DATA SCIENCE METHODS

Assignment 02

25/03/2022

Group 01:

Bilge Kasapoğlu – u941664

Hoan Van Nguyen – u1274449

Jiahe Wang – u489199

Roshini Sudhaharan – u725261

Clean environment

```
rm(list=ls())
```

Load the required libraries

```
library(foreign)
library(data.table)
library(dplyr)
library(fastDummies)
library(stringr)
library(readxl)
library(leaps)
library(glmnet)
library(pls)
library(fastDummies)
library(randomForest)
```

QUESTION 01:

Not to have scientific notation

```
options(scipen = 999)
```

Load and inspect the data set

```
soccer <- read_excel("soccer.xlsx")
summary(soccer)

##       Item            ID               Name                 Age
##  Min.   :    0   Min.   :    16   Length:18207       Min.   :16.00
##  1st Qu.: 4552   1st Qu.:200316   Class :character   1st Qu.:21.00
##  Median : 9103   Median :221759   Mode  :character   Median :25.00
##  Mean   : 9103   Mean   :214298                      Mean   :25.12
##  3rd Qu.:13654   3rd Qu.:236530                      3rd Qu.:28.00
##  Max.   :18206   Max.   :246620                      Max.   :45.00
##
##  Nationality           Overall        Potential        Value
##  Length:18207       Min.   :46.00   Min.   :48.00   Length:18207
```

```
##  Class :character   1st Qu.:62.00   1st Qu.:67.00   Class :character  
##  Mode  :character   Median :66.00   Median :71.00   Mode  :character  
##                     Mean   :66.24   Mean   :71.31                     
##                     3rd Qu.:71.00   3rd Qu.:75.00                     
##                     Max.   :94.00   Max.   :95.00                     
##                                                                       
##      Wage           Preferred Foot     International Reputation   Weak Foot    
##  Length:18207       Length:18207       Min.   :1.000            Min.   :1.000  
##  Class :character   Class :character   1st Qu.:1.000            1st Qu.:3.000  
##  Mode  :character   Mode  :character   Median :1.000            Median :3.000  
##                                        Mean   :1.113            Mean   :2.947  
##                                        3rd Qu.:1.000            3rd Qu.:3.000  
##                                        Max.   :5.000            Max.   :5.000  
##                                        NA's   :48               NA's   :48     
##   Skill Moves     Work Rate          Body Type           Position        
##  Min.   :1.000   Length:18207       Length:18207       Length:18207      
##  1st Qu.:2.000   Class :character   Class :character   Class :character  
##  Median :2.000   Mode  :character   Mode  :character   Mode  :character  
##  Mean   :2.361                                                           
##  3rd Qu.:3.000                                                           
##  Max.   :5.000                                                           
##  NA's   :48                                                              
##     Height             Weight             Crossing       Finishing    
##  Length:18207       Length:18207       Min.   : 5.00   Min.   : 2.00  
##  Class :character   Class :character   1st Qu.:38.00   1st Qu.:30.00  
##  Mode  :character   Mode  :character   Median :54.00   Median :49.00  
##                                        Mean   :49.73   Mean   :45.55  
##                                        3rd Qu.:64.00   3rd Qu.:62.00  
##                                        Max.   :93.00   Max.   :95.00  
##                                        NA's   :48      NA's   :48     
##  HeadingAccuracy  ShortPassing      Volleys        Dribbling    
##  Min.   : 4.0    Min.   : 7.00   Min.   : 4.00   Min.   : 4.00  
##  1st Qu.:44.0    1st Qu.:54.00   1st Qu.:30.00   1st Qu.:49.00  
##  Median :56.0    Median :62.00   Median :44.00   Median :61.00  
##  Mean   :52.3    Mean   :58.69   Mean   :42.91   Mean   :55.37  
##  3rd Qu.:64.0    3rd Qu.:68.00   3rd Qu.:57.00   3rd Qu.:68.00  
##  Max.   :94.0    Max.   :93.00   Max.   :90.00   Max.   :97.00  
##  NA's   :48      NA's   :48      NA's   :48      NA's   :48     
##      Curve         FKAccuracy     LongPassing     BallControl   
##  Min.   : 6.00   Min.   : 3.00   Min.   : 9.00   Min.   : 5.00  
##  1st Qu.:34.00   1st Qu.:31.00   1st Qu.:43.00   1st Qu.:54.00  
##  Median :48.00   Median :41.00   Median :56.00   Median :63.00  
##  Mean   :47.17   Mean   :42.86   Mean   :52.71   Mean   :58.37  
##  3rd Qu.:62.00   3rd Qu.:57.00   3rd Qu.:64.00   3rd Qu.:69.00  
##  Max.   :94.00   Max.   :94.00   Max.   :93.00   Max.   :96.00  
##  NA's   :48      NA's   :48      NA's   :48      NA's   :48     
##   Acceleration    SprintSpeed       Agility       Reactions        Balance     
##  Min.   :12.00   Min.   :12.00   Min.   :14.0   Min.   :21.00   Min.   :16.00  
##  1st Qu.:57.00   1st Qu.:57.00   1st Qu.:55.0   1st Qu.:56.00   1st Qu.:56.00  
##  Median :67.00   Median :67.00   Median :66.0   Median :62.00   Median :66.00  
##  Mean   :64.61   Mean   :64.73   Mean   :63.5   Mean   :61.84   Mean   :63.97  
##  3rd Qu.:75.00   3rd Qu.:75.00   3rd Qu.:74.0   3rd Qu.:68.00   3rd Qu.:74.00  
```

```
##  Max.   :97.00   Max.   :96.00   Max.   :96.0    Max.   :96.00   Max.   :96.00  
##  NA's   :48      NA's   :48      NA's   :48      NA's   :48      NA's   :48     
##    ShotPower        Jumping          Stamina          Strength    
##  Min.   : 2.00   Min.   :15.00   Min.   :12.00   Min.   :17.00  
##  1st Qu.:45.00   1st Qu.:58.00   1st Qu.:56.00   1st Qu.:58.00  
##  Median :59.00   Median :66.00   Median :66.00   Median :67.00  
##  Mean   :55.46   Mean   :65.09   Mean   :63.22   Mean   :65.31  
##  3rd Qu.:68.00   3rd Qu.:73.00   3rd Qu.:74.00   3rd Qu.:74.00  
##  Max.   :95.00   Max.   :95.00   Max.   :96.00   Max.   :97.00  
##  NA's   :48      NA's   :48      NA's   :48      NA's   :48     
##    LongShots       Aggression     Interceptions   Positioning        Vision    
##  Min.   : 3.00   Min.   :11.00   Min.   : 3.0    Min.   : 2.00   Min.   :10.0  
##  1st Qu.:33.00   1st Qu.:44.00   1st Qu.:26.0    1st Qu.:38.00   1st Qu.:44.0  
##  Median :51.00   Median :59.00   Median :52.0    Median :55.00   Median :55.0  
##  Mean   :47.11   Mean   :55.87   Mean   :46.7    Mean   :49.96   Mean   :53.4  
##  3rd Qu.:62.00   3rd Qu.:69.00   3rd Qu.:64.0    3rd Qu.:64.00   3rd Qu.:64.0  
##  Max.   :94.00   Max.   :95.00   Max.   :92.0    Max.   :95.00   Max.   :94.0  
##  NA's   :48      NA's   :48      NA's   :48      NA's   :48      NA's   :48    
##    Penalties       Composure        Marking       StandingTackle SlidingTackle  
##  Min.   : 5.00   Min.   : 3.00   Min.   : 3.00   Min.   : 2.0   Min.   : 3.00  
##  1st Qu.:39.00   1st Qu.:51.00   1st Qu.:30.00   1st Qu.:27.0   1st Qu.:24.00  
##  Median :49.00   Median :60.00   Median :53.00   Median :55.0   Median :52.00  
##  Mean   :48.55   Mean   :58.65   Mean   :47.28   Mean   :47.7   Mean   :45.66  
##  3rd Qu.:60.00   3rd Qu.:67.00   3rd Qu.:64.00   3rd Qu.:66.0   3rd Qu.:64.00  
##  Max.   :92.00   Max.   :96.00   Max.   :94.00   Max.   :93.0   Max.   :91.00  
##  NA's   :48      NA's   :48      NA's   :48      NA's   :48     NA's   :48     
##     GKDiving       GKHandling       GKKicking      GKPositioning  
##  Min.   : 1.00   Min.   : 1.00   Min.   : 1.00   Min.   : 1.00  
##  1st Qu.: 8.00   1st Qu.: 8.00   1st Qu.: 8.00   1st Qu.: 8.00  
##  Median :11.00   Median :11.00   Median :11.00   Median :11.00  
##  Mean   :16.62   Mean   :16.39   Mean   :16.23   Mean   :16.39  
##  3rd Qu.:14.00   3rd Qu.:14.00   3rd Qu.:14.00   3rd Qu.:14.00  
##  Max.   :90.00   Max.   :92.00   Max.   :91.00   Max.   :90.00  
##  NA's   :48      NA's   :48      NA's   :48      NA's   :48     
##    GKReflexes    Release Clause    
##  Min.   : 1.00   Length:18207      
##  1st Qu.: 8.00   Class :character  
##  Median :11.00   Mode  :character  
##  Mean   :16.71                     
##  3rd Qu.:14.00                     
##  Max.   :94.00                     
##  NA's   :48
```

Drop the missing and duplicated values

```
dta <-  na.omit(soccer)
dta <- dta[!duplicated(dta$Name),]
dta <- dta  %>% select(c(4:20))
```

Transform variables if necessary

```
# Work Rate
unique(dta$`Work Rate`)
dta$`Work Rate` <- ordered(dta$`Work Rate`, levels =
```

```
                              c("Low/ Low", "Low/ Medium", "Low/ High",
                                "Medium/ Low", "Medium/ Medium", "Medium/ High",
                                "High/ Low", "High/ Medium","High/ High"))
dta$`Work Rate` <- unclass(dta$`Work Rate`)

# Body Types
unique(dta$`Body Type`)
table(dta$`Body Type`)
dta$`Body Type`[dta$`Body Type` == "PLAYER_BODY_TYPE_25"] <- "Other"
dta <- dummy_cols(dta, select_columns = "Body Type")
dta <- dta  %>% select(-c("Body Type"))

# Position
unique(dta$Position)
table(dta$Position)
dta <- dummy_cols(dta, select_columns = "Position")
dta <- dta  %>% select(-c("Position"))

# Preferred Foot
unique(dta$`Preferred Foot`)
table(dta$`Preferred Foot`)
dta <- dummy_cols(dta, select_columns = "Preferred Foot")
dta <- dta  %>% select(-c("Preferred Foot"))

# Convert Wage into Numeric
sum(!str_detect(dta$Wage, "€"))
sum(!str_detect(dta$Wage, "K"))
dta$Wage <- str_remove(dta$Wage, "€")
dta$Wage <- str_remove(dta$Wage, "K") %>% as.numeric()
dta$Wage <- dta$Wage * 1000

# Convert Value into Numeric
money_convert <- function(x)  {
  value = as.numeric(str_sub(x, 2, -2))
  suffix = str_sub(x, nchar(x), -1)
  if (suffix == "K"){
    value = value*1000
  } else if ((suffix == "M"))  {
    value = value*1000000
  }
}
sum(!str_detect(dta$Value, "€"))
sum(!str_detect(dta$Value, "K"))
sum(!str_detect(dta$Value, "M"))
dta$Value <- sapply(dta$Value, money_convert)

# Nationality
dta <- dummy_cols(dta, select_columns = "Nationality")
dta <- dta  %>% select(-c("Nationality"))

# Weight
dta$Weight <- str_remove(dta$Weight, "lbs") %>% as.numeric()
```

```
# Height
dta$Height <- str_replace(dta$Height, "'", ".") %>% as.numeric()
head(dta)

data <- lapply(dta, as.numeric) %>% as.data.frame()
```

1a.

Linear Regression

```
# Split data into two parts
set.seed(21) # for reproducibility
sum(is.na(data)) # check whether there are missing values

## [1] 0

train.size = dim(data)[1] / 2 # Learn how many observations we need to sample
train = sample(1:dim(data)[1], train.size) # get indexes of observations for
training set
test = -train # get indexes of observations for test set
data.train = data[train, ] # extract training observations
data.test = data[test, ] # extract test observations

# Linear Regression
lm.fit = lm(Wage~., data=data.train) # apply linear regression
lm.pred = predict(lm.fit, data.test)  %>% as.numeric() # get predicted outcomes
testerror.ls = mean((data.test[, "Wage"] - lm.pred)^2) # find MSE
testerror.ls

## [1] 118531610
```

Lasso

```
# Convert data to matrix format
train.mat = model.matrix(Wage~., data=data.train)
test.mat = model.matrix(Wage~., data=data.test)

# Create a grid of lambdas from a large range
grid = 10^seq(2, -3, length=100)

# Lasso
mod.lasso = cv.glmnet(train.mat, data.train[, "Wage"], alpha=1, lambda=grid,
thresh=1e-12, nfolds = 5)
lambda.best.lasso = mod.lasso$lambda.min
lambda.best.lasso

## [1] 100

lasso.pred = predict(mod.lasso, newx=test.mat, s=lambda.best.lasso)
testerror.lasso=mean((data.test[, "Wage"] - lasso.pred)^2)
testerror.lasso

## [1] 117873828
```

```r
# estimate lasso regression with full data and present coefficients
mod.lasso = glmnet(model.matrix(Wage~., data=data), data[, "Wage"], alpha=1)
coef.lasso=predict(mod.lasso, s=lambda.best.lasso, type="coefficients")
coef.lasso
```

```
## 213 x 1 sparse Matrix of class "dgCMatrix"
##                                              s1
## (Intercept)                    -19529.221253574
## (Intercept)                          .
## Age                               184.514478608
## Overall                            57.402264242
## Potential                           3.936971196
## Value                               0.002810146
## International.Reputation         9819.432044598
## Weak.Foot                            .
## Skill.Moves                      -373.758869558
## Work.Rate                          16.320494118
## Height                            100.509695415
## Weight                              9.093830527
## Crossing                           20.846424846
## Finishing                            .
## Body.Type_Akinfenwa                  .
## Body.Type_C..Ronaldo           136711.375069051
## Body.Type_Courtois              43948.549484346
## Body.Type_Lean                     24.170958905
## Body.Type_Messi                200632.048838682
## Body.Type_Neymar               -73487.004001360
## Body.Type_Normal                     .
## Body.Type_Other                 25247.210082757
## Body.Type_Shaqiri               43257.607186249
## Body.Type_Stocky                 -459.668585589
## Position_CAM                     -759.437695977
## Position_CB                       636.574460170
## Position_CDM                      357.976658925
## Position_CF                      -474.419856295
## Position_CM                          .
## Position_GK                        -7.049049916
## Position_LAM                         .
## Position_LB                       881.822147549
## Position_LCB                     -397.253669401
## Position_LCM                    -1857.434185634
## Position_LDM                     -398.717921965
## Position_LF                     -8798.191668584
## Position_LM                      -682.388138278
## Position_LS                     -2817.097231881
## Position_LW                       396.833401371
## Position_LWB                         .
## Position_RAM                         .
## Position_RB                       605.438113714
## Position_RCB                         .
## Position_RCM                     -830.522807105
## Position_RDM                         .
```

```
## Position_RF                          -2124.104959230
## Position_RM                           -523.446917780
## Position_RS                          -1040.099499173
## Position_RW                            508.059795713
## Position_RWB                             .
## Position_ST                             92.965176392
## Preferred.Foot_Left                      .
## Preferred.Foot_Right                     .
## Nationality_Afghanistan                  .
## Nationality_Albania                      .
## Nationality_Algeria                      .
## Nationality_Andorra                      .
## Nationality_Angola                       .
## Nationality_Antigua...Barbuda            .
## Nationality_Argentina                   -4.853373043
## Nationality_Armenia                   1954.764732489
## Nationality_Australia                    .
## Nationality_Austria                    524.188063263
## Nationality_Azerbaijan                   .
## Nationality_Barbados                     .
## Nationality_Belarus                      .
## Nationality_Belgium                    551.566871237
## Nationality_Belize                       .
## Nationality_Benin                        .
## Nationality_Bermuda                      .
## Nationality_Bolivia                      .
## Nationality_Bosnia.Herzegovina           .
## Nationality_Botswana                     .
## Nationality_Brazil                     766.171614043
## Nationality_Bulgaria                 -1037.098388635
## Nationality_Burkina.Faso                 .
## Nationality_Burundi                      .
## Nationality_Cameroon                     .
## Nationality_Canada                       .
## Nationality_Cape.Verde               -1118.551797137
## Nationality_Central.African.Rep.     -7134.146835703
## Nationality_Chad                         .
## Nationality_Chile                        .
## Nationality_China.PR                   235.266717132
## Nationality_Colombia                  -921.635328289
## Nationality_Comoros                      .
## Nationality_Congo                        .
## Nationality_Costa.Rica                 189.548374268
## Nationality_Croatia                    679.904781659
## Nationality_Cuba                         .
## Nationality_Curacao                      .
## Nationality_Cyprus                       .
## Nationality_Czech.Republic           -2197.956835381
## Nationality_Denmark                      .
## Nationality_Dominican.Republic       31751.867279432
## Nationality_DR.Congo                     .
## Nationality_Ecuador                   3290.314436635
```

```
## Nationality_Egypt                           .
## Nationality_El.Salvador                     .
## Nationality_England                    3833.217440922
## Nationality_Equatorial.Guinea           540.793834508
## Nationality_Eritrea                         .
## Nationality_Estonia                         .
## Nationality_Ethiopia                        .
## Nationality_Faroe.Islands                   .
## Nationality_Fiji                            .
## Nationality_Finland                         .
## Nationality_France                          .
## Nationality_FYR.Macedonia                   .
## Nationality_Gabon                      1745.865955640
## Nationality_Gambia                          .
## Nationality_Georgia                    -884.544520423
## Nationality_Germany                         .
## Nationality_Ghana                           .
## Nationality_Greece                    -3131.479792121
## Nationality_Grenada                         .
## Nationality_Guam                            .
## Nationality_Guatemala                       .
## Nationality_Guinea                          .
## Nationality_Guinea.Bissau                   .
## Nationality_Guyana                          .
## Nationality_Haiti                           .
## Nationality_Honduras                        .
## Nationality_Hong.Kong                       .
## Nationality_Hungary                         .
## Nationality_Iceland                         .
## Nationality_Indonesia                       .
## Nationality_Iran                            .
## Nationality_Iraq                            .
## Nationality_Israel                          .
## Nationality_Italy                           .
## Nationality_Ivory.Coast                     .
## Nationality_Jamaica                         .
## Nationality_Japan                           .
## Nationality_Jordan                          .
## Nationality_Kazakhstan                      .
## Nationality_Kenya                           .
## Nationality_Korea.DPR                       .
## Nationality_Korea.Republic             -871.027516634
## Nationality_Kosovo                          .
## Nationality_Kuwait                          .
## Nationality_Latvia                          .
## Nationality_Lebanon                         .
## Nationality_Libya                           .
## Nationality_Liechtenstein                   .
## Nationality_Lithuania                       .
## Nationality_Luxembourg                      .
## Nationality_Madagascar                      .
## Nationality_Mali                            .
```

```
## Nationality_Mauritania                  .
## Nationality_Mauritius                   .
## Nationality_Mexico                    944.400342818
## Nationality_Moldova                     .
## Nationality_Montenegro                  .
## Nationality_Montserrat                  .
## Nationality_Morocco                     .
## Nationality_Mozambique                  .
## Nationality_Namibia                     .
## Nationality_Netherlands              -406.977552111
## Nationality_New.Caledonia               .
## Nationality_New.Zealand                 .
## Nationality_Nicaragua                   .
## Nationality_Niger                       .
## Nationality_Nigeria                     .
## Nationality_Northern.Ireland         1542.644349819
## Nationality_Norway                    -25.480712754
## Nationality_Oman                        .
## Nationality_Palestine                   .
## Nationality_Panama                      .
## Nationality_Paraguay                    .
## Nationality_Peru                     -787.525050394
## Nationality_Philippines                 .
## Nationality_Poland                      .
## Nationality_Portugal                -4451.264602794
## Nationality_Puerto.Rico                 .
## Nationality_Qatar                       .
## Nationality_Republic.of.Ireland       935.613742321
## Nationality_Romania                     .
## Nationality_Russia                  -8312.310309193
## Nationality_Rwanda                      .
## Nationality_São.Tomé...Príncipe         .
## Nationality_Saudi.Arabia             2249.233566503
## Nationality_Scotland                 2214.321129065
## Nationality_Senegal                     .
## Nationality_Serbia                   -650.040071472
## Nationality_Sierra.Leone                .
## Nationality_Slovakia                -1250.412913662
## Nationality_Slovenia                -4631.582762604
## Nationality_South.Africa             -931.431610064
## Nationality_South.Sudan                 .
## Nationality_Spain                       .
## Nationality_St.Kitts.Nevis              .
## Nationality_Sudan                       .
## Nationality_Suriname                    .
## Nationality_Sweden                   -171.645174726
## Nationality_Switzerland                 .
## Nationality_Syria                       .
## Nationality_Tanzania                    .
## Nationality_Thailand                    .
## Nationality_Togo                        .
## Nationality_Trinidad...Tobago           .
```

```
## Nationality_Tunisia                          .
## Nationality_Turkey                  2000.812420090
## Nationality_Uganda                           .
## Nationality_Ukraine                -6890.255602278
## Nationality_United.Arab.Emirates             .
## Nationality_United.States           -888.980175804
## Nationality_Uruguay                          .
## Nationality_Uzbekistan                       .
## Nationality_Venezuela               -814.177982834
## Nationality_Wales                   3965.252059828
## Nationality_Zambia                           .
## Nationality_Zimbabwe                         .
```

```
# the number of non-zero coefficient (of predictors)
length(which(coef.lasso[-(1:2),]!=0))
```

```
## [1] 79
```

```
which(coef.lasso[-(1:2),]!=0)
```

```
##                                Age                          Overall
##                                  1                                2
##                          Potential                            Value
##                                  3                                4
##           International.Reputation                      Skill.Moves
##                                  5                                7
##                          Work.Rate                           Height
##                                  8                                9
##                             Weight                         Crossing
##                                 10                               11
##               Body.Type_C..Ronaldo              Body.Type_Courtois
##                                 14                               15
##                     Body.Type_Lean                  Body.Type_Messi
##                                 16                               17
##                   Body.Type_Neymar                  Body.Type_Other
##                                 18                               20
##                  Body.Type_Shaqiri                 Body.Type_Stocky
##                                 21                               22
##                       Position_CAM                      Position_CB
##                                 23                               24
##                       Position_CDM                      Position_CF
##                                 25                               26
##                        Position_GK                      Position_LB
##                                 28                               30
##                       Position_LCB                     Position_LCM
##                                 31                               32
##                       Position_LDM                      Position_LF
##                                 33                               34
##                        Position_LM                      Position_LS
##                                 35                               36
##                        Position_LW                      Position_RB
##                                 37                               40
##                       Position_RCM                      Position_RF
```

```
##                              42                               44
##                     Position_RM                      Position_RS
##                              45                               46
##                     Position_RW                      Position_ST
##                              47                               49
##           Nationality_Argentina              Nationality_Armenia
##                              58                               59
##             Nationality_Austria              Nationality_Belgium
##                              61                               65
##              Nationality_Brazil             Nationality_Bulgaria
##                              72                               73
##          Nationality_Cape.Verde Nationality_Central.African.Rep.
##                              78                               79
##            Nationality_China.PR             Nationality_Colombia
##                              82                               83
##          Nationality_Costa.Rica              Nationality_Croatia
##                              86                               87
##      Nationality_Czech.Republic   Nationality_Dominican.Republic
##                              91                               93
##             Nationality_Ecuador              Nationality_England
##                              95                               98
##   Nationality_Equatorial.Guinea                Nationality_Gabon
##                              99                              108
##             Nationality_Georgia               Nationality_Greece
##                             110                              113
##      Nationality_Korea.Republic               Nationality_Mexico
##                             137                              150
##         Nationality_Netherlands     Nationality_Northern.Ireland
##                             157                              163
##              Nationality_Norway                 Nationality_Peru
##                             164                              169
##            Nationality_Portugal  Nationality_Republic.of.Ireland
##                             172                              175
##              Nationality_Russia         Nationality_Saudi.Arabia
##                             177                              180
##            Nationality_Scotland               Nationality_Serbia
##                             181                              183
##            Nationality_Slovakia             Nationality_Slovenia
##                             185                              186
##        Nationality_South.Africa               Nationality_Sweden
##                             187                              193
##              Nationality_Turkey              Nationality_Ukraine
##                             201                              203
##       Nationality_United.States            Nationality_Venezuela
##                             205                              208
##               Nationality_Wales
##                             209
```

```
coef_lasso <- coef.lasso[-(1:2),] %>% as.data.frame()
names(coef_lasso) <- "Coef.Est."
coef_lasso <- coef_lasso %>% arrange(abs(coef_lasso$Coef.Est.))
```

Ridge

```r
# Ridge
mod.ridge = cv.glmnet(train.mat, data.train[, "Wage"], alpha=0, lambda=grid,
thresh=1e-12, nfolds = 5)
lambda.best.ridge = mod.ridge$lambda.min
lambda.best.ridge
```

```
## [1] 79.24829
```

```r
ridge.pred = predict(mod.ridge, newx=test.mat, s=lambda.best.ridge)
testerror.ridge = mean((data.test[, "Wage"] - ridge.pred)^2)
testerror.ridge
```

```
## [1] 118420695
```

```r
# estimate ridge regression with full data and present coefficients
mod.ridge = glmnet(model.matrix(Wage~., data=data), data[, "Wage"], alpha=1)
coef.ridge=predict(mod.ridge, s=lambda.best.ridge, type="coefficients")
coef.ridge
```

```
## 213 x 1 sparse Matrix of class "dgCMatrix"
##                                                   s1
## (Intercept)                      -21155.2291835064970655
## (Intercept)                           .
## Age                                 199.9653290308133364
## Overall                              51.3595832598032729
## Potential                            18.9460054265721354
## Value                                 0.0028116445117049
## International.Reputation           9791.2657512245987164
## Weak.Foot                             .
## Skill.Moves                        -509.3747380644920781
## Work.Rate                            30.9667784291440675
## Height                              151.7361546265980223
## Weight                               11.4707167430354584
## Crossing                             24.0739856165341379
## Finishing                             .
## Body.Type_Akinfenwa                   .
## Body.Type_C..Ronaldo             139287.1939886672771536
## Body.Type_Courtois                46246.8231726534650079
## Body.Type_Lean                       73.0586547217811955
## Body.Type_Messi                  204205.6455278349167202
## Body.Type_Neymar                 -76415.5472667725989595
## Body.Type_Normal                      .
## Body.Type_Other                   27853.7464074500458082
## Body.Type_Shaqiri                 46188.5273870402670582
## Body.Type_Stocky                   -554.4959708901798194
## Position_CAM                       -801.9967335756008424
## Position_CB                         652.4815201591922005
## Position_CDM                        413.2547788757198646
## Position_CF                        -724.4447933512340114
## Position_CM                          61.1174956758651646
## Position_GK                        -192.5527991420316880
## Position_LAM                          .
```

```
## Position_LB                          938.3588936769497195
## Position_LCB                        -587.0891176696935645
## Position_LCM                       -2018.2975113316492752
## Position_LDM                        -591.1499650035382274
## Position_LF                        -9422.8278330104185443
## Position_LM                         -728.2449920498286247
## Position_LS                        -3031.5392498330647868
## Position_LW                          633.7011527776261346
## Position_LWB                           .
## Position_RAM                           .
## Position_RB                          630.5788759742354159
## Position_RCB                           .
## Position_RCM                        -991.3837415691752994
## Position_RDM                           .
## Position_RF                        -3138.5475913640334511
## Position_RM                         -576.6380572854802722
## Position_RS                        -1248.5459137256286795
## Position_RW                          723.9669908511991707
## Position_RWB                           .
## Position_ST                          163.2551533179317573
## Preferred.Foot_Left                  -43.2716340465480584
## Preferred.Foot_Right                   0.0000000002150304
## Nationality_Afghanistan                .
## Nationality_Albania                    .
## Nationality_Algeria                    .
## Nationality_Andorra                    .
## Nationality_Angola                     .
## Nationality_Antigua...Barbuda          .
## Nationality_Argentina               -104.6827042824260730
## Nationality_Armenia                 2835.5386577184294765
## Nationality_Australia                  .
## Nationality_Austria                  695.4131107244789973
## Nationality_Azerbaijan                 .
## Nationality_Barbados                   .
## Nationality_Belarus                    .
## Nationality_Belgium                  713.3336085736502810
## Nationality_Belize                     .
## Nationality_Benin                     15.5060601241888474
## Nationality_Bermuda                    .
## Nationality_Bolivia                    .
## Nationality_Bosnia.Herzegovina         .
## Nationality_Botswana                   .
## Nationality_Brazil                   848.8358889060687034
## Nationality_Bulgaria               -1768.3553338266272021
## Nationality_Burkina.Faso               .
## Nationality_Burundi                    .
## Nationality_Cameroon                   .
## Nationality_Canada                     .
## Nationality_Cape.Verde             -1754.8522053831013636
## Nationality_Central.African.Rep.   -8660.7248105068501900
## Nationality_Chad                       .
## Nationality_Chile                    -52.0711165889367322
```

```
## Nationality_China.PR                    400.1442957552147845
## Nationality_Colombia                  -1016.7383176647890650
## Nationality_Comoros                       .
## Nationality_Congo                         .
## Nationality_Costa.Rica                  709.9658393454252518
## Nationality_Croatia                     879.5100340692451937
## Nationality_Cuba                          .
## Nationality_Curacao                       .
## Nationality_Cyprus                        .
## Nationality_Czech.Republic            -2534.2905730154711819
## Nationality_Denmark                       .
## Nationality_Dominican.Republic        33499.4597449186549056
## Nationality_DR.Congo                     11.4456028963948668
## Nationality_Ecuador                    3813.4340850079265692
## Nationality_Egypt                         .
## Nationality_El.Salvador                   .
## Nationality_England                    3908.4720066626714470
## Nationality_Equatorial.Guinea          1713.4894873529246979
## Nationality_Eritrea                       .
## Nationality_Estonia                       .
## Nationality_Ethiopia                      .
## Nationality_Faroe.Islands                 .
## Nationality_Fiji                          .
## Nationality_Finland                       .
## Nationality_France                        .
## Nationality_FYR.Macedonia                 .
## Nationality_Gabon                      2512.7851754799744413
## Nationality_Gambia                        .
## Nationality_Georgia                   -1359.3071483257608634
## Nationality_Germany                       .
## Nationality_Ghana                         .
## Nationality_Greece                    -3444.8205715860890450
## Nationality_Grenada                       .
## Nationality_Guam                          .
## Nationality_Guatemala                     .
## Nationality_Guinea                        .
## Nationality_Guinea.Bissau                 .
## Nationality_Guyana                        .
## Nationality_Haiti                         .
## Nationality_Honduras                   -303.6912997164046715
## Nationality_Hong.Kong                     .
## Nationality_Hungary                    -496.1294968885154049
## Nationality_Iceland                      26.6639850632590623
## Nationality_Indonesia                     .
## Nationality_Iran                          .
## Nationality_Iraq                          .
## Nationality_Israel                        .
## Nationality_Italy                         .
## Nationality_Ivory.Coast                   .
## Nationality_Jamaica                       .
## Nationality_Japan                       -29.4997787859302960
## Nationality_Jordan                        .
```

```
## Nationality_Kazakhstan                .
## Nationality_Kenya                     .
## Nationality_Korea.DPR                 .
## Nationality_Korea.Republic        -1015.8015353838673036
## Nationality_Kosovo                    .
## Nationality_Kuwait                    .
## Nationality_Latvia                    .
## Nationality_Lebanon                   .
## Nationality_Libya                     .
## Nationality_Liechtenstein             .
## Nationality_Lithuania                 .
## Nationality_Luxembourg                .
## Nationality_Madagascar                .
## Nationality_Mali                      .
## Nationality_Mauritania                .
## Nationality_Mauritius                 .
## Nationality_Mexico                 1101.5498550642232658
## Nationality_Moldova                   .
## Nationality_Montenegro                .
## Nationality_Montserrat                .
## Nationality_Morocco                   .
## Nationality_Mozambique                .
## Nationality_Namibia                   .
## Nationality_Netherlands            -533.6034747162877920
## Nationality_New.Caledonia             .
## Nationality_New.Zealand               .
## Nationality_Nicaragua                 .
## Nationality_Niger                     .
## Nationality_Nigeria                   .
## Nationality_Northern.Ireland       1883.0015745612774936
## Nationality_Norway                 -177.7034338836225800
## Nationality_Oman                      .
## Nationality_Palestine                 .
## Nationality_Panama                    .
## Nationality_Paraguay                  .
## Nationality_Peru                  -1272.5334076209082923
## Nationality_Philippines            1258.2173982549272750
## Nationality_Poland                  -86.1080363032450151
## Nationality_Portugal              -4638.8155428014943027
## Nationality_Puerto.Rico               .
## Nationality_Qatar                     .
## Nationality_Republic.of.Ireland    1112.4137976484134924
## Nationality_Romania                   .
## Nationality_Russia                -8664.4519571483415348
## Nationality_Rwanda                    .
## Nationality_São.Tomé...Príncipe       .
## Nationality_Saudi.Arabia           2463.3129163874455116
## Nationality_Scotland               2401.2434199851413723
## Nationality_Senegal                 122.3181147333869490
## Nationality_Serbia                 -947.9531552762744013
## Nationality_Sierra.Leone              .
## Nationality_Slovakia              -1650.1593556799791713
```

```
## Nationality_Slovenia                 -5051.3065033215716539
## Nationality_South.Africa             -1249.6666357632636846
## Nationality_South.Sudan                  .
## Nationality_Spain                        .
## Nationality_St.Kitts.Nevis             776.7396036760649167
## Nationality_Sudan                        .
## Nationality_Suriname                     .
## Nationality_Sweden                    -305.5958207569042884
## Nationality_Switzerland                  .
## Nationality_Syria                        0.4542967748221321
## Nationality_Tanzania                     .
## Nationality_Thailand                     .
## Nationality_Togo                         .
## Nationality_Trinidad...Tobago            .
## Nationality_Tunisia                      .
## Nationality_Turkey                    2162.1795497841599172
## Nationality_Uganda                       .
## Nationality_Ukraine                  -7241.7979829341866207
## Nationality_United.Arab.Emirates         .
## Nationality_United.States            -1044.4198572383943429
## Nationality_Uruguay                      .
## Nationality_Uzbekistan                   .
## Nationality_Venezuela                -1204.7617073567621446
## Nationality_Wales                     4216.1231917654749850
## Nationality_Zambia                       .
## Nationality_Zimbabwe                   -54.4907238864165677
```

```
# the number of non-zero coefficient (of predictors)
length(which(coef.ridge[-(1:2),]!=0))
```

```
## [1] 95
```

```
which(coef.ridge[-(1:2),]!=0)
```

```
##                                  Age                              Overall
##                                    1                                    2
##                            Potential                                Value
##                                    3                                    4
##             International.Reputation                          Skill.Moves
##                                    5                                    7
##                            Work.Rate                               Height
##                                    8                                    9
##                               Weight                             Crossing
##                                   10                                   11
##                 Body.Type_C..Ronaldo                    Body.Type_Courtois
##                                   14                                   15
##                       Body.Type_Lean                       Body.Type_Messi
##                                   16                                   17
##                     Body.Type_Neymar                      Body.Type_Other
##                                   18                                   20
##                    Body.Type_Shaqiri                     Body.Type_Stocky
##                                   21                                   22
##                         Position_CAM                          Position_CB
```

```
##                                 23                                  24
##                       Position_CDM                         Position_CF
##                                 25                                  26
##                        Position_CM                         Position_GK
##                                 27                                  28
##                        Position_LB                        Position_LCB
##                                 30                                  31
##                       Position_LCM                        Position_LDM
##                                 32                                  33
##                        Position_LF                         Position_LM
##                                 34                                  35
##                        Position_LS                         Position_LW
##                                 36                                  37
##                        Position_RB                        Position_RCM
##                                 40                                  42
##                        Position_RF                         Position_RM
##                                 44                                  45
##                        Position_RS                         Position_RW
##                                 46                                  47
##                        Position_ST                 Preferred.Foot_Left
##                                 49                                  50
##                Preferred.Foot_Right               Nationality_Argentina
##                                 51                                  58
##                 Nationality_Armenia                 Nationality_Austria
##                                 59                                  61
##                 Nationality_Belgium                   Nationality_Benin
##                                 65                                  67
##                  Nationality_Brazil                Nationality_Bulgaria
##                                 72                                  73
##              Nationality_Cape.Verde Nationality_Central.African.Rep.
##                                 78                                  79
##                   Nationality_Chile                Nationality_China.PR
##                                 81                                  82
##                Nationality_Colombia              Nationality_Costa.Rica
##                                 83                                  86
##                 Nationality_Croatia          Nationality_Czech.Republic
##                                 87                                  91
##      Nationality_Dominican.Republic               Nationality_DR.Congo
##                                 93                                  94
##                 Nationality_Ecuador                 Nationality_England
##                                 95                                  98
##       Nationality_Equatorial.Guinea                   Nationality_Gabon
##                                 99                                 108
##                 Nationality_Georgia                  Nationality_Greece
##                                110                                 113
##                Nationality_Honduras                 Nationality_Hungary
##                                121                                 123
##                 Nationality_Iceland                   Nationality_Japan
##                                124                                 132
##          Nationality_Korea.Republic                  Nationality_Mexico
##                                137                                 150
##             Nationality_Netherlands        Nationality_Northern.Ireland
```

```
##                             157                              163
##               Nationality_Norway                 Nationality_Peru
##                             164                              169
##          Nationality_Philippines              Nationality_Poland
##                             170                              171
##             Nationality_Portugal  Nationality_Republic.of.Ireland
##                             172                              175
##               Nationality_Russia        Nationality_Saudi.Arabia
##                             177                              180
##             Nationality_Scotland             Nationality_Senegal
##                             181                              182
##               Nationality_Serbia            Nationality_Slovakia
##                             183                              185
##             Nationality_Slovenia        Nationality_South.Africa
##                             186                              187
##       Nationality_St.Kitts.Nevis              Nationality_Sweden
##                             190                              193
##                Nationality_Syria              Nationality_Turkey
##                             195                              201
##              Nationality_Ukraine       Nationality_United.States
##                             203                              205
##            Nationality_Venezuela               Nationality_Wales
##                             208                              209
##             Nationality_Zimbabwe
##                             211
```

```
coef_ridge <- coef.ridge[-(1:2),] %>% as.data.frame()
names(coef_ridge) <- "Coef.Est."
coef_ridge<- coef_ridge %>% arrange(abs(coef_ridge$Coef.Est.))
```

Random Forest

```
# Random Forest
mod.rf = randomForest(Wage~., data=data, subset=train, mtry=7, importance=TRUE)
yhat.rf = predict(mod.rf, newdata=data[-train,])
rf.test = data[-train, "Wage"]
testerror.rf = mean((yhat.rf-rf.test)^2)
importance(mod.rf)
```

```
##                                %IncMSE  IncNodePurity
## Age                        13.13315366  59074339390.6
## Overall                    21.76741215 577335448331.9
## Potential                  17.58954910 349085154703.4
## Value                      21.61971732 585453742853.4
## International.Reputation   17.44845456 338084313088.2
## Weak.Foot                   1.97390397  22662946783.0
## Skill.Moves                 7.34469033  91301489671.1
## Work.Rate                   2.80482875  30772209982.7
## Height                      6.49212315  21727608937.2
## Weight                      5.06804327  37555339688.1
## Crossing                   12.45831779 156938470465.7
## Finishing                  10.39944420 139354309445.0
## Body.Type_Akinfenwa         0.00000000      2667381.5
## Body.Type_C..Ronaldo        0.00000000            0.0
```

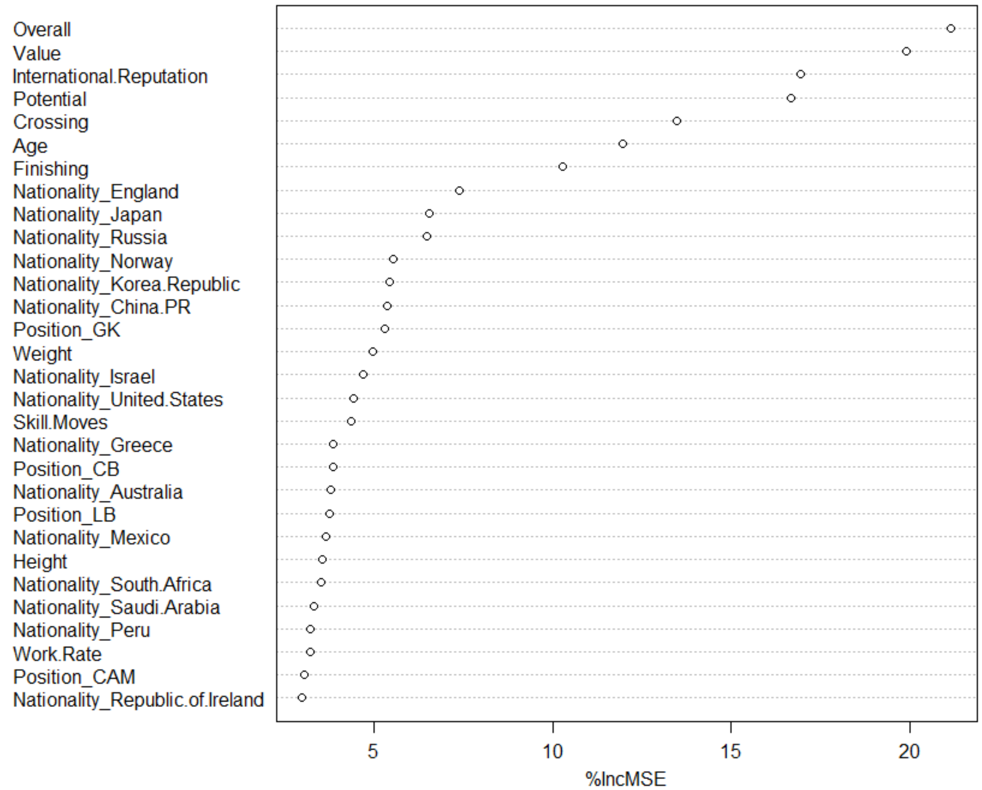
```
## Body.Type_Courtois                0.00000000          0.0
## Body.Type_Lean                    2.06775695 6706864127.5
## Body.Type_Messi                   0.00000000          0.0
## Body.Type_Neymar                  0.00000000          0.0
## Body.Type_Normal                 -1.66374312 5575828022.9
## Body.Type_Other                   0.00000000 9244821988.2
## Body.Type_Shaqiri                 0.00000000          0.0
## Body.Type_Stocky                  0.17294342 2931583006.7
## Position_CAM                     -0.18466021 2138303920.8
## Position_CB                       4.28114441 2496368675.4
## Position_CDM                      0.14898222 6214842317.2
## Position_CF                       0.97901282  439381512.9
## Position_CM                       1.57377426 3097211594.9
## Position_GK                       4.29778497 3312562371.3
## Position_LAM                      0.41265076  153340440.4
## Position_LB                       2.89625098 1818593802.7
## Position_LCB                     -0.47427207 3164383886.3
## Position_LCM                      0.49493778 1517822699.2
## Position_LDM                     -0.94678385 2625104785.3
## Position_LF                      -2.27994322 5590494018.0
## Position_LM                      -0.83659707 3062775309.9
## Position_LS                       2.60022362 1394241032.8
## Position_LW                      -1.04947057 3365315095.7
## Position_LWB                     -0.21370470  479842015.9
## Position_RAM                     -1.13658896 1021376743.7
## Position_RB                       2.62565268 2606470794.9
## Position_RCB                      1.52337464 4701865838.3
## Position_RCM                     -0.65235482 5407684368.2
## Position_RDM                      0.35464131 1700365294.3
## Position_RF                      -1.38698631  596630129.4
## Position_RM                      -0.97319820 3062700621.2
## Position_RS                       1.28955581 7234775572.9
## Position_RW                      -2.58828394 3607200597.5
## Position_RWB                      0.80612463  450049235.5
## Position_ST                       2.64991076 6457888801.5
## Preferred.Foot_Left               0.95285286 4858824245.2
## Preferred.Foot_Right             -0.01336665 4456260577.7
## Nationality_Afghanistan           0.00000000          0.0
## Nationality_Albania               0.23405248   64753312.6
## Nationality_Algeria              -1.06813509 2234592071.9
## Nationality_Andorra               0.00000000    2333381.2
## Nationality_Angola                0.98938492    8270901.1
## Nationality_Antigua...Barbuda    -0.43520278    3611817.6
## Nationality_Argentina             1.38063353 2347022713.5
## Nationality_Armenia              -0.93360992 2127555566.7
## Nationality_Australia             3.41946738  647894193.3
## Nationality_Austria              -0.04165737  710091822.4
## Nationality_Azerbaijan           -1.00100150    7445168.1
## Nationality_Barbados              0.00000000    1589637.2
## Nationality_Belarus               1.28731818    9310479.5
## Nationality_Belgium              -1.66452912 8781890532.4
## Nationality_Belize                0.00000000    3644492.6
```

```
## Nationality_Benin                  1.35799268     10999125.2
## Nationality_Bermuda                0.00000000            0.0
## Nationality_Bolivia               -0.94536857      8001445.9
## Nationality_Bosnia.Herzegovina     1.00409858   1905635681.8
## Nationality_Botswana               0.00000000            0.0
## Nationality_Brazil                 0.62973322   7835687319.8
## Nationality_Bulgaria               1.08079306     55514722.7
## Nationality_Burkina.Faso           1.01506925    514649570.1
## Nationality_Burundi               -2.11554560     44133086.9
## Nationality_Cameroon              -1.58715675    773512973.5
## Nationality_Canada                 2.25414803    201447278.2
## Nationality_Cape.Verde            -0.25800943    428793625.7
## Nationality_Central.African.Rep. -2.04566011    229994801.8
## Nationality_Chad                   0.00000000      1427012.0
## Nationality_Chile                  2.46462643    888499670.2
## Nationality_China.PR               4.23405682   1317000952.6
## Nationality_Colombia               2.22166878   2242323164.0
## Nationality_Comoros                1.24578114      9403697.4
## Nationality_Congo                  1.15224161     44788933.3
## Nationality_Costa.Rica            -0.01540030    210061533.4
## Nationality_Croatia               -0.56092284   9631500857.4
## Nationality_Cuba                   0.00000000            0.0
## Nationality_Curacao                0.70319136     44941873.1
## Nationality_Cyprus                -0.85529295     16621107.6
## Nationality_Czech.Republic         2.35468384   1008807102.2
## Nationality_Denmark               -1.12643210   1685112164.2
## Nationality_Dominican.Republic    -0.55464800   3999602871.7
## Nationality_DR.Congo               4.86523126    237289921.3
## Nationality_Ecuador                0.71991015   2011123692.5
## Nationality_Egypt                  0.86464298   3500714765.0
## Nationality_El.Salvador            0.00000000       923717.8
## Nationality_England                6.86834500   8136087335.1
## Nationality_Equatorial.Guinea     -0.85050259      4932342.0
## Nationality_Eritrea                0.00000000      1641780.7
## Nationality_Estonia               -2.57789745     56385815.0
## Nationality_Ethiopia               0.00000000      5498630.6
## Nationality_Faroe.Islands          0.00000000            0.0
## Nationality_Fiji                   0.00000000      3917047.1
## Nationality_Finland                2.16002352    117565203.3
## Nationality_France                -1.89925849   5803809708.5
## Nationality_FYR.Macedonia          0.06687352     91586334.4
## Nationality_Gabon                 -2.04984498   5451557983.3
## Nationality_Gambia                 0.60495531    115425259.4
## Nationality_Georgia                1.99959924     98900670.7
## Nationality_Germany                1.74865236   2397169987.1
## Nationality_Ghana                  2.36713022    277588404.2
## Nationality_Greece                 3.18935656   1014710223.2
## Nationality_Grenada                0.00000000       906345.8
## Nationality_Guam                   0.00000000      1108791.6
## Nationality_Guatemala             -0.53823229      4469337.9
## Nationality_Guinea                -1.56429223    208734632.9
## Nationality_Guinea.Bissau          1.26979722     56829195.0
```

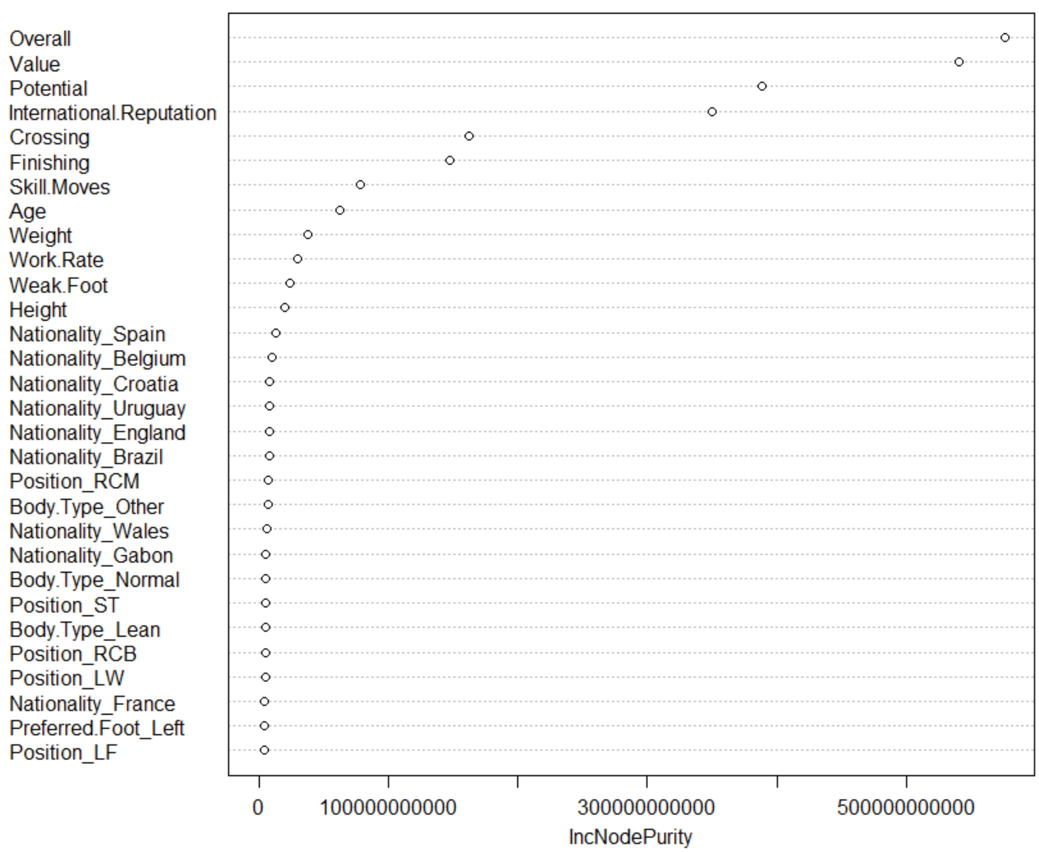
```
## Nationality_Guyana                  1.01476177     3900378.0
## Nationality_Haiti                   1.04804728    48451989.0
## Nationality_Honduras               -0.26691513   109791675.9
## Nationality_Hong.Kong               0.00000000           0.0
## Nationality_Hungary                 0.59473529   217265534.3
## Nationality_Iceland                -1.05935933   294924754.7
## Nationality_Indonesia               0.00000000     1952568.8
## Nationality_Iran                   -1.95880317    47979733.7
## Nationality_Iraq                   -1.66418107     4977804.7
## Nationality_Israel                  2.97329976   456397591.0
## Nationality_Italy                  -1.72904525  3175882984.7
## Nationality_Ivory.Coast             0.48972540   974072782.3
## Nationality_Jamaica                 0.09649999   273528502.5
## Nationality_Japan                   5.12682932  2214944113.6
## Nationality_Jordan                  0.00000000      515350.9
## Nationality_Kazakhstan              0.95631527     6358305.6
## Nationality_Kenya                   0.81867546    11879018.2
## Nationality_Korea.DPR              -1.10124169     4437200.4
## Nationality_Korea.Republic          6.50918812  2078233854.1
## Nationality_Kosovo                 -1.07474241    45548608.4
## Nationality_Kuwait                  0.00000000           0.0
## Nationality_Latvia                  1.24598140     5962007.5
## Nationality_Lebanon                 0.00000000      563305.6
## Nationality_Libya                   0.00000000    24720952.3
## Nationality_Liechtenstein          -0.22585378     3323249.8
## Nationality_Lithuania               0.28347215    22941869.9
## Nationality_Luxembourg             -1.03302560     4165749.8
## Nationality_Madagascar             -2.21029283   252257262.9
## Nationality_Mali                   -1.06270793   212709872.3
## Nationality_Mauritania              0.00000000     2017292.1
## Nationality_Mauritius               0.00000000      998823.9
## Nationality_Mexico                  4.19557220  1400165613.4
## Nationality_Moldova                 1.75994765    28779486.4
## Nationality_Montenegro              0.84778478    39773476.4
## Nationality_Montserrat              0.00000000           0.0
## Nationality_Morocco                 1.00820856   494905783.1
## Nationality_Mozambique              1.01706895    56220380.1
## Nationality_Namibia                 0.00000000     4112796.0
## Nationality_Netherlands             2.52000029  1737186725.0
## Nationality_New.Caledonia           0.00000000     1623255.2
## Nationality_New.Zealand             1.14382020    81301886.6
## Nationality_Nicaragua               0.00000000     2404996.5
## Nationality_Niger                   1.24604735     1844525.1
## Nationality_Nigeria                 0.73014389   456956111.7
## Nationality_Northern.Ireland        0.61622067   531602761.8
## Nationality_Norway                  5.08606035  1556371904.8
## Nationality_Oman                    0.00000000     7384648.9
## Nationality_Palestine               0.00000000           0.0
## Nationality_Panama                 -0.38271957    31137643.0
## Nationality_Paraguay               -0.43203964   809912694.2
## Nationality_Peru                    1.01092809   230432607.0
## Nationality_Philippines             0.00000000           0.0
```

```
## Nationality_Poland                  0.76203322   1193507728.3
## Nationality_Portugal                1.58671633   3275814146.5
## Nationality_Puerto.Rico             0.00000000          0.0
## Nationality_Qatar                   0.00000000       981990.6
## Nationality_Republic.of.Ireland     3.54381876   1148482764.1
## Nationality_Romania                -1.51382869    301052291.5
## Nationality_Russia                  4.88560490   1103435117.3
## Nationality_Rwanda                  0.00000000       756789.3
## Nationality_São.Tomé...Príncipe     0.00000000      8163242.8
## Nationality_Saudi.Arabia            3.15872159    526490803.0
## Nationality_Scotland                1.40225407   1006159049.7
## Nationality_Senegal                -1.41317804   1676312493.9
## Nationality_Serbia                  0.71479664   1323273298.0
## Nationality_Sierra.Leone           -1.00100150      5934458.0
## Nationality_Slovakia                0.50436168    199627602.5
## Nationality_Slovenia                1.63216108   1216728586.7
## Nationality_South.Africa            2.47113071    298466113.1
## Nationality_South.Sudan             0.00000000      2807769.1
## Nationality_Spain                  -3.86779796  11954561696.2
## Nationality_St.Kitts.Nevis          0.00000000      3752809.9
## Nationality_Sudan                   1.12632412      8431421.0
## Nationality_Suriname                1.41577932      6734468.3
## Nationality_Sweden                  4.15260873   1831696486.3
## Nationality_Switzerland             1.32749486   1061472722.9
## Nationality_Syria                  -2.03033997    167656032.0
## Nationality_Tanzania               -0.51621077     27505871.2
## Nationality_Thailand                0.00000000      2710287.1
## Nationality_Togo                    1.56743549    180776152.0
## Nationality_Trinidad...Tobago       1.37884512     11820406.9
## Nationality_Tunisia                 1.07416343     85380647.1
## Nationality_Turkey                  0.96779346   1110453433.8
## Nationality_Uganda                 -1.71587764       971674.1
## Nationality_Ukraine                -0.69032073    889269027.2
## Nationality_United.Arab.Emirates    0.00000000    115200971.6
## Nationality_United.States           4.63515260   1292140879.2
## Nationality_Uruguay                -1.21461895   5787149516.2
## Nationality_Uzbekistan             -2.50140842     17690866.1
## Nationality_Venezuela              -0.86687176    376631572.2
## Nationality_Wales                   0.56279480   8314717096.4
## Nationality_Zambia                  1.45507106     10444495.1
## Nationality_Zimbabwe               -0.84652380    123865674.7

varImpPlot(mod.rf)
```



Overall
Value
International.Reputation
Potential
Crossing
Age
Finishing
Nationality_England
Nationality_Japan
Nationality_Russia
Nationality_Norway
Nationality_Korea.Republic
Nationality_China.PR
Position_GK
Weight
Nationality_Israel
Nationality_United.States
Skill.Moves
Nationality_Greece
Position_CB
Nationality_Australia
Position_LB
Nationality_Mexico
Height
Nationality_South.Africa
Nationality_Saudi.Arabia
Nationality_Peru
Work.Rate
Position_CAM
Nationality_Republic.of.Ireland
5
10
15
20
%IncMSE



Overall
Value
Potential
International.Reputation
Crossing
Finishing
Skill.Moves
Age
Weight
Work.Rate
Weak.Foot
Height
Nationality_Spain
Nationality_Belgium
Nationality_Croatia
Nationality_Uruguay
Nationality_England
Nationality_Brazil
Position_RCM
Body.Type_Other
Nationality_Wales
Nationality_Gabon
Body.Type_Normal
Position_ST
Body.Type_Lean
Position_RCB
Position_LW
Nationality_France
Preferred.Foot_Left
Position_LF
0
100000000000
300000000000
500000000000
IncNodePurity

Overview

```r
# Overview of test errors
testerror.ls

## [1] 118531610

testerror.lasso

## [1] 117873828

testerror.ridge

## [1] 118420695

testerror.rf

## [1] 171586047

MSE <- c(testerror.ls, testerror.lasso, testerror.ridge, testerror.rf)
names(MSE) <- c("Linear Regression", "Lasso", "Ridge", "Random Forest")
MSE

## Linear Regression             Lasso             Ridge     Random Forest
##         118531610         117873828         118420695         171586047

which.min(MSE)

## Lasso
##     2
```

Conclusion: Lasso performs the best. Lasso tends to select one regressor and not the other.

Since we have perfect multicollinearity among our dummy variables,

Lasso performs the best by setting perfectly correlated variables to zero.

1b.

Linear Regression

The best 10 covariates from linear regression are: Value, International.Reputation, Age, Nationality_Dominican.Republic, Position_LS, Position_CAM, Skill.Moves, Position_LF, Position_RF, Body.Type_Other.

Lasso

The best 10 covariates from Lasso are: Body.Type_Messi, Body.Type_C..Ronaldo, Body.Type_Neymar, Body.Type_Shaqiri, Body.Type_Courtois, Nationality_Dominican.Republic, Body.Type_Other, Nationality_Central.African.Rep. Position_LF, Nationality_Russia

Ridge

The best 10 covariates from Ridge are: Body.Type_Messi, Body.Type_C..Ronaldo, Body.Type_Neymar, Body.Type_Shaqiri, Body.Type_Courtois, Nationality_Dominican.Republic, Body.Type_Other, Nationality_Central.African.Rep. Position_LF, Nationality_Russia

Random Forest

The best 10 covariates from random forest are: Age, Overall, Potential, Value, International.Reputation, Weak.Foot, Skill.Moves, Work.Rate,Height, Weight.

1c.

Messi, Ronaldo and Neymar are the represent players. They have the highest coefficients in the

Lasso estimation meaning that they have the highest importance in explaining the

wages.

QUESTION 02:

a) $\min\limits_{\beta} \sum_{i=1}^{N} (y_i - \beta_i)^2 + \lambda \sum_{i=3}^{N} \left[(\beta_i - \beta_{i-1}) - (\beta_{i-1} - \beta_{i-2}) \right]^2$

$= \min\limits_{\beta} \sum_{i=1}^{N} (y_i - \beta_i)^2 + \lambda \sum_{i=3}^{N} \left[\beta_i - 2\beta_{i-1} + \beta_{i-2} \right]^2$

b) $y = (y_1, \dots, y_n)'$, $\beta = (\beta_1, \dots, \beta_n)'$

$\min\limits_{\beta} (y - \beta)'(y - \beta) + \lambda (Q\beta)'(Q\beta)$

$= \min\limits_{\beta} (y' - \beta')(y - \beta) + \lambda \beta' Q' Q \beta$

$= \min\limits_{\beta} y'y - y'\beta - \beta' y + \beta'\beta + \lambda \beta' Q' Q \beta$

$= \min\limits_{\beta} \beta'(I + \lambda Q'Q)\beta - 2y'\beta + y'y$

where $Q = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 & 0 \\ & & \vdots & & & & \\ 0 & 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix}$

c) FOC wrt β:

$\frac{\partial}{\partial \beta} \left[\beta'(I + \lambda Q'Q)\beta - 2y'\beta + y'y \right] \Big|_{\beta = \hat{\beta}} = 0$

$\Rightarrow \cancel{2}(I + \lambda Q'Q)\hat{\beta} - \cancel{2}y = 0$

$\Rightarrow \hat{\beta} = (I + \lambda Q'Q)^{-1} y$

It look similar to the Ridge estimator, which is given by $\hat{\beta}_{RIDGE} = (X'X + \lambda I_p)^{-1} X^T y$

d) $\hat{\beta} = (I + \lambda Q'Q)^{-1} y$

Expectation

$$E[\hat{\beta} \mid x] = E\left[(I + \lambda Q'Q)^{-1}(\beta + \varepsilon) \mid x\right] \quad \pm (Q'Q\lambda + I)^{-1}\lambda Q'Q\beta$$

$$= E[(I + \lambda Q'Q)^{-1}\beta + (I + \lambda Q'Q)^{-1}\varepsilon \mid x]$$

$$= E[(I + \lambda Q'Q)^{-1}\beta + (I + \lambda Q'Q)^{-1}\lambda Q'Q\beta + (I + \lambda Q'Q)^{-1}\varepsilon - (I + \lambda Q'Q)^{-1}\lambda Q'Q\beta \mid x]$$

$$= E[(I + \lambda Q'Q)^{-1}(\beta + \lambda Q'Q\beta) + (I + \lambda Q'Q)^{-1}(\varepsilon + \lambda Q'Q\beta) \mid x]$$

$$= E[(I + \lambda Q'Q)^{-1}(I + \lambda Q'Q)\beta + \overbrace{(I + \lambda Q'Q)^{-1}}^{W(\lambda)}(\varepsilon + \lambda Q'Q\beta) \mid x]$$

$$= \beta + W(\lambda)\underbrace{E(\varepsilon \mid x)}_{=0} + W(\lambda)\lambda Q'Q\beta = \beta + W(\lambda)\lambda Q'Q\beta$$

Variance

$$\hat{\beta} = (I + \lambda Q'Q)^{-1} y = W(\lambda) y \qquad \mathrm{Var}(\beta + \varepsilon) = I$$

$$\mathrm{Var}(\hat{\beta} \mid x) = \mathrm{Var}(W(\lambda) y \mid x) = W'(\lambda)\underbrace{\mathrm{Var}(y \mid x)}_{= \mathrm{Var}(\beta+\varepsilon)} W(\lambda)$$

$$= W(\lambda)' W(\lambda) = W(\lambda) W(\lambda)$$

e) If $\lambda = 0$, the expectation & variance will be as if we didn't have penalty.

$\lambda = 0$: $E(\hat{\beta} \mid x) = \beta$

$\mathrm{Var}(\hat{\beta} \mid x) = I$

QUESTION 03:

The objective function for lasso for a linear regression model:

$$\hat{\beta}_{lasso} = \arg\min_{\beta} \left\| y - \sum_{j=1}^{p} x_j \beta_j \right\|^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

The objective function for adaptive lasso for a linear regression model:

$$\hat{\beta}_{a.lasso} = \arg\min_{\beta} \left\| y - \sum_{j=1}^{p} x_j \beta_j \right\|^2 + \lambda \sum_{j=1}^{p} w_j |\beta_j|$$

where $w_j = \frac{1}{|\beta_j|^{\gamma}}, \gamma \in [1,2]$ is a known weights vector.

A researcher would prefer to use adaptive lasso when the lasso is inconsistent for variable selection. The adaptive lasso where adaptive weights w_j are introduced to penalize different coefficients in the ℓ_1 penalty (i.e., the second term in the objective function for lasso) enjoys the oracle properties, which means that it performs as well as if the true underlying model was given in advance. Nevertheless, the adaptive lasso entails a finite sample bias in the non-zero coefficient estimates. As sample size grows, the data-dependent weights for zero-coefficient predictors get inflated (to infinity), whereas the weights for nonzero coefficient predictors converge to a finite constant. The adaptive lasso should therefore be adopted so the researcher could select the true zeros only, unbiasedly (asymptotically) re-estimate large coefficients and small threshold estimates in finite samples.