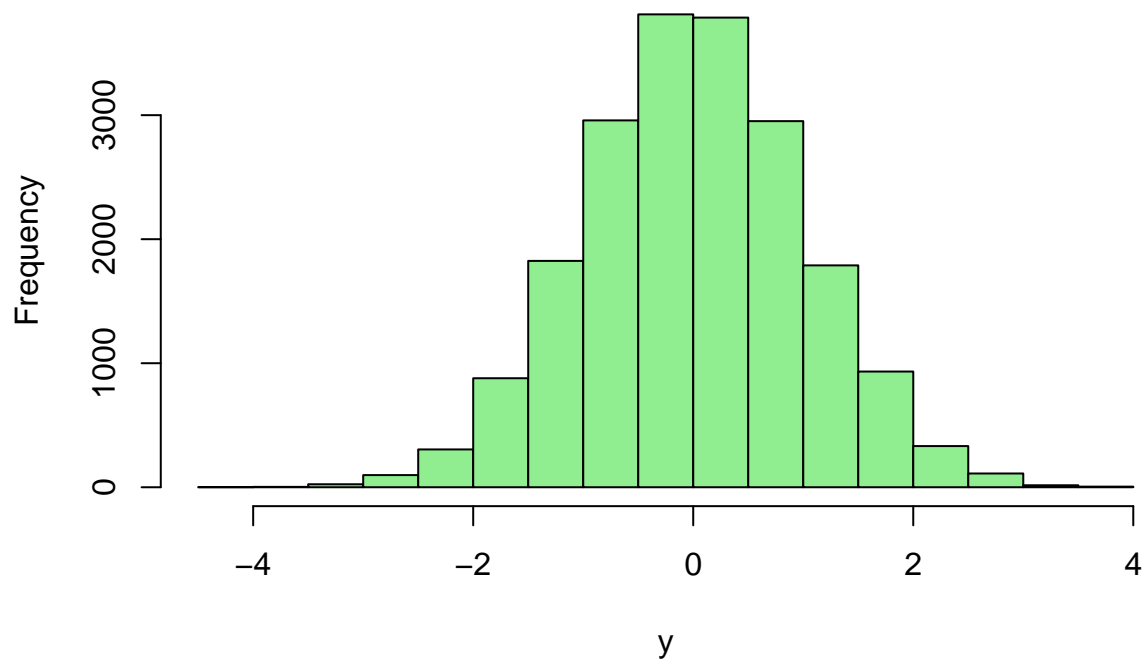
```
Ave <- mean(results1.men$Age)
SD <- sd(results1.men$Age)
c(Average = Ave, SD = SD)
```

```
## Average      SD
## 25.862942 4.969973
```

Men - normal distribution histogram of age

```
y <- rnorm(results1.men$Age)
hist(y, main = 'Age Normal Distribution', col = "light green")
```

Age Normal Distribution



late standard units

calcu-

```
z <- scale(results1.men$Age)

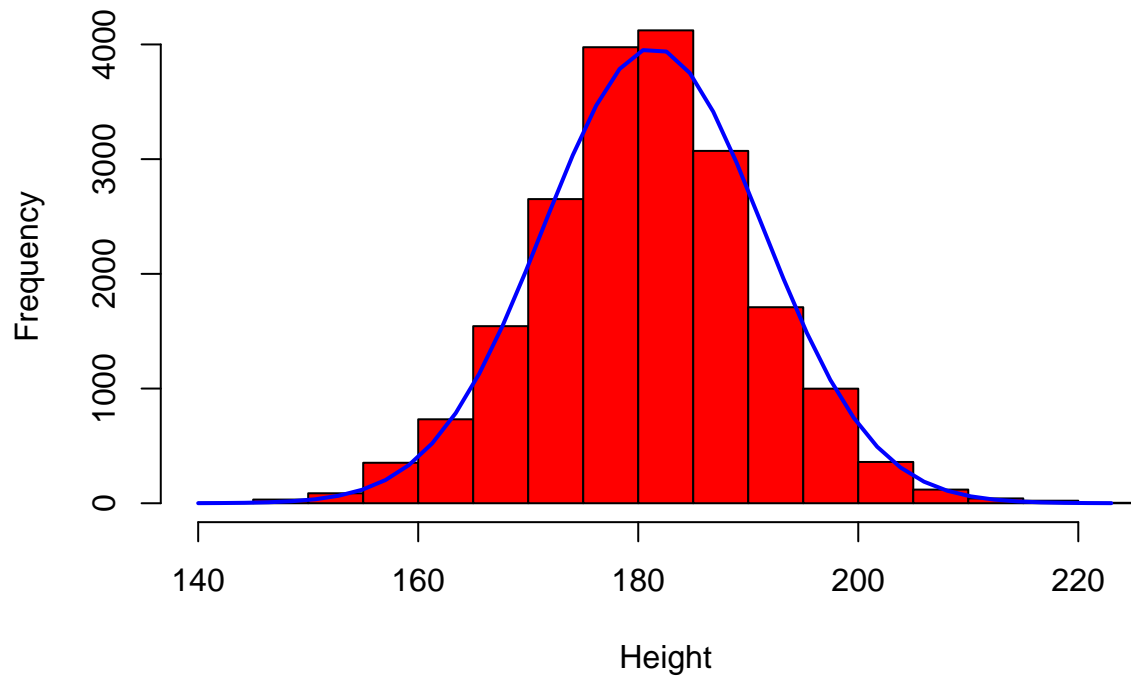
# calculate proportion of values within 2 SD of mean
mean(abs(z) < 2)

## [1] 0.9589532
```

Men - histogram with normal curve of height

```
h<-hist(results1.men$Height, breaks=20, col="red", xlab="Height",
        main="Histogram with Normal Curve of Men's Height")
xfit<-seq(min(results1.men$Height),max(results1.men$Height),length=40)
yfit<-dnorm(xfit,mean=mean(results1.men$Height),sd=sd(results1.men$Height))
yfit <- yfit*diff(h$mids[1:2])*length(results1.men$Height)
lines(xfit, yfit, col="blue", lwd=2)
```

Histogram with Normal Curve of Men's Height



mean and standard deviation of height

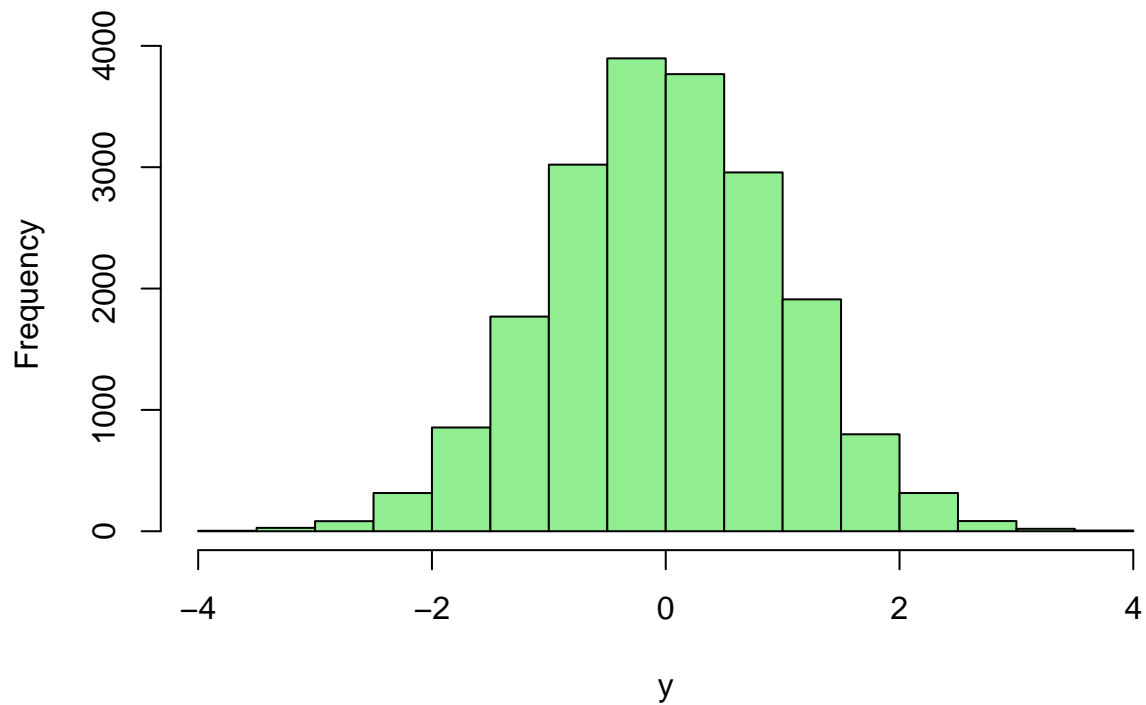
```
Ave <- mean(results1.men$Height)
SD <- sd(results1.men$Height)
c(Average = Ave, SD = SD)
```

```
##      Average      SD
## 181.352831  9.970714
```

Men - normal distribution histogram of height

```
y <- rnorm(results1.men$Height)
hist(y, main = 'Height Normal Distribution', col = "light green")
```

Height Normal Distribution



late standard units

calcu-

```
z <- scale(results1.men$Height)

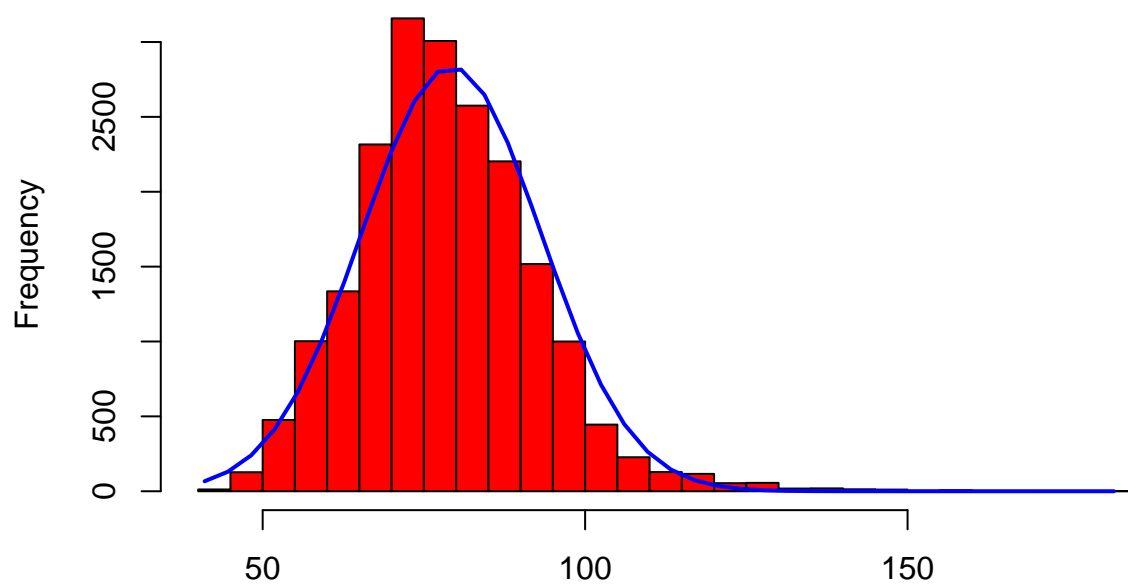
# calculate proportion of values within 2 SD of mean
mean(abs(z) < 2)

## [1] 0.9494226
```

Men - histogram with normal curve of weight

```
h<-hist(results1.men$Weight, breaks=30, col="red", xlab="Weight",
        main="Histogram with Normal Curve of Men's Weight")
xfit<-seq(min(results1.men$Weight),max(results1.men$Weight),length=40)
yfit<-dnorm(xfit,mean=mean(results1.men$Weight),sd=sd(results1.men$Weight))
yfit <- yfit*diff(h$mids[1:2])*length(results1.men$Weight)
lines(xfit, yfit, col="blue", lwd=2)
```

Histogram with Normal Curve of Men's Weight



Weight

Men -

mean and standard deviation of weight

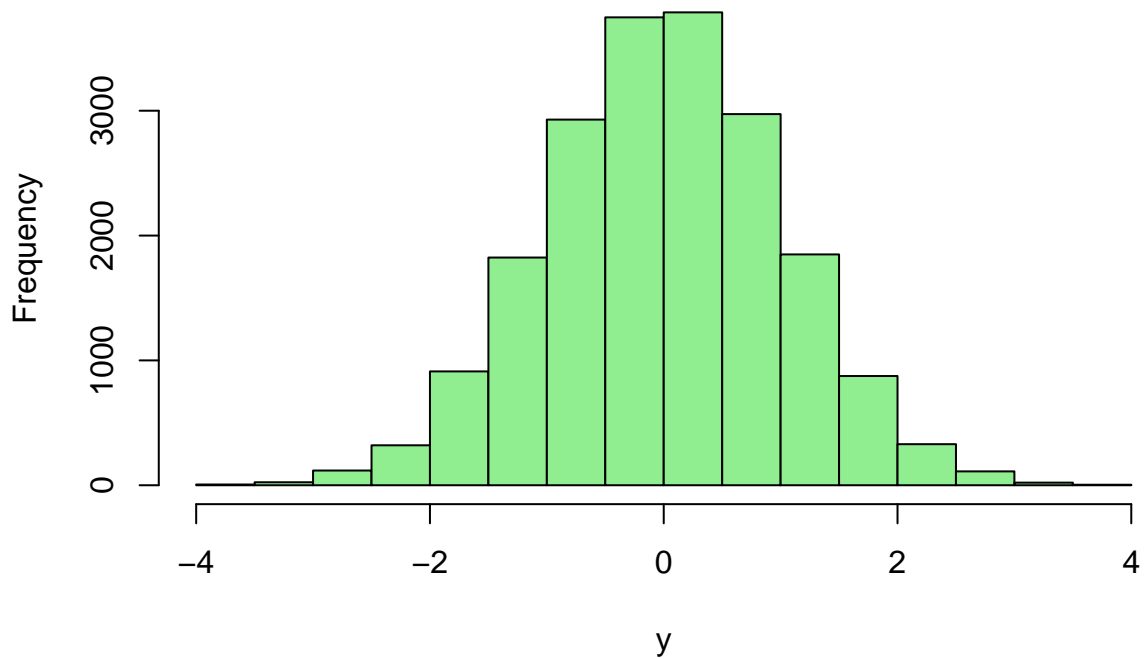
```
Ave <- mean(results1.men$Weight)
SD <- sd(results1.men$Weight)
c(Average = Ave, SD = SD)
```

```
## Average      SD
## 79.25132 13.96353
```

Men - normal distribution histogram of weight

```
y <- rnorm(results1.men$Weight)
hist(y, main = 'Weight Normal Distribution', col = "light green")
```


Weight Normal Distribution



```
late standard units
# calcu-
z <- scale(results1.men$Weight)

# calculate proportion of values within 2 SD of mean
mean(abs(z) < 2)

## [1] 0.9609198
```

Remove ID, Name, Sex, Team, NOC, Games, Year, Season, City, Medal columns

```
results2.men <- results1.men[, c(-1:-3, -7:-12, -14)]
str(results2.men)

## 'data.frame':    19831 obs. of  4 variables:
## $ Age   : int  28 28 28 28 28 32 20 20 22 22 ...
## $ Height: int  184 175 175 175 175 175 176 176 176 176 ...
## $ Weight: num   85 64 64 64 64 64 85 85 85 85 ...
## $ Sport : Factor w/ 52 levels "Alpine Skiing",...: 26 23 23 23 23 23 1 1 1 1 ...
```

Athlete_Events - randomForest analysis with Age, Height, Weight, and Team

```
results2.men.rf <- randomForest(results2.men[, -4], results2.men[, 4])
options(max.print=999999)
results2.men.rf
```

```

##
## Call:
## randomForest(x = results2.men[, -4], y = results2.men[, 4])
##           Type of random forest: classification
##           Number of trees: 500
## No. of variables tried at each split: 1
##
##           OOB estimate of  error rate: 70.96%
## Confusion matrix:
##
##           Alpine Skiing Archery Art Competitions Athletics
## Alpine Skiing           15         0             0         35
## Archery                  0         3             0         42
## Art Competitions         0         0             0          0
## Athletics                1         1             0        1546
## Badminton                0         0             0         45
## Baseball                 2         0             0         50
## Basketball               1         0             0         58
## Beach Volleyball         0         0             0          2
## Biathlon                 0         0             0        140
## Bobsleigh                1         0             0         22
## Boxing                   0         0             0        249
## Canoeing                 0         0             0        141
## Cross Country Skiing     0         0             0        200
## Curling                  0         0             0         13
## Cycling                  1         0             0        311
## Diving                   0         0             0         55
## Equestrianism            0         0             0        111
## Fencing                  0         0             0        310
## Figure Skating           0         0             0         69
## Football                 2         0             0        344
## Freestyle Skiing         0         0             0         10
## Golf                     0         0             0          0
## Gymnastics               0         0             0         87
## Handball                 0         0             0         48
## Hockey                   0         1             0        283
## Ice Hockey               3         0             0        116
## Judo                     0         0             0         73
## Lacrosse                 0         0             0          2
## Luge                     0         1             0         24
## Modern Pentathlon        0         0             0         86
## Nordic Combined          0         0             0         96
## Rowing                   0         0             0        196
## Rugby                    0         1             0          7
## Rugby Sevens             0         0             0          2
## Sailing                  2         1             0        173
## Shooting                 0         0             0         69
## Short Track Speed Skating 0         0             0         52
## Skeleton                 0         0             0          4
## Ski Jumping              0         0             0        104
## Snowboarding             0         0             0         12
## Speed Skating            0         0             0         77
## Swimming                 0         0             0        287
## Table Tennis             0         0             0         43
## Taekwondo                0         0             0         41

```

## Tennis	0	0	0	41
## Trampolining	0	0	0	7
## Triathlon	0	0	0	12
## Tug-Of-War	0	0	0	2
## Volleyball	0	0	0	69
## Water Polo	1	0	0	75
## Weightlifting	0	0	0	20
## Wrestling	1	2	0	152
##	Badminton	Baseball	Basketball	Beach Volleyball
## Alpine Skiing	0	0	0	0
## Archery	0	0	0	0
## Art Competitions	0	0	0	0
## Athletics	0	2	6	0
## Badminton	0	0	0	0
## Baseball	0	2	1	1
## Basketball	0	0	205	1
## Beach Volleyball	0	1	4	1
## Biathlon	0	0	0	0
## Bobsleigh	0	0	2	0
## Boxing	0	0	4	0
## Canoeing	0	0	1	0
## Cross Country Skiing	0	0	0	0
## Curling	0	0	0	0
## Cycling	1	0	2	0
## Diving	0	0	0	0
## Equestrianism	0	0	0	0
## Fencing	0	0	1	0
## Figure Skating	0	0	0	0
## Football	0	1	0	0
## Freestyle Skiing	0	0	0	0
## Golf	0	0	0	0
## Gymnastics	0	0	0	0
## Handball	0	0	26	0
## Hockey	0	1	0	0
## Ice Hockey	0	7	0	0
## Judo	0	1	2	0
## Lacrosse	0	0	0	0
## Luge	0	0	0	0
## Modern Pentathlon	0	0	0	0
## Nordic Combined	0	0	0	0
## Rowing	0	2	35	1
## Rugby	0	0	0	0
## Rugby Sevens	0	0	0	0
## Sailing	0	5	2	0
## Shooting	0	1	0	0
## Short Track Speed Skating	0	0	0	0
## Skeleton	0	0	0	0
## Ski Jumping	0	0	0	0
## Snowboarding	0	1	0	0
## Speed Skating	0	0	0	0
## Swimming	0	0	7	0
## Table Tennis	0	1	0	0
## Taekwondo	0	0	3	0
## Tennis	0	0	1	0

## Trampolineing	0	0	0	0
## Triathlon	0	0	0	0
## Tug-Of-War	0	0	0	0
## Volleyball	0	1	50	0
## Water Polo	0	2	16	0
## Weightlifting	0	6	1	0
## Wrestling	0	4	2	0
##	Biathlon	Bobsleigh	Boxing	Canoeing
## Alpine Skiing	0	1	0	6
## Archery	0	1	2	1
## Art Competitions	0	0	0	0
## Athletics	0	14	28	7
## Badminton	0	0	0	2
## Baseball	0	2	0	9
## Basketball	0	0	0	3
## Beach Volleyball	0	0	0	1
## Biathlon	11	0	0	3
## Bobsleigh	0	53	0	3
## Boxing	0	0	143	1
## Canoeing	3	3	0	54
## Cross Country Skiing	1	0	1	1
## Curling	0	1	0	2
## Cycling	2	3	4	8
## Diving	0	0	7	1
## Equestrianism	4	1	0	1
## Fencing	3	6	0	11
## Figure Skating	1	0	2	0
## Football	0	0	6	10
## Freestyle Skiing	0	0	0	1
## Golf	0	0	0	0
## Gymnastics	0	1	10	0
## Handball	0	0	2	6
## Hockey	4	1	6	18
## Ice Hockey	1	15	1	12
## Judo	0	4	1	3
## Lacrosse	0	0	0	0
## Luge	0	0	0	1
## Modern Pentathlon	1	0	0	1
## Nordic Combined	0	0	5	1
## Rowing	0	1	30	10
## Rugby	0	0	0	1
## Rugby Sevens	0	1	0	0
## Sailing	1	10	1	8
## Shooting	0	7	0	7
## Short Track Speed Skating	0	0	5	0
## Skeleton	2	0	0	0
## Ski Jumping	0	0	8	0
## Snowboarding	0	0	0	0
## Speed Skating	0	0	0	9
## Swimming	0	0	2	3
## Table Tennis	1	0	1	0
## Taekwondo	0	0	1	1
## Tennis	0	0	0	3
## Trampolineing	0	0	3	0

## Triathlon	0	0	0	0	
## Tug-Of-War	0	0	0	0	
## Volleyball	0	0	0	3	
## Water Polo	1	5	0	8	
## Weightlifting	0	0	0	4	
## Wrestling	0	13	12	9	
##	Cross Country	Skiing	Curling	Cycling	Diving
## Alpine Skiing		0	0	1	0
## Archery		0	0	1	0
## Art Competitions		0	0	0	0
## Athletics		11	1	9	0
## Badminton		1	0	2	0
## Baseball		2	0	1	0
## Basketball		1	0	3	0
## Beach Volleyball		0	0	0	0
## Biathlon		2	0	3	0
## Bobsleigh		0	0	1	0
## Boxing		0	0	5	1
## Canoeing		2	1	4	0
## Cross Country Skiing		39	0	2	0
## Curling		2	0	0	0
## Cycling		5	0	23	0
## Diving		2	0	1	6
## Equestrianism		8	0	1	1
## Fencing		9	0	0	0
## Figure Skating		3	0	0	0
## Football		2	0	6	0
## Freestyle Skiing		0	0	0	0
## Golf		0	0	0	0
## Gymnastics		3	0	0	2
## Handball		1	0	1	0
## Hockey		5	0	2	0
## Ice Hockey		1	0	7	0
## Judo		0	0	2	0
## Lacrosse		0	0	0	0
## Luge		0	0	0	0
## Modern Pentathlon		4	0	1	0
## Nordic Combined		0	0	1	0
## Rowing		3	0	3	1
## Rugby		0	0	0	0
## Rugby Sevens		0	0	0	0
## Sailing		4	0	2	0
## Shooting		4	1	2	0
## Short Track Speed Skating		0	0	0	0
## Skeleton		0	0	0	0
## Ski Jumping		0	0	1	0
## Snowboarding		0	0	0	0
## Speed Skating		3	0	2	0
## Swimming		0	0	5	0
## Table Tennis		1	0	0	0
## Taekwondo		0	0	0	0
## Tennis		0	0	0	0
## Trampolining		0	0	0	0
## Triathlon		0	0	0	0

## Tug-Of-War	0	0	0	0
## Volleyball	2	0	0	0
## Water Polo	2	0	0	0
## Weightlifting	1	0	2	0
## Wrestling	0	0	2	0
##	Equestrianism	Fencing	Figure Skating	Football
## Alpine Skiing	0	2	0	8
## Archery	4	1	0	6
## Art Competitions	1	0	0	0
## Athletics	15	37	0	58
## Badminton	1	4	0	2
## Baseball	4	13	0	14
## Basketball	0	7	0	2
## Beach Volleyball	1	0	0	0
## Biathlon	10	19	0	12
## Bobsleigh	5	8	0	0
## Boxing	1	6	1	29
## Canoeing	12	31	0	29
## Cross Country Skiing	18	14	0	14
## Curling	4	6	0	0
## Cycling	12	19	0	56
## Diving	2	2	0	14
## Equestrianism	226	25	1	8
## Fencing	34	135	0	29
## Figure Skating	6	6	1	9
## Football	4	33	1	141
## Freestyle Skiing	1	2	0	11
## Golf	0	0	0	0
## Gymnastics	5	0	0	20
## Handball	1	16	0	1
## Hockey	7	38	0	63
## Ice Hockey	4	33	0	39
## Judo	0	7	0	14
## Lacrosse	0	0	0	0
## Luge	2	5	0	7
## Modern Pentathlon	9	13	0	6
## Nordic Combined	1	0	0	3
## Rowing	7	19	0	10
## Rugby	0	1	0	5
## Rugby Sevens	0	1	0	1
## Sailing	58	39	0	10
## Shooting	46	28	0	12
## Short Track Speed Skating	0	2	0	24
## Skeleton	0	0	0	0
## Ski Jumping	0	1	0	4
## Snowboarding	0	0	0	2
## Speed Skating	1	15	0	19
## Swimming	0	5	1	22
## Table Tennis	1	1	0	4
## Taekwondo	0	1	0	1
## Tennis	2	2	0	2
## Trampolining	0	0	0	1
## Triathlon	1	0	0	1
## Tug-Of-War	0	0	0	0

## Volleyball	4	19	0	4	
## Water Polo	0	14	0	7	
## Weightlifting	1	1	0	10	
## Wrestling	7	13	0	27	
##	Freestyle Skiing	Golf	Gymnastics	Handball	Hockey
## Alpine Skiing	0	0	7	0	0
## Archery	0	0	2	1	0
## Art Competitions	0	0	0	0	0
## Athletics	0	0	64	8	11
## Badminton	0	0	4	0	3
## Baseball	0	0	0	6	3
## Basketball	0	0	0	7	0
## Beach Volleyball	0	0	0	0	0
## Biathlon	0	0	4	0	4
## Bobsleigh	0	0	0	4	4
## Boxing	0	0	56	3	5
## Canoeing	0	0	8	4	4
## Cross Country Skiing	0	0	18	1	9
## Curling	0	0	1	0	2
## Cycling	0	0	8	1	8
## Diving	0	0	42	0	3
## Equestrianism	0	0	6	0	3
## Fencing	0	0	9	3	13
## Figure Skating	0	0	5	0	2
## Football	0	0	37	1	10
## Freestyle Skiing	0	0	2	0	3
## Golf	0	0	0	0	0
## Gymnastics	0	0	457	0	7
## Handball	0	0	1	52	0
## Hockey	1	0	43	2	16
## Ice Hockey	0	0	6	9	10
## Judo	0	0	35	1	5
## Lacrosse	0	0	0	0	0
## Luge	0	0	4	2	4
## Modern Pentathlon	0	0	1	0	3
## Nordic Combined	0	0	6	0	0
## Rowing	0	0	5	30	3
## Rugby	0	0	1	0	1
## Rugby Sevens	0	0	0	3	0
## Sailing	0	0	5	5	4
## Shooting	0	0	9	1	2
## Short Track Speed Skating	0	0	13	0	6
## Skeleton	0	0	0	0	0
## Ski Jumping	0	0	4	0	0
## Snowboarding	0	0	5	0	0
## Speed Skating	0	0	4	1	7
## Swimming	0	0	12	3	0
## Table Tennis	0	0	10	0	3
## Taekwondo	0	0	1	0	1
## Tennis	0	0	3	1	0
## Trampoline	0	0	0	0	1
## Triathlon	0	0	0	0	0
## Tug-Of-War	0	0	0	0	0
## Volleyball	0	0	0	2	3

## Water Polo	0	0	2	28	2
## Weightlifting	0	0	40	0	1
## Wrestling	0	0	131	3	7
##	Ice Hockey	Judo	Lacrosse	Luge	Modern Pentathlon
## Alpine Skiing	68	0	0	0	0
## Archery	15	0	0	0	0
## Art Competitions	0	0	0	0	0
## Athletics	133	12	0	0	1
## Badminton	4	0	0	0	0
## Baseball	111	3	0	0	0
## Basketball	22	0	0	0	0
## Beach Volleyball	4	0	0	0	0
## Biathlon	4	0	0	0	1
## Bobsleigh	95	2	0	0	0
## Boxing	36	0	0	0	0
## Canoeing	215	0	0	0	1
## Cross Country Skiing	14	0	0	0	1
## Curling	18	0	0	0	1
## Cycling	88	0	0	0	0
## Diving	0	0	0	0	0
## Equestrianism	16	0	0	0	2
## Fencing	66	0	0	0	1
## Figure Skating	6	0	0	0	0
## Football	47	1	0	0	0
## Freestyle Skiing	11	0	0	0	0
## Golf	0	0	0	0	0
## Gymnastics	1	2	0	0	0
## Handball	94	1	0	1	0
## Hockey	93	1	0	2	0
## Ice Hockey	472	3	0	1	0
## Judo	77	13	0	0	0
## Lacrosse	0	0	0	0	0
## Luge	60	0	0	3	0
## Modern Pentathlon	3	0	0	0	4
## Nordic Combined	0	0	0	0	0
## Rowing	136	0	0	0	1
## Rugby	7	2	0	0	0
## Rugby Sevens	19	0	0	0	0
## Sailing	83	0	0	0	1
## Shooting	71	5	0	0	0
## Short Track Speed Skating	7	0	0	1	0
## Skeleton	2	0	0	0	0
## Ski Jumping	0	1	0	0	0
## Snowboarding	16	0	0	0	0
## Speed Skating	32	0	0	0	0
## Swimming	43	0	0	0	0
## Table Tennis	2	0	0	0	0
## Taekwondo	2	0	0	0	0
## Tennis	4	0	0	0	0
## Trampolining	1	0	0	0	0
## Triathlon	0	0	0	0	0
## Tug-Of-War	1	0	0	0	0
## Volleyball	22	0	0	0	0
## Water Polo	94	1	0	0	0

## Weightlifting	35	6	0	1		0	
## Wrestling	132	8	0	0		0	
##	Nordic	Combined	Rowing	Rugby	Rugby	Sevens	Sailing
## Alpine Skiing		0	9	0		0	2
## Archery		0	1	1		0	5
## Art Competitions		0	0	0		0	0
## Athletics		0	127	0		0	13
## Badminton		0	4	0		0	1
## Baseball		0	48	0		0	7
## Basketball		0	174	0		0	2
## Beach Volleyball		0	13	0		0	0
## Biathlon		0	2	0		0	5
## Bobsleigh		0	18	0		0	6
## Boxing		0	57	0		0	0
## Canoeing		0	135	0		0	6
## Cross Country Skiing		0	14	0		0	3
## Curling		0	2	0		0	6
## Cycling		0	19	0		0	5
## Diving		0	3	0		0	0
## Equestrianism		0	10	0		0	19
## Fencing		0	56	0		0	10
## Figure Skating		0	3	0		0	1
## Football		0	26	0		0	2
## Freestyle Skiing		0	2	0		0	0
## Golf		0	1	0		0	1
## Gymnastics		0	3	0		0	0
## Handball		0	190	0		0	3
## Hockey		1	35	0		0	5
## Ice Hockey		0	123	0		0	14
## Judo		0	19	1		0	1
## Lacrosse		0	0	0		0	0
## Luge		0	7	0		0	5
## Modern Pentathlon		0	2	0		0	0
## Nordic Combined		8	2	0		0	1
## Rowing		0	665	0		0	12
## Rugby		0	0	0		0	0
## Rugby Sevens		0	8	0		0	0
## Sailing		0	42	0		0	29
## Shooting		0	11	0		0	12
## Short Track Speed Skating		0	1	0		0	0
## Skeleton		0	3	0		0	1
## Ski Jumping		0	1	0		0	0
## Snowboarding		0	1	0		0	1
## Speed Skating		0	29	0		0	0
## Swimming		1	160	0		0	3
## Table Tennis		0	3	0		0	1
## Taekwondo		0	12	0		0	0
## Tennis		0	16	0		0	0
## Trampoline		0	0	0		0	0
## Triathlon		0	1	0		0	0
## Tug-Of-War		0	1	0		0	0
## Volleyball		0	174	0		0	2
## Water Polo		0	210	0		0	5
## Weightlifting		0	1	0		0	5

## Wrestling	0	41	0	0	5
##	Shooting	Short Track	Speed Skating	Skeleton	
## Alpine Skiing	0		0	0	
## Archery	0		0	0	
## Art Competitions	0		0	0	
## Athletics	4		1	0	
## Badminton	0		1	0	
## Baseball	5		0	0	
## Basketball	0		0	0	
## Beach Volleyball	0		0	0	
## Biathlon	0		1	0	
## Bobsleigh	6		0	0	
## Boxing	0		1	0	
## Canoeing	5		1	0	
## Cross Country Skiing	0		0	0	
## Curling	4		0	0	
## Cycling	0		0	0	
## Diving	3		0	0	
## Equestrianism	11		1	0	
## Fencing	5		0	0	
## Figure Skating	0		0	0	
## Football	1		4	0	
## Freestyle Skiing	0		0	0	
## Golf	0		0	0	
## Gymnastics	0		1	0	
## Handball	1		0	0	
## Hockey	2		4	0	
## Ice Hockey	9		1	0	
## Judo	3		0	0	
## Lacrosse	0		0	0	
## Luge	1		1	0	
## Modern Pentathlon	0		0	0	
## Nordic Combined	0		0	0	
## Rowing	1		0	0	
## Rugby	0		0	0	
## Rugby Sevens	0		0	0	
## Sailing	25		1	0	
## Shooting	59		0	0	
## Short Track Speed Skating	0		7	0	
## Skeleton	0		0	0	
## Ski Jumping	0		0	0	
## Snowboarding	0		0	0	
## Speed Skating	3		0	0	
## Swimming	0		2	0	
## Table Tennis	2		0	0	
## Taekwondo	0		0	0	
## Tennis	1		0	0	
## Trampoline	1		0	0	
## Triathlon	0		0	0	
## Tug-Of-War	0		0	0	
## Volleyball	0		0	0	
## Water Polo	3		0	0	
## Weightlifting	8		0	0	
## Wrestling	12		0	0	

##	Ski Jumping	Snowboarding	Speed Skating	Swimming
## Alpine Skiing	0	0	0	9
## Archery	0	0	0	12
## Art Competitions	0	0	0	0
## Athletics	8	0	5	194
## Badminton	0	0	1	4
## Baseball	0	1	0	23
## Basketball	0	0	0	84
## Beach Volleyball	0	0	0	2
## Biathlon	0	0	0	12
## Bobsleigh	0	0	0	2
## Boxing	8	0	0	31
## Canoeing	0	0	1	78
## Cross Country Skiing	0	0	0	8
## Curling	0	0	0	1
## Cycling	1	0	0	72
## Diving	0	0	0	8
## Equestrianism	1	0	0	4
## Fencing	1	0	0	53
## Figure Skating	0	0	0	7
## Football	0	0	1	79
## Freestyle Skiing	0	0	0	3
## Golf	0	0	0	0
## Gymnastics	0	1	1	7
## Handball	0	0	1	37
## Hockey	0	0	2	35
## Ice Hockey	0	0	1	50
## Judo	1	0	0	7
## Lacrosse	0	0	0	0
## Luge	0	0	0	3
## Modern Pentathlon	0	0	0	2
## Nordic Combined	2	0	0	9
## Rowing	2	0	2	167
## Rugby	0	0	0	0
## Rugby Sevens	0	0	0	0
## Sailing	3	0	0	12
## Shooting	0	0	3	8
## Short Track Speed Skating	0	0	0	10
## Skeleton	0	0	0	0
## Ski Jumping	41	0	0	6
## Snowboarding	0	0	0	5
## Speed Skating	0	0	22	26
## Swimming	2	0	1	740
## Table Tennis	0	0	0	2
## Taekwondo	1	0	0	7
## Tennis	0	0	0	9
## Trampoline	0	0	0	0
## Triathlon	0	0	0	0
## Tug-Of-War	0	0	0	1
## Volleyball	0	0	0	56
## Water Polo	0	0	0	63
## Weightlifting	1	0	1	2
## Wrestling	1	0	0	7
##	Table Tennis	Taekwondo	Tennis	Trampoline
##	Triathlon			

## Alpine Skiing	0	0	0	0	0
## Archery	0	0	0	0	0
## Art Competitions	0	0	0	0	0
## Athletics	0	0	0	0	0
## Badminton	0	0	0	0	0
## Baseball	0	0	0	0	0
## Basketball	0	1	0	0	0
## Beach Volleyball	0	0	0	0	0
## Biathlon	0	0	0	0	0
## Bobsleigh	0	0	0	0	0
## Boxing	0	0	0	0	0
## Canoeing	0	0	0	0	0
## Cross Country Skiing	0	0	0	0	0
## Curling	0	0	0	0	0
## Cycling	0	0	0	0	0
## Diving	0	0	0	0	0
## Equestrianism	0	0	0	0	0
## Fencing	0	1	0	0	0
## Figure Skating	0	0	0	0	0
## Football	0	0	0	0	0
## Freestyle Skiing	0	0	0	0	0
## Golf	0	0	0	0	0
## Gymnastics	0	0	0	0	0
## Handball	0	0	0	0	0
## Hockey	0	0	0	0	0
## Ice Hockey	0	0	0	0	0
## Judo	0	0	0	0	0
## Lacrosse	0	0	0	0	0
## Luge	0	0	0	0	0
## Modern Pentathlon	0	0	0	0	0
## Nordic Combined	0	0	0	0	0
## Rowing	0	0	0	0	0
## Rugby	0	0	0	0	0
## Rugby Sevens	0	0	0	0	0
## Sailing	0	0	0	0	0
## Shooting	0	0	0	0	0
## Short Track Speed Skating	0	0	0	0	0
## Skeleton	0	0	0	0	0
## Ski Jumping	0	0	0	0	0
## Snowboarding	0	0	0	0	0
## Speed Skating	0	0	0	0	0
## Swimming	0	0	0	0	0
## Table Tennis	2	0	0	0	0
## Taekwondo	0	0	0	0	0
## Tennis	0	0	2	0	0
## Trampoline	0	0	0	0	0
## Triathlon	0	0	0	0	0
## Tug-Of-War	0	0	0	0	0
## Volleyball	0	0	0	0	0
## Water Polo	0	0	0	0	0
## Weightlifting	1	0	0	0	0
## Wrestling	0	0	0	0	0
##					
## Alpine Skiing	0	0	1	3	

## Archery	0	0	0	2
## Art Competitions	0	0	0	0
## Athletics	0	9	2	18
## Badminton	0	0	0	0
## Baseball	0	0	0	9
## Basketball	0	33	6	0
## Beach Volleyball	0	4	0	0
## Biathlon	0	3	0	0
## Bobsleigh	0	0	1	5
## Boxing	0	1	1	7
## Canoeing	0	6	3	2
## Cross Country Skiing	0	3	0	0
## Curling	0	0	0	1
## Cycling	0	0	0	7
## Diving	0	0	0	1
## Equestrianism	0	1	0	0
## Fencing	0	6	0	1
## Figure Skating	0	0	0	1
## Football	0	1	1	3
## Freestyle Skiing	0	0	0	2
## Golf	0	0	0	0
## Gymnastics	0	0	0	5
## Handball	0	8	15	1
## Hockey	0	3	1	9
## Ice Hockey	0	1	6	13
## Judo	0	0	0	24
## Lacrosse	0	0	0	0
## Luge	0	0	0	0
## Modern Pentathlon	0	1	0	0
## Nordic Combined	0	0	0	0
## Rowing	0	24	11	1
## Rugby	0	0	0	2
## Rugby Sevens	0	0	1	2
## Sailing	0	2	1	7
## Shooting	0	0	1	16
## Short Track Speed Skating	0	0	0	0
## Skeleton	0	0	0	0
## Ski Jumping	0	0	0	0
## Snowboarding	0	0	0	0
## Speed Skating	0	0	0	1
## Swimming	0	12	0	1
## Table Tennis	0	2	0	1
## Taekwondo	0	0	0	0
## Tennis	0	2	0	0
## Trampoline	0	0	0	0
## Triathlon	0	0	0	0
## Tug-Of-War	0	0	0	1
## Volleyball	0	74	3	0
## Water Polo	0	6	16	2
## Weightlifting	0	0	1	220
## Wrestling	1	0	6	49
##	Wrestling class.error			
## Alpine Skiing	7	0.9137931		
## Archery	6	0.9716981		

## Art Competitions	1	1.0000000
## Athletics	63	0.3582399
## Badminton	2	1.0000000
## Baseball	16	0.9939940
## Basketball	0	0.6639344
## Beach Volleyball	0	0.9696970
## Biathlon	5	0.9543568
## Bobsleigh	21	0.7953668
## Boxing	39	0.7912409
## Canoeing	20	0.9298701
## Cross Country Skiing	10	0.8948787
## Curling	1	1.0000000
## Cycling	13	0.9656203
## Diving	7	0.9617834
## Equestrianism	10	0.5201699
## Fencing	20	0.8275862
## Figure Skating	2	0.9919355
## Football	19	0.8199234
## Freestyle Skiing	3	1.0000000
## Golf	0	1.0000000
## Gymnastics	40	0.3001531
## Handball	9	0.8992248
## Hockey	30	0.9775910
## Ice Hockey	43	0.5284715
## Judo	44	0.9615385
## Lacrosse	0	1.0000000
## Luge	3	0.9774436
## Modern Pentathlon	1	0.9710145
## Nordic Combined	2	0.9416058
## Rowing	18	0.5236390
## Rugby	2	1.0000000
## Rugby Sevens	0	1.0000000
## Sailing	16	0.9474638
## Shooting	22	0.8513854
## Short Track Speed Skating	10	0.9492754
## Skeleton	0	1.0000000
## Ski Jumping	1	0.7616279
## Snowboarding	1	1.0000000
## Speed Skating	1	0.9126984
## Swimming	7	0.4389689
## Table Tennis	2	0.9759036
## Taekwondo	0	1.0000000
## Tennis	0	0.9775281
## Trampolining	1	1.0000000
## Triathlon	0	1.0000000
## Tug-Of-War	2	1.0000000
## Volleyball	1	0.8486708
## Water Polo	10	0.9720768
## Weightlifting	58	0.4847775
## Wrestling	252	0.7196885

```
importance(results2.men.rf)
```

```
##           MeanDecreaseGini
## Age           1591.764
```

```
## Height      2124.690
## Weight      2257.845
```

randomForest analysis for the following Sport

(Gymnastics, Athletics, Swimming, Weightlifting, and Equestrianism)

based on Men's Age, Height, Weight, and Team

```
length(which(olympics$Sport == "Gymnastics"))
```

```
## [1] 1161
```

```
olympics.gymnastics <- olympics[olympics$Sport == "Gymnastics", ]
str(olympics.gymnastics)
```

```
## 'data.frame':    1161 obs. of  15 variables:
## $ ID      : int  17 17 17 17 17 394 455 455 455 455 ...
## $ Name    : chr  "Paavo Johannes Aaltonen" "Paavo Johannes Aaltonen" "Paavo Johannes Aaltonen" "Paavo
## $ Sex     : chr  "M" "M" "M" "M" ...
## $ Age     : int  28 28 28 28 32 19 19 19 24 24 ...
## $ Height  : int  175 175 175 175 175 160 161 161 161 161 ...
## $ Weight  : num  64 64 64 64 64 48 62 62 62 62 ...
## $ Team    : chr  "Finland" "Finland" "Finland" "Finland" ...
## $ NOC     : chr  "FIN" "FIN" "FIN" "FIN" ...
## $ Games   : chr  "1948 Summer" "1948 Summer" "1948 Summer" "1948 Summer" ...
## $ Year    : int  1948 1948 1948 1948 1952 1972 2012 2012 2016 2016 ...
## $ Season  : chr  "Summer" "Summer" "Summer" "Summer" ...
## $ City    : chr  "London" "London" "London" "London" ...
## $ Sport   : chr  "Gymnastics" "Gymnastics" "Gymnastics" "Gymnastics" ...
## $ Event   : chr  "Gymnastics Men's Individual All-Around" "Gymnastics Men's Team All-Around" "Gymnast
## $ Medal   : chr  "Bronze" "Gold" "Gold" "Gold" ...
## - attr(*, "na.action")= 'omit' Named int [1:240935] 1 2 3 4 5 6 7 8 9 10 ...
## ..- attr(*, "names")= chr [1:240935] "1" "2" "3" "4" ...
```

```
length(which(olympics$Sport == "Athletics"))
```

```
## [1] 3648
```

```
olympics.athletics <- olympics[olympics$Sport == "Athletics", ]
str(olympics.athletics)
```

```
## 'data.frame':    3648 obs. of  15 variables:
## $ ID      : int  67 86 359 379 411 416 416 428 454 519 ...
## $ Name    : chr  "Mariya Vasilyevna Abakumova (-Tarabina)" "Jos Manuel Abascal Gmez" "Andrey Khakimov
## $ Sex     : chr  "F" "M" "M" "M" ...
## $ Age     : int  22 26 26 21 22 19 19 25 19 24 ...
## $ Height  : int  179 182 190 160 166 181 181 159 182 183 ...
## $ Weight  : num  80 67 125 50 58 78 78 40 68 75 ...
## $ Team    : chr  "Russia" "Spain" "Unified Team" "Ethiopia" ...
## $ NOC     : chr  "RUS" "ESP" "EUN" "ETH" ...
## $ Games   : chr  "2008 Summer" "1984 Summer" "1992 Summer" "1992 Summer" ...
## $ Year    : int  2008 1984 1992 1992 2000 1912 1912 2008 1932 1924 ...
```

```
## $ Season: chr "Summer" "Summer" "Summer" "Summer" ...
## $ City : chr "Beijing" "Los Angeles" "Barcelona" "Barcelona" ...
## $ Sport : chr "Athletics" "Athletics" "Athletics" "Athletics" ...
## $ Event : chr "Athletics Women's Javelin Throw" "Athletics Men's 1,500 metres" "Athletics Men's Ha
## $ Medal : chr "Silver" "Bronze" "Gold" "Bronze" ...
## - attr(*, "na.action")= 'omit' Named int [1:240935] 1 2 3 4 5 6 7 8 9 10 ...
## ..- attr(*, "names")= chr [1:240935] "1" "2" "3" "4" ...
```

```
length(which(olympics$Sport == "Swimming"))
```

```
## [1] 2486
```

```
olympics.swimming <- olympics[olympics$Sport == "Swimming", ]
str(olympics.swimming)
```

```
## 'data.frame': 2486 obs. of 15 variables:
## $ ID : int 259 460 460 465 509 980 980 980 980 1017 ...
## $ Name : chr "Reema Abdo" "Viktor Andreyevich Aboimov" "Viktor Andreyevich Aboimov" "Matthew \"Ma
## $ Sex : chr "F" "M" "M" "M" ...
## $ Age : int 21 22 22 30 21 19 19 23 23 19 ...
## $ Height: int 173 190 190 197 175 179 179 179 179 198 ...
## $ Weight: num 59 78 78 92 64 70 70 70 70 100 ...
## $ Team : chr "Canada" "Soviet Union" "Soviet Union" "Australia" ...
## $ NOC : chr "CAN" "URS" "URS" "AUS" ...
## $ Games : chr "1984 Summer" "1972 Summer" "1972 Summer" "2016 Summer" ...
## $ Year : int 1984 1972 1972 2016 1980 2008 2008 2012 2012 2008 ...
## $ Season: chr "Summer" "Summer" "Summer" "Summer" ...
## $ City : chr "Los Angeles" "Munich" "Munich" "Rio de Janeiro" ...
## $ Sport : chr "Swimming" "Swimming" "Swimming" "Swimming" ...
## $ Event : chr "Swimming Women's 4 x 100 metres Medley Relay" "Swimming Men's 4 x 100 metres Freesty
## $ Medal : chr "Bronze" "Silver" "Bronze" "Bronze" ...
## - attr(*, "na.action")= 'omit' Named int [1:240935] 1 2 3 4 5 6 7 8 9 10 ...
## ..- attr(*, "names")= chr [1:240935] "1" "2" "3" "4" ...
```

```
length(which(olympics$Sport == "Weightlifting"))
```

```
## [1] 532
```

```
olympics.weightlifting <- olympics[olympics$Sport == "Weightlifting", ]
str(olympics.weightlifting)
```

```
## 'data.frame': 532 obs. of 15 variables:
## $ ID : int 1301 1698 1698 2306 2483 2511 2511 3854 4013 4449 ...
## $ Name : chr "Sri Wahyuni Agustiani" "Khadzhimurat Magomedovich Akkayev" "Khadzhimurat Magomedovi
## $ Sex : chr "F" "M" "M" "M" ...
## $ Age : int 21 19 23 24 20 30 34 24 26 28 ...
## $ Height: int 147 178 178 192 176 185 185 178 154 192 ...
## $ Weight: num 47 105 105 156 89 160 160 141 59 105 ...
## $ Team : chr "Indonesia" "Russia" "Russia" "Russia" ...
## $ NOC : chr "INA" "RUS" "RUS" "RUS" ...
## $ Games : chr "2016 Summer" "2004 Summer" "2008 Summer" "2012 Summer" ...
## $ Year : int 2016 2004 2008 2012 1980 1972 1976 1956 1976 2012 ...
## $ Season: chr "Summer" "Summer" "Summer" "Summer" ...
## $ City : chr "Rio de Janeiro" "Athina" "Beijing" "London" ...
## $ Sport : chr "Weightlifting" "Weightlifting" "Weightlifting" "Weightlifting" ...
## $ Event : chr "Weightlifting Women's Flyweight" "Weightlifting Men's Middle-Heavyweight" "Weightli
## $ Medal : chr "Silver" "Silver" "Bronze" "Bronze" ...
```



```
## - attr(*, "na.action")= 'omit' Named int [1:240935] 1 2 3 4 5 6 7 8 9 10 ...
## ..- attr(*, "names")= chr [1:240935] "1" "2" "3" "4" ...

length(which(olympics$Sport == "Equestrianism"))

## [1] 663

olympics.equestrian <- olympics[olympics$Sport == "Equestrianism", ]
str(olympics.equestrian)

## 'data.frame': 663 obs. of 15 variables:
## $ ID : int 1356 1356 1885 1889 2134 2651 2894 2894 3477 4335 ...
## $ Name : chr "Christian Ahlmann" "Christian Ahlmann" "Ramzy Al-Duhami" "Khaled Al-Eid" ...
## $ Sex : chr "M" "M" "M" "M" ...
## $ Age : int 29 41 40 31 27 37 54 54 43 26 ...
## $ Height: int 189 189 162 175 178 163 180 180 183 177 ...
## $ Weight: num 80 80 70 57 69 56 71 71 71 64 ...
## $ Team : chr "Germany" "Germany" "Saudi Arabia" "Saudi Arabia" ...
## $ NOC : chr "GER" "GER" "KSA" "KSA" ...
## $ Games : chr "2004 Summer" "2016 Summer" "2012 Summer" "2000 Summer" ...
## $ Year : int 2004 2016 2012 2000 2012 2012 1968 1968 1976 1964 ...
## $ Season: chr "Summer" "Summer" "Summer" "Summer" ...
## $ City : chr "Athina" "Rio de Janeiro" "London" "Sydney" ...
## $ Sport : chr "Equestrianism" "Equestrianism" "Equestrianism" "Equestrianism" ...
## $ Event : chr "Equestrianism Mixed Jumping, Team" "Equestrianism Mixed Jumping, Team" "Equestrianism Mixed Jumping, Team" ...
## $ Medal : chr "Bronze" "Bronze" "Bronze" "Bronze" ...
## - attr(*, "na.action")= 'omit' Named int [1:240935] 1 2 3 4 5 6 7 8 9 10 ...
## ..- attr(*, "names")= chr [1:240935] "1" "2" "3" "4" ...

olympics.combine <- rbind(olympics.gymnastics, olympics.swimming, olympics.weightlifting,
                          olympics.athletics, olympics.equestrian)
str(olympics.combine)

## 'data.frame': 8490 obs. of 15 variables:
## $ ID : int 17 17 17 17 17 394 455 455 455 455 ...
## $ Name : chr "Paavo Johannes Aaltonen" "Paavo Johannes Aaltonen" "Paavo Johannes Aaltonen" "Paavo Johannes Aaltonen" ...
## $ Sex : chr "M" "M" "M" "M" ...
## $ Age : int 28 28 28 28 32 19 19 19 24 24 ...
## $ Height: int 175 175 175 175 175 160 161 161 161 161 ...
## $ Weight: num 64 64 64 64 64 48 62 62 62 62 ...
## $ Team : chr "Finland" "Finland" "Finland" "Finland" ...
## $ NOC : chr "FIN" "FIN" "FIN" "FIN" ...
## $ Games : chr "1948 Summer" "1948 Summer" "1948 Summer" "1948 Summer" ...
## $ Year : int 1948 1948 1948 1948 1952 1972 2012 2012 2016 2016 ...
## $ Season: chr "Summer" "Summer" "Summer" "Summer" ...
## $ City : chr "London" "London" "London" "London" ...
## $ Sport : chr "Gymnastics" "Gymnastics" "Gymnastics" "Gymnastics" ...
## $ Event : chr "Gymnastics Men's Individual All-Around" "Gymnastics Men's Team All-Around" "Gymnastics Men's Individual All-Around" ...
## $ Medal : chr "Bronze" "Gold" "Gold" "Gold" ...
## - attr(*, "na.action")= 'omit' Named int [1:240935] 1 2 3 4 5 6 7 8 9 10 ...
## ..- attr(*, "names")= chr [1:240935] "1" "2" "3" "4" ...

olympics.combine1 <- olympics.combine[, c(-1:-3, -7:-12, -14, -15)]
str(olympics.combine1)

## 'data.frame': 8490 obs. of 4 variables:
## $ Age : int 28 28 28 28 32 19 19 19 24 24 ...
```

```
## $ Height: int 175 175 175 175 175 160 161 161 161 161 ...
## $ Weight: num 64 64 64 64 64 48 62 62 62 62 ...
## $ Sport : chr "Gymnastics" "Gymnastics" "Gymnastics" "Gymnastics" ...

olympics.combine1$Sport <- as.factor(olympics.combine1$Sport)
str(olympics.combine1)

## 'data.frame': 8490 obs. of 4 variables:
## $ Age : int 28 28 28 28 32 19 19 19 24 24 ...
## $ Height: int 175 175 175 175 175 160 161 161 161 161 ...
## $ Weight: num 64 64 64 64 64 48 62 62 62 62 ...
## $ Sport : Factor w/ 5 levels "Athletics","Equestrianism",...: 3 3 3 3 3 3 3 3 3 3 ...

olympics.combine1.rf <- randomForest(olympics.combine1[, -4], olympics.combine1[, 4])
olympics.combine1.rf

##
## Call:
## randomForest(x = olympics.combine1[, -4], y = olympics.combine1[, 4])
## Type of random forest: classification
## Number of trees: 500
## No. of variables tried at each split: 1
##
## OOB estimate of error rate: 27.14%
## Confusion matrix:
## Athletics Equestrianism Gymnastics Swimming Weightlifting
## Athletics 2809 50 154 558 77
## Equestrianism 246 379 13 21 4
## Gymnastics 212 6 862 65 16
## Swimming 629 6 54 1792 5
## Weightlifting 104 8 54 22 344
## class.error
## Athletics 0.2299890
## Equestrianism 0.4283560
## Gymnastics 0.2575366
## Swimming 0.2791633
## Weightlifting 0.3533835

importance(olympics.combine1.rf)

## MeanDecreaseGini
## Age 1342.259
## Height 1215.336
## Weight 1116.341
```

Predicting Sport based on randomForest analysis

Row 1000 & 4000

```
olympics.combine1[1000,]

## Age Height Weight Sport
## 244057 26 164 59 Gymnastics
```

```
olympics.combine1[4000,]
```

```
##      Age Height Weight      Sport
## 197817  34   181   105 Weightlifting
```

```
#Predicting Row 1000 & 4000
```

```
predict(olympics.combine1.rf, olympics.combine1[1000,-5])
```

```
##      244057
```

```
## Gymnastics
```

```
## Levels: Athletics Equestrianism Gymnastics Swimming Weightlifting
```

```
predict(olympics.combine1.rf, olympics.combine1[4000,-5])
```

```
##      197817
```

```
## Weightlifting
```

```
## Levels: Athletics Equestrianism Gymnastics Swimming Weightlifting
```

Predicting for Age = 30, Height = 170, Weight = 50

```
predict(olympics.combine1.rf, c(30, 170, 50))
```

```
##      1
```

```
## Athletics
```

```
## Levels: Athletics Equestrianism Gymnastics Swimming Weightlifting
```

Heatmap with relative change equally across all variables

Rescale Age, Height and Weight 0 to 100

```
str(olympics.combine1)
```

```
## 'data.frame': 8490 obs. of 4 variables:
```

```
## $ Age : int 28 28 28 28 32 19 19 19 24 24 ...
```

```
## $ Height: int 175 175 175 175 175 160 161 161 161 161 ...
```

```
## $ Weight: num 64 64 64 64 64 48 62 62 62 62 ...
```

```
## $ Sport : Factor w/ 5 levels "Athletics","Equestrianism",...: 3 3 3 3 3 3 3 3 3 3 ...
```

```
olympics.combine1$Age <- rescale(olympics.combine1$Age, to = c(0, 100))
```

```
olympics.combine1$Height <- rescale(olympics.combine1$Height, to = c(0, 100))
```

```
olympics.combine1$Weight <- rescale(olympics.combine1$Weight, to = c(0, 100))
```

```
str(olympics.combine1)
```

```
## 'data.frame': 8490 obs. of 4 variables:
```

```
## $ Age : num 31.2 31.2 31.2 31.2 39.6 ...
```

```
## $ Height: num 54.2 54.2 54.2 54.2 54.2 ...
```

```
## $ Weight: num 24.3 24.3 24.3 24.3 24.3 ...
```

```
## $ Sport : Factor w/ 5 levels "Athletics","Equestrianism",...: 3 3 3 3 3 3 3 3 3 3 ...
```

Melt Age, Weight, Height against Sport

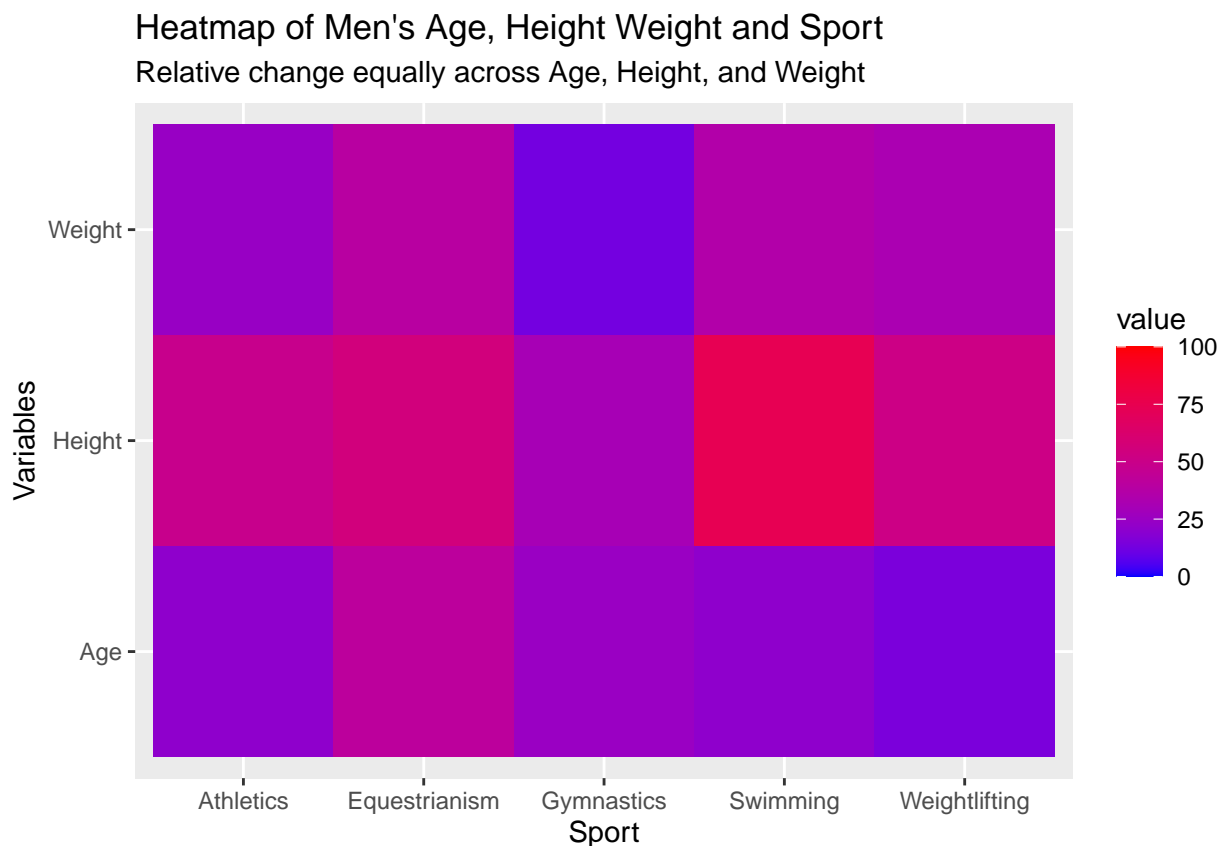
```

melt <- reshape::melt(olympics.combine1, id=c("Sport"))
str(melt)

## 'data.frame': 25470 obs. of 3 variables:
## $ Sport : Factor w/ 5 levels "Athletics","Equestrianism",...: 3 3 3 3 3 3 3 3 3 3 ...
## $ variable: Factor w/ 3 levels "Age","Height",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ value : num 31.2 31.2 31.2 31.2 39.6 ...

gplot <- ggplot(melt, aes(x = Sport, y = variable,
                          group = variable, fill = value))
gplot <- gplot + geom_tile()
gplot <- gplot + scale_fill_gradient(low = 'blue', high = 'red')
gplot <- gplot + ggtitle("Heatmap of Men's Age, Height Weight and Sport",
                        subtitle = "Relative change equally across Age, Height, and Weight")
gplot <- gplot + labs(x = "Sport", y='Variables')
gplot

```



Scatter plot Men's Age, Height, Weight,
and Sport (Athletics, Equestrianism, Gymnastics, Swimming, &
Weightlifting)

```

str(olympics.combine)

## 'data.frame': 8490 obs. of 15 variables:
## $ ID : int 17 17 17 17 17 394 455 455 455 455 ...

```

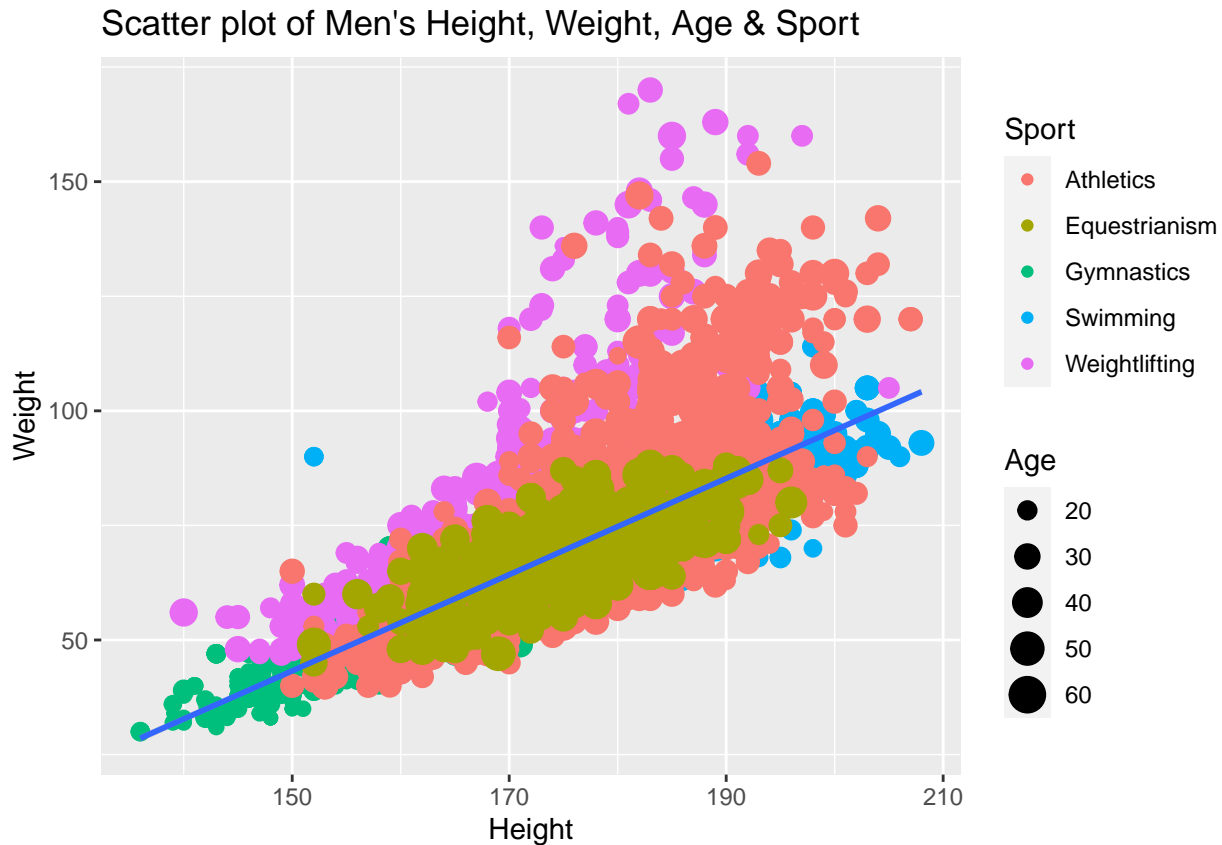
```
## $ Name : chr "Paavo Johannes Aaltonen" "Paavo Johannes Aaltonen" "Paavo Johannes Aaltonen" "Paavo
## $ Sex : chr "M" "M" "M" "M" ...
## $ Age : int 28 28 28 28 32 19 19 19 24 24 ...
## $ Height: int 175 175 175 175 175 160 161 161 161 161 ...
## $ Weight: num 64 64 64 64 64 48 62 62 62 62 ...
## $ Team : chr "Finland" "Finland" "Finland" "Finland" ...
## $ NOC : chr "FIN" "FIN" "FIN" "FIN" ...
## $ Games : chr "1948 Summer" "1948 Summer" "1948 Summer" "1948 Summer" ...
## $ Year : int 1948 1948 1948 1948 1952 1972 2012 2012 2016 2016 ...
## $ Season: chr "Summer" "Summer" "Summer" "Summer" ...
## $ City : chr "London" "London" "London" "London" ...
## $ Sport : chr "Gymnastics" "Gymnastics" "Gymnastics" "Gymnastics" ...
## $ Event : chr "Gymnastics Men's Individual All-Around" "Gymnastics Men's Team All-Around" "Gymnast
## $ Medal : chr "Bronze" "Gold" "Gold" "Gold" ...
## - attr(*, "na.action")= 'omit' Named int [1:240935] 1 2 3 4 5 6 7 8 9 10 ...
## ..- attr(*, "names")= chr [1:240935] "1" "2" "3" "4" ...
```

```
olympics.combine2 <- olympics.combine[, c(-1:-3, -7:-12, -14, -15)]
str(olympics.combine2)
```

```
## 'data.frame': 8490 obs. of 4 variables:
## $ Age : int 28 28 28 28 32 19 19 19 24 24 ...
## $ Height: int 175 175 175 175 175 160 161 161 161 161 ...
## $ Weight: num 64 64 64 64 64 48 62 62 62 62 ...
## $ Sport : chr "Gymnastics" "Gymnastics" "Gymnastics" "Gymnastics" ...
```

```
gplot <- ggplot(olympics.combine2, aes(x=Height, y=Weight))
gplot <- gplot + geom_point(aes(size=Age, color=as.factor(Sport)))
gplot <- gplot + stat_smooth(method = "lm", aes(x=Height, y=Weight))
gplot <- gplot + labs(x = 'Height', y = 'Weight', size = 'Age',
                     color = "Sport")
gplot <- gplot + ggtitle("Scatter plot of Men's Height, Weight, Age & Sport")
gplot
```

```
## `geom_smooth()` using formula 'y ~ x'
```



randomForest analysis for the following Sport

(Equestrianism, Gymnastics, Swimming, and Weightlifting)

based on Men's Age, Height, Weight, and Team

```
olympics.combine3 <- rbind(olympics.gymnastics, olympics.swimming, olympics.weightlifting, olympics.equestrianism)
str(olympics.combine3)
```

```
## 'data.frame':   4842 obs. of  15 variables:
## $ ID      : int  17 17 17 17 17 394 455 455 455 455 ...
## $ Name    : chr  "Paavo Johannes Aaltonen" "Paavo Johannes Aaltonen" "Paavo Johannes Aaltonen" "Paavo Johannes Aaltonen" ...
## $ Sex     : chr  "M" "M" "M" "M" ...
## $ Age     : int  28 28 28 28 32 19 19 19 24 24 ...
## $ Height  : int  175 175 175 175 175 160 161 161 161 161 ...
## $ Weight  : num  64 64 64 64 64 48 62 62 62 62 ...
## $ Team    : chr  "Finland" "Finland" "Finland" "Finland" ...
## $ NOC     : chr  "FIN" "FIN" "FIN" "FIN" ...
## $ Games   : chr  "1948 Summer" "1948 Summer" "1948 Summer" "1948 Summer" ...
## $ Year    : int  1948 1948 1948 1948 1952 1972 2012 2012 2016 2016 ...
## $ Season  : chr  "Summer" "Summer" "Summer" "Summer" ...
## $ City    : chr  "London" "London" "London" "London" ...
## $ Sport   : chr  "Gymnastics" "Gymnastics" "Gymnastics" "Gymnastics" ...
## $ Event   : chr  "Gymnastics Men's Individual All-Around" "Gymnastics Men's Team All-Around" "Gymnastics Men's Individual All-Around" ...
## $ Medal   : chr  "Bronze" "Gold" "Gold" "Gold" ...
```

```
## - attr(*, "na.action")= 'omit' Named int [1:240935] 1 2 3 4 5 6 7 8 9 10 ...
## ..- attr(*, "names")= chr [1:240935] "1" "2" "3" "4" ...

olympics.combine4 <- olympics.combine3[, c(-1:-3, -7:-12, -14, -15)]
str(olympics.combine4)

## 'data.frame': 4842 obs. of 4 variables:
## $ Age : int 28 28 28 28 32 19 19 19 24 24 ...
## $ Height: int 175 175 175 175 175 160 161 161 161 161 ...
## $ Weight: num 64 64 64 64 64 48 62 62 62 62 ...
## $ Sport : chr "Gymnastics" "Gymnastics" "Gymnastics" "Gymnastics" ...

olympics.combine4$Sport <- as.factor(olympics.combine4$Sport)
str(olympics.combine4)

## 'data.frame': 4842 obs. of 4 variables:
## $ Age : int 28 28 28 28 32 19 19 19 24 24 ...
## $ Height: int 175 175 175 175 175 160 161 161 161 161 ...
## $ Weight: num 64 64 64 64 64 48 62 62 62 62 ...
## $ Sport : Factor w/ 4 levels "Equestrianism",...: 2 2 2 2 2 2 2 2 2 2 ...

olympics.combine4.rf <- randomForest(olympics.combine4[, -4], olympics.combine4[, 4])
olympics.combine4.rf

##
## Call:
## randomForest(x = olympics.combine4[, -4], y = olympics.combine4[, 4])
## Type of random forest: classification
## Number of trees: 500
## No. of variables tried at each split: 1
##
## OOB estimate of error rate: 11.46%
## Confusion matrix:
## Equestrianism Gymnastics Swimming Weightlifting class.error
## Equestrianism 524 42 92 5 0.20965309
## Gymnastics 26 1006 108 21 0.13350560
## Swimming 35 96 2341 14 0.05832663
## Weightlifting 16 65 35 416 0.21804511

importance(olympics.combine4.rf)

## MeanDecreaseGini
## Age 828.6792
## Height 944.9063
## Weight 822.0378
```

Scatter plot Men's Age, Height, Weight, and Sport (Equestrianism, Gymnastics, Swimming, & Weightlifting)

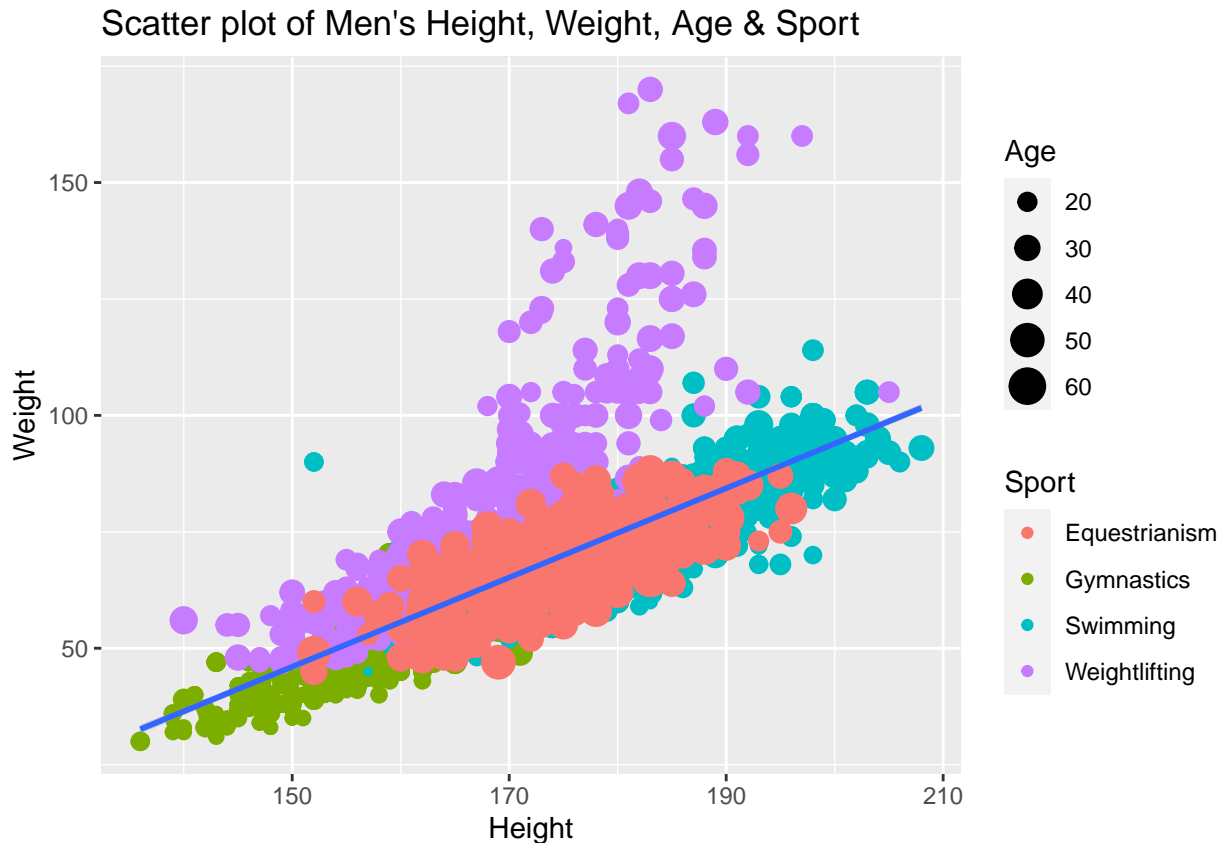
```
str(olympics.combine4)

## 'data.frame': 4842 obs. of 4 variables:
## $ Age : int 28 28 28 28 32 19 19 19 24 24 ...
## $ Height: int 175 175 175 175 175 160 161 161 161 161 ...
## $ Weight: num 64 64 64 64 64 48 62 62 62 62 ...
```

```
## $ Sport : Factor w/ 4 levels "Equestrianism",...: 2 2 2 2 2 2 2 2 2 2 ...

gplot <- ggplot(olympics.combine4, aes(x=Height, y=Weight))
gplot <- gplot + geom_point(aes(size=Age, color=as.factor(Sport)))
gplot <- gplot + stat_smooth(method = "lm", aes(x=Height, y=Weight))
gplot <- gplot + labs(x = 'Height', y = 'Weight', size = 'Age',
                     color = "Sport")
gplot <- gplot + ggtitle("Scatter plot of Men's Height, Weight, Age & Sport")
gplot
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Create train and test data sets

based on train dataset for the following Sport

(Equestrianism, Gymnastics, Swimming, and Weightlifting)

based on Men's Age, Height, Weight, and Team

```
str(olympics.combine4)
```

```
## 'data.frame': 4842 obs. of 4 variables:
## $ Age : int 28 28 28 28 32 19 19 19 24 24 ...
## $ Height: int 175 175 175 175 175 160 161 161 161 161 ...
## $ Weight: num 64 64 64 64 64 48 62 62 62 62 ...
```



```
## $ Sport : Factor w/ 4 levels "Equestrianism",...: 2 2 2 2 2 2 2 2 2 ...
split.olympics.combine4 <- sample(nrow(olympics.combine4), nrow(olympics.combine4) * .67)
olympics.combine4.Train <- olympics.combine4[split.olympics.combine4, ]
olympics.combine4.Test <- olympics.combine4[-split.olympics.combine4, ]
str(olympics.combine4.Train)

## 'data.frame': 3244 obs. of 4 variables:
## $ Age : int 21 21 17 25 27 17 25 14 29 19 ...
## $ Height: int 178 175 172 178 175 180 178 181 161 166 ...
## $ Weight: num 67 60 59 66 80 63 59 66 52 69 ...
## $ Sport : Factor w/ 4 levels "Equestrianism",...: 1 3 3 3 3 3 3 3 2 2 ...
str(olympics.combine4.Test)

## 'data.frame': 1598 obs. of 4 variables:
## $ Age : int 28 24 26 21 16 16 20 20 19 19 ...
## $ Height: int 175 161 148 154 158 158 158 158 166 166 ...
## $ Weight: num 64 62 46 53 44 44 44 44 60 60 ...
## $ Sport : Factor w/ 4 levels "Equestrianism",...: 2 2 2 2 2 2 2 2 2 2 ...
```

randomForest analysis based on train dataset for the following Sport
(Equestrianism, Gymnastics, Swimming, and Weightlifting)
based on Men's Age, Height, Weight, and Team

```
olympics.combine4.Train$Sport <- as.factor(olympics.combine4.Train$Sport)
str(olympics.combine4.Train)

## 'data.frame': 3244 obs. of 4 variables:
## $ Age : int 21 21 17 25 27 17 25 14 29 19 ...
## $ Height: int 178 175 172 178 175 180 178 181 161 166 ...
## $ Weight: num 67 60 59 66 80 63 59 66 52 69 ...
## $ Sport : Factor w/ 4 levels "Equestrianism",...: 1 3 3 3 3 3 3 3 2 2 ...
olympics.combine4.Test$Sport <- as.factor(olympics.combine4.Test$Sport)
str(olympics.combine4.Test)

## 'data.frame': 1598 obs. of 4 variables:
## $ Age : int 28 24 26 21 16 16 20 20 19 19 ...
## $ Height: int 175 161 148 154 158 158 158 158 166 166 ...
## $ Weight: num 64 62 46 53 44 44 44 44 60 60 ...
## $ Sport : Factor w/ 4 levels "Equestrianism",...: 2 2 2 2 2 2 2 2 2 2 ...
olympics.combine4.Train.rf <- randomForest(olympics.combine4.Train[,-4], olympics.combine4.Train[,4])
olympics.combine4.Train.rf

##
## Call:
## randomForest(x = olympics.combine4.Train[, -4], y = olympics.combine4.Train[, 4])
## Type of random forest: classification
## Number of trees: 500
## No. of variables tried at each split: 1
##
```

```
##          OOB estimate of  error rate: 11.99%
## Confusion matrix:
##          Equestrianism Gymnastics Swimming Weightlifting class.error
## Equestrianism          350          30          56           5 0.20634921
## Gymnastics              19         648          79          17 0.15072084
## Swimming                19          65        1565          15 0.05949519
## Weightlifting           10          49          25         292 0.22340426

importance(olympics.combine4.Train.rf)

##          MeanDecreaseGini
## Age                575.3124
## Height              653.6929
## Weight              544.9007
```

Predicting Sport based on randomForest analysis

Row 500 & 1300

```
olympics.combine4.Test[500,]

##          Age Height Weight      Sport
## 45405    21    173     63 Swimming

olympics.combine4.Test[1300,]

##          Age Height Weight      Sport
## 157473    27    175     82 Weightlifting

#Predicting Row 5000 & 1300
predict(olympics.combine4.Train.rf, olympics.combine4.Test[500,-4])

##          45405
## Swimming
## Levels: Equestrianism Gymnastics Swimming Weightlifting

predict(olympics.combine4.Train.rf, olympics.combine4.Test[1300,-4])

##          157473
## Weightlifting
## Levels: Equestrianism Gymnastics Swimming Weightlifting
```

Predicting for Age = 30, Height = 170, Weight = 50

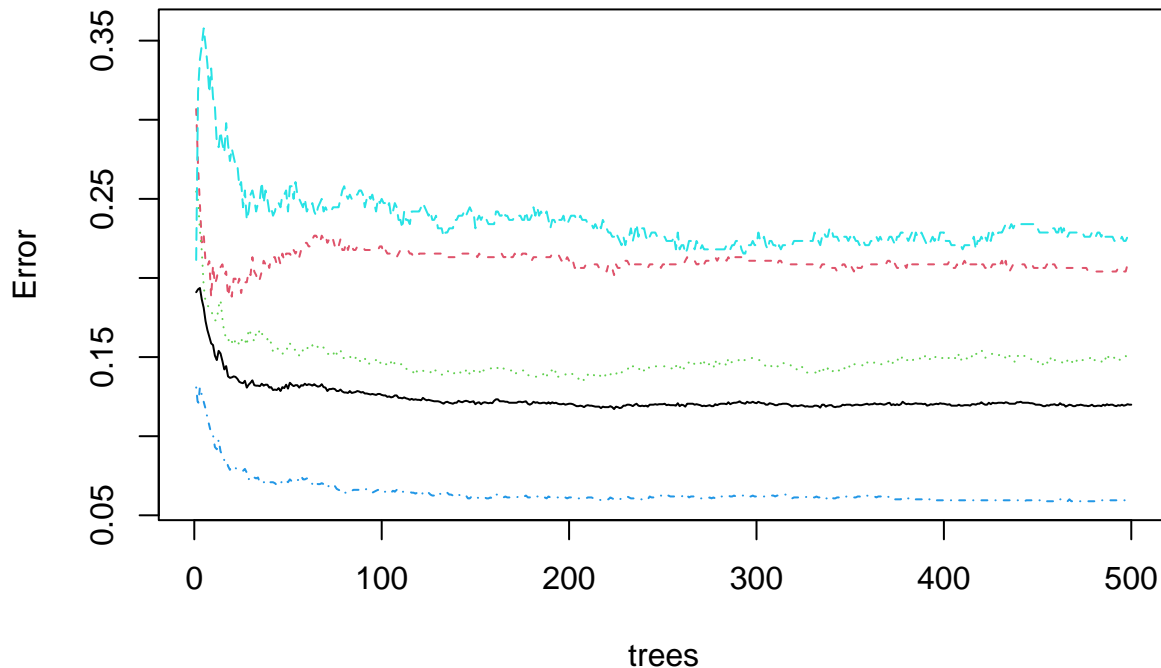
```
predict(olympics.combine4.Train.rf, c(30, 170, 50))

##          1
## Equestrianism
## Levels: Equestrianism Gymnastics Swimming Weightlifting
```

Plot on random Forest model

```
plot(olympics.combine4.Train.rf)
```

olympics.combine4.Train.rf



lm() regression models

```
str(olympics.combine4)
```

```
## 'data.frame':  4842 obs. of  4 variables:
## $ Age   : int  28 28 28 28 32 19 19 19 24 24 ...
## $ Height: int  175 175 175 175 175 160 161 161 161 161 ...
## $ Weight: num  64 64 64 64 64 48 62 62 62 62 ...
## $ Sport : Factor w/ 4 levels "Equestrianism",...: 2 2 2 2 2 2 2 2 2 2 ...
```

```
olympics.combine4$Sport <- as.factor(olympics.combine4$Sport)
str(olympics.combine4)
```

```
## 'data.frame':  4842 obs. of  4 variables:
## $ Age   : int  28 28 28 28 32 19 19 19 24 24 ...
## $ Height: int  175 175 175 175 175 160 161 161 161 161 ...
## $ Weight: num  64 64 64 64 64 48 62 62 62 62 ...
## $ Sport : Factor w/ 4 levels "Equestrianism",...: 2 2 2 2 2 2 2 2 2 2 ...
```

```
olympics.combine4$Sport1 <- c(0:0)
str(olympics.combine4)
```

```
## 'data.frame':  4842 obs. of  5 variables:
## $ Age   : int  28 28 28 28 32 19 19 19 24 24 ...
## $ Height: int  175 175 175 175 175 160 161 161 161 161 ...
## $ Weight: num  64 64 64 64 64 48 62 62 62 62 ...
## $ Sport : Factor w/ 4 levels "Equestrianism",...: 2 2 2 2 2 2 2 2 2 2 ...
## $ Sport1: int   0 0 0 0 0 0 0 0 0 0 ...
```

```
df.sport <- c('Equestrianism', 'Gymnastics', 'Swimming', 'Weightlifting')
str(df.sport)
```

```
## chr [1:4] "Equestrianism" "Gymnastics" "Swimming" "Weightlifting"
```

```
olympics.combine4$Sport1 <- match(olympics.combine4$Sport, df.sport, nomatch = 0)
```

```
knitr::kable(head(olympics.combine4[which(olympics.combine4$Sport == "Equestrianism"),]))
```

	Age	Height	Weight	Sport	Sport1
2424	29	189	80	Equestrianism	1
2430	41	189	80	Equestrianism	1
3355	40	162	70	Equestrianism	1
3361	31	175	57	Equestrianism	1
3722	27	178	69	Equestrianism	1
4659	37	163	56	Equestrianism	1

```
knitr::kable(head(olympics.combine4[which(olympics.combine4$Sport == "Gymnastics"),]))
```

	Age	Height	Weight	Sport	Sport1
42	28	175	64	Gymnastics	2
43	28	175	64	Gymnastics	2
45	28	175	64	Gymnastics	2
49	28	175	64	Gymnastics	2
51	32	175	64	Gymnastics	2
678	19	160	48	Gymnastics	2

```
knitr::kable(head(olympics.combine4[which(olympics.combine4$Sport == "Swimming"),]))
```

	Age	Height	Weight	Sport	Sport1
477	21	173	59	Swimming	3
804	22	190	78	Swimming	3
805	22	190	78	Swimming	3
815	30	197	92	Swimming	3
893	21	175	64	Swimming	3
1773	19	179	70	Swimming	3

```
knitr::kable(head(olympics.combine4[which(olympics.combine4$Sport == "Weightlifting"),]))
```

	Age	Height	Weight	Sport	Sport1
2332	21	147	47	Weightlifting	4
3046	19	178	105	Weightlifting	4
3047	23	178	105	Weightlifting	4
4002	24	192	156	Weightlifting	4
4361	20	176	89	Weightlifting	4
4405	30	185	160	Weightlifting	4

```
predict.sport.age <- lm(Sport1 ~ Age, olympics.combine4)
summary(predict.sport.age)
```

```
##
## Call:
## lm(formula = Sport1 ~ Age, data = olympics.combine4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.85905 -0.63140  0.08403  0.36860  2.33614
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.940413   0.039869   98.83  <2e-16 ***
## Age        -0.056914   0.001623  -35.08  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7653 on 4840 degrees of freedom
## Multiple R-squared:  0.2027, Adjusted R-squared:  0.2025
## F-statistic: 1230 on 1 and 4840 DF,  p-value: < 2.2e-16
predict.sport.height <- lm(Sport1 ~ Height, olympics.combine4)
summary(predict.sport.height)
```

```
##
## Call:
## lm(formula = Sport1 ~ Height, data = olympics.combine4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8751 -0.4427  0.2266  0.4047  1.8371
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.3825325  0.1699020   2.251  0.0244 *
## Height       0.0127173  0.0009736  13.063  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8424 on 4840 degrees of freedom
## Multiple R-squared:  0.03405, Adjusted R-squared:  0.03385
## F-statistic: 170.6 on 1 and 4840 DF,  p-value: < 2.2e-16
predict.sport.weight <- lm(Sport1 ~ Weight, olympics.combine4)
summary(predict.sport.weight)
```

```
##
## Call:
## lm(formula = Sport1 ~ Weight, data = olympics.combine4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.9709 -0.3765  0.1281  0.4650  1.8415
##
```

```
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2272015  0.0516132   23.78  <2e-16 ***
## Weight      0.0198152  0.0007284   27.20  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7982 on 4840 degrees of freedom
## Multiple R-squared:  0.1326, Adjusted R-squared:  0.1325
## F-statistic: 740.1 on 1 and 4840 DF,  p-value: < 2.2e-16

predict.sport.height.weight <- lm(Sport1 ~ Height + Weight, olympics.combine4)
summary(predict.sport.height.weight)
```

```
##
## Call:
## lm(formula = Sport1 ~ Height + Weight, data = olympics.combine4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.0943 -0.4394  0.2076  0.5188  1.7268
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.173886   0.192160   16.52  <2e-16 ***
## Height      -0.014658   0.001395  -10.51  <2e-16 ***
## Weight       0.028570   0.001101   25.94  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7894 on 4839 degrees of freedom
## Multiple R-squared:  0.152, Adjusted R-squared:  0.1516
## F-statistic: 433.6 on 2 and 4839 DF,  p-value: < 2.2e-16
```

```
predict.sport <- lm(Sport1 ~ Age + Height + Weight, olympics.combine4)
summary(predict.sport)
```

```
##
## Call:
## lm(formula = Sport1 ~ Age + Height + Weight, data = olympics.combine4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.12804 -0.42757  0.02107  0.33133  2.54187
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.6514538  0.1619392   28.72  <2e-16 ***
## Age         -0.0663255  0.0014018  -47.32  <2e-16 ***
## Height      -0.0163120  0.0011540  -14.13  <2e-16 ***
## Weight       0.0340233  0.0009179   37.06  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6527 on 4838 degrees of freedom
```

```
## Multiple R-squared:  0.4203, Adjusted R-squared:  0.4199  
## F-statistic: 1169 on 3 and 4838 DF,  p-value: < 2.2e-16
```