FIFA 선수 분석

[5조] 조성우 심소은
Department of E-business, Ajou School of Business, Ajou University, South Korea

2020-12-22

1. 개요

주제

최근 손흥민 선수는 토트넘 소속으로 무서운 기세의 활약으로 한국선수중 최초로 푸스카스 상을 수상하는 등 만개하는 기량을 보여주며 어려운 시국 중 우리에게 몇없는 즐거움과 자부심을 안겨주고 있기 때문에, 본 조원들 뿐 아니라 많은 한국인들이 해외축구에 대한 관심과 애정을 높이고 있는것은 분명하다.

이렇게 높아지는 해외축구에 대한 국내 팬들의 열정을 빌어, 본 조는 해당 보고서에서 FIFA 선수 데이터와 UEFA 축구클럽 순위 데이터를 연계시켜 선수의 Feature 및 stat에 따라 선수가 어떤 등급의 팀에 소속될지 예측하는 분류모델을 제작하는것을 주 연구과제로 EDA,다중회귀분석,군집분석 등 추가적인 다양한 분석기법을 활용하며 FIFA 데이터에서 축구팬들이 정량적인 STAT를 통해 흥미를 가질 수 있는 다양한 insight를 제시해 내는 것을 연구 목표로 세우고자 한다.

데이터

데이터셋
FIFA에서 제공하는 **"FIFA Dataset"** 과 UEFA의 **"20-21 Club Ranking"** 데이터를 사용하고자 한다.

주요 변수 및 특징

- *position* :
  선수의 주 포지션을 표기하는 변수

- *stat* :
  경기내에서 측정되는 선수의 각종 수많은 stat을 관측한 변수들

- *club* :
  선수가 소속된 구단, UEFA에서 가져온 club순위 변수, 그리고 이를 통해 해당 보고서에서 만들어낼 클럽등급 파생변수

- *wage & value* :
  선수의 임금 및 가치 변수

연구법(분석기법)

- 1. EDA
- 2. 유사도분석
- 2. 군집분석
- 3. 다중회귀분석
- 4. 분류분석

세부 연구주제 & 가설

(1) 유사도분석
손흥민,케인 선수와 유사한 능력 및 포지션의 선수는 누구일까

(2) 군집분석
FIFA 데이터상 선수들의 군집은 어떻게 나눠질 것인가

(3) 다중회귀분석
FIFA 데이터상 선수들의 overall에 가장 큰 영향을 주는 Feature은 무엇이며
나머지 선수개인특성의 변수들에 따라 overall이 어느정도로 설명될 것인가?

(4) 분류분석
사전에 UEFA순위에 따라 임의로 나눠준 선수들이 현재 소속된 팀의 등급을 label로
분류를 실시하여, 손흥민 선수가 현재 소속된 팀의 grade에 적합한 선수인지 판별하는 분류기를 학습 및 테스트한다

2. 본문

2.1 데이터 수집 및 전처리

2.1.1 패키지 준비

```
library(tidyverse)
library(magrittr)
library(DataExplorer)
library(maps)
library(plotly)
library(DT)
library(tidytext)
library(gridExtra)
library(factoextra)

library(readxl)
library(dplyr)
library(ggpubr)
library(caret)
library(corrplot)
library(car)
library(olsrr)
library(MASS)
library(ggplot2)
```

2.1.2 데이터 불러오기

```
#데이터 불러오기 df
setwd('C:\\Users\\JSW\\Desktop\\학교\\1_학기\\3학년 2학기_2020_2\\강의
df <- read.csv("complete_data.csv", encoding = "UTF-8")[-1]

UEFA = read.csv('club_ranking.csv')
```

```
dim(df)
```

```
## [1] 18207    88
```

FIFA 원데이터의 차원 갯수

```
glimpse(df)
```

```
## Rows: 18,207
## Columns: 88
## $ ID                        <int> 158023, 20801, 190871, 193080, 1
## $ Name                      <chr> "L. Messi", "Cristiano Ronaldo",
## $ Age                       <int> 31, 33, 26, 27, 27, 27, 32, 31,
## $ Photo                     <chr> "https://cdn.sofifa.org/players/
## $ Nationality               <chr> "Argentina", "Portugal", "Brazil
## $ Flag                      <chr> "https://cdn.sofifa.org/flags/52
## $ Overall                   <int> 94, 94, 92, 91, 91, 91, 91, 91,
## $ Potential                 <int> 94, 94, 93, 93, 92, 91, 91, 91,
## $ Club                      <chr> "FC Barcelona", "Juventus", "Par
## $ Club.Logo                 <chr> "https://cdn.sofifa.org/teams/2/
## $ Value                     <chr> "110.5M", "77M", "118.5M", "72M"
## $ Wage                      <chr> "565K", "405K", "290K", "260K",
## $ Special                   <int> 2202, 2228, 2143, 1471, 2281, 21
## $ Preferred.Foot            <chr> "Left", "Right", "Right", "Right
## $ International.Reputation  <int> 5, 5, 5, 4, 4, 4, 4, 5, 4, 3, 4,
## $ Weak.Foot                 <int> 4, 4, 5, 3, 5, 4, 4, 4, 3, 3, 4,
## $ Skill.Moves               <int> 4, 5, 5, 1, 4, 4, 4, 3, 3, 1, 4,
## $ Work.Rate                 <chr> "Medium/ Medium", "High/ Low", "
## $ Body.Type                 <chr> "Messi", "C. Ronaldo", "Neymar",
## $ Real.Face                 <chr> "Yes", "Yes", "Yes", "Yes", "Yes
## $ Position                  <chr> "RF", "ST", "LW", "GK", "RCM", "
## $ Jersey.Number             <int> 10, 7, 10, 1, 7, 10, 10, 9, 15,
## $ Joined                    <chr> "01-Jul-04", "10-Jul-18", "03-Au
## $ Loaned.From               <chr> "", "", "", "", "", "", "", "",
## $ Contract.Valid.Until      <chr> "2021", "2022", "2022", "2020",
## $ Height                    <chr> "5'7", "6'2", "5'9", "6'4", "5'1
## $ Weight                    <chr> "159lbs", "183lbs", "150lbs", "1
## $ LS                        <chr> "88+2", "91+3", "84+3", "", "82+
```

```
## $ ST                   <chr> "88+2", "91+3", "84+3", "", "82+
## $ RS                   <chr> "88+2", "91+3", "84+3", "", "82+
## $ LW                   <chr> "92+2", "89+3", "89+3", "", "87+
## $ LF                   <chr> "93+2", "90+3", "89+3", "", "87+
## $ CF                   <chr> "93+2", "90+3", "89+3", "", "87+
## $ RF                   <chr> "93+2", "90+3", "89+3", "", "87+
## $ RW                   <chr> "92+2", "89+3", "89+3", "", "87+
## $ LAM                  <chr> "93+2", "88+3", "89+3", "", "88+
## $ CAM                  <chr> "93+2", "88+3", "89+3", "", "88+
## $ RAM                  <chr> "93+2", "88+3", "89+3", "", "88+
## $ LM                   <chr> "91+2", "88+3", "88+3", "", "88+
## $ LCM                  <chr> "84+2", "81+3", "81+3", "", "87+
## $ CM                   <chr> "84+2", "81+3", "81+3", "", "87+
## $ RCM                  <chr> "84+2", "81+3", "81+3", "", "87+
## $ RM                   <chr> "91+2", "88+3", "88+3", "", "88+
## $ LWB                  <chr> "64+2", "65+3", "65+3", "", "77+
## $ LDM                  <chr> "61+2", "61+3", "60+3", "", "77+
## $ CDM                  <chr> "61+2", "61+3", "60+3", "", "77+
## $ RDM                  <chr> "61+2", "61+3", "60+3", "", "77+
## $ RWB                  <chr> "64+2", "65+3", "65+3", "", "77+

## $ LB                   <chr> "59+2", "61+3", "60+3", "", "73+
## $ LCB                  <chr> "47+2", "53+3", "47+3", "", "66+
## $ CB                   <chr> "47+2", "53+3", "47+3", "", "66+
## $ RCB                  <chr> "47+2", "53+3", "47+3", "", "66+
## $ RB                   <chr> "59+2", "61+3", "60+3", "", "73+
## $ Crossing             <int> 84, 84, 79, 17, 93, 81, 86, 77,
## $ Finishing            <int> 95, 94, 87, 13, 82, 84, 72, 93,
## $ HeadingAccuracy      <int> 70, 89, 62, 21, 55, 61, 55, 77,
## $ ShortPassing         <int> 90, 81, 84, 50, 92, 89, 93, 82,
## $ Volleys              <int> 86, 87, 84, 13, 82, 80, 76, 88,
## $ Dribbling            <int> 97, 88, 96, 18, 86, 95, 90, 87,
## $ Curve                <int> 93, 81, 88, 21, 85, 83, 85, 86,
## $ FKAccuracy           <int> 94, 76, 87, 19, 83, 79, 78, 84,
## $ LongPassing          <int> 87, 77, 78, 51, 91, 83, 88, 64,
## $ BallControl          <int> 96, 94, 95, 42, 91, 94, 93, 90,
## $ Acceleration         <int> 91, 89, 94, 57, 78, 94, 80, 86,
## $ SprintSpeed          <int> 86, 91, 90, 58, 76, 88, 72, 75,
## $ Agility              <int> 91, 87, 96, 60, 79, 95, 93, 82,
## $ Reactions            <int> 95, 96, 94, 90, 91, 90, 90, 92,
## $ Balance              <int> 95, 70, 84, 43, 77, 94, 94, 83,
## $ ShotPower            <int> 85, 95, 80, 31, 91, 82, 79, 86,
## $ Jumping              <int> 68, 95, 61, 67, 63, 56, 68, 69,
```

```
## $ Stamina                <int> 72, 88, 81, 43, 90, 83, 89, 90,
## $ Strength               <int> 59, 79, 49, 64, 75, 66, 58, 83,
## $ LongShots              <int> 94, 93, 82, 12, 91, 80, 82, 85,
## $ Aggression             <int> 48, 63, 56, 38, 76, 54, 62, 87,
## $ Interceptions          <int> 22, 29, 36, 30, 61, 41, 83, 41,
## $ Positioning            <int> 94, 95, 89, 12, 87, 87, 79, 92,
## $ Vision                 <int> 94, 82, 87, 68, 94, 89, 92, 84,
## $ Penalties              <int> 75, 85, 81, 40, 79, 86, 82, 85,
## $ Composure              <int> 96, 95, 94, 68, 88, 91, 84, 85,
## $ Marking                <int> 33, 28, 27, 15, 68, 34, 60, 62,
## $ StandingTackle         <int> 28, 31, 24, 21, 58, 27, 76, 45,
## $ SlidingTackle          <int> 26, 23, 33, 13, 51, 22, 73, 38,
## $ GKDiving               <int> 6, 7, 9, 90, 15, 11, 13, 27, 11,
## $ GKHandling             <int> 11, 11, 9, 85, 13, 12, 9, 25, 8,
## $ GKKicking              <int> 15, 15, 15, 87, 5, 6, 7, 31, 9,
## $ GKPositioning          <int> 14, 14, 15, 88, 10, 8, 14, 33, 7
## $ GKReflexes             <int> 8, 11, 11, 94, 13, 8, 9, 37, 11,
## $ Release.Clause         <chr> "226.5M", "127.1M", "228.1M", "1
```

FIFA 원데이터의 모습

```
glimpse(UEFA)
```

```
## Rows: 429
## Columns: 10
## $ rank <chr> "1?", "2?", "3?", "4?", "5?", "6?", "7?", "8?", "9?"
## $ club <chr> "FC Bayern Munchen", "FC Barcelona", "Juventus", "Re
## $ X    <chr> "GER", "ESP", "ITA", "ESP", "ESP", "ENG", "FRA", "ES
## $ X.1  <chr> "22", "23", "33", "33", "29", "18", "20", "19", "26"
## $ X.2  <chr> "29", "25", "23", "32", "28", "22", "19", "21", "20"
## $ X.3  <chr> "20", "30", "21", "19", "20", "25", "19", "13", "19"
## $ X.4  <chr> "36", "24", "22", "17", "22", "25", "31", "26", "22"
## $ X.5  <chr> "20", "19", "19", "16", "16", "20", "17", "18", "10"
## $ X.6  <dbl> 127.0, 121.0, 118.0, 117.0, 115.0, 110.0, 106.0, 97.
## $ X.7  <dbl> 13.799, 17.885, 14.030, 17.885, 17.885, 17.713, 10.8
```

UEFA 클럽 순위 데이터의 모습

2.1.3 데이터 전처리

1.club_ranking data 전처리 및 임의의 클럽등급 파생변수 생성

이후 분류분석에 label로 활용하기 위한 선수 소속의 클럽등급 파생변수를 생성합니다.

```
UEFA = UEFA[,c(1,2)]

for(i in 1:429){
  UEFA$rank[i] = i

}


#해당 데이터에서 구간등급도 나눠 파생변수를 만들어준다
UEFA$rank = as.numeric(UEFA$rank)

club_grade = c(1:10)


UEFA$club_grade = c(rep('S',times=10),rep('A',times=10),rep('B',time
```

*2.*리그변수 할당

```
bundesliga <- c(
  "1. FC Nürnberg", "1. FSV Mainz 05", "Bayer 04 Leverkusen", "FC Ba
  "Borussia Dortmund", "Borussia Mönchengladbach", "Eintracht Frankf
  "FC Augsburg", "FC Schalke 04", "Fortuna Düsseldorf", "Hannover 96
  "Hertha BSC", "RB Leipzig", "SC Freiburg", "TSG 1899 Hoffenheim",
  "VfB Stuttgart", "VfL Wolfsburg", "SV Werder Bremen"
)
```

```
premierLeague <- c(
  "Arsenal", "Bournemouth", "Brighton & Hove Albion", "Burnley",
  "Cardiff City", "Chelsea", "Crystal Palace", "Everton", "Fulham",
  "Huddersfield Town", "Leicester City", "Liverpool", "Manchester Ci
  "Manchester United", "Newcastle United", "Southampton",
  "Tottenham Hotspur", "Watford", "West Ham United", "Wolverhampton

)

laliga <- c(
  "Athletic Club de Bilbao", "Atlético Madrid", "CD Leganés",
  "Deportivo Alavés", "FC Barcelona", "Getafe CF", "Girona FC",
  "Levante UD", "Rayo Vallecano", "RC Celta", "RCD Espanyol",
  "Real Betis", "Real Madrid", "Real Sociedad", "Real Valladolid CF"
  "SD Eibar", "SD Huesca", "Sevilla FC", "Valencia CF", "Villarreal
)

seriea <- c(
  "Atalanta","Bologna","Cagliari","Chievo Verona","Empoli", "Fiorent
  "Inter","Juventus","Lazio","Milan","Napoli","Parma","Roma","Sampdo
  "Torino","Udinese"

)

superlig <- c(
  "Akhisar Belediyespor","Alanyaspor", "Antalyaspor","Medipol Başakş
  "Bursaspor","Çaykur Rizespor","Fenerbahçe SK", "Galatasaray SK","G
  "Kayserispor","Atiker Konyaspor","MKE Ankaragücü", "Sivasspor","Tr
)

ligue1 <- c(
  "Amiens SC", "Angers SCO", "AS Monaco", "AS Saint-Étienne", "Dijon
  "FC Nantes", "FC Girondins de Bordeaux", "LOSC Lille", "Montpellie
  "OGC Nice", "Olympique Lyonnais","Olympique de Marseille", "Paris
  "RC Strasbourg Alsace", "Stade Malherbe Caen", "Stade de Reims", "
)

eredivisie <- c(
  "ADO Den Haag","Ajax", "AZ Alkmaar", "De Graafschap","Excelsior","
  "FC Utrecht", "Feyenoord","Fortuna Sittard", "Heracles Almelo","NA
  "PEC Zwolle", "PSV","SC Heerenveen","Vitesse","VVV-Venlo","Willem
)
```

```
liganos <- c(
  "Os Belenenses", "Boavista FC", "CD Feirense", "CD Tondela", "CD A
  "CD Nacional", "GD Chaves", "Clube Sport Marítimo", "Moreirense FC
  "Santa Clara", "SC Braga", "SL Benfica", "Sporting CP", "Vitória G
)


df %<>% mutate(
  League = case_when(
    Club %in% bundesliga ~ "Bundesliga",
    Club %in% premierLeague ~ "Premier League",
    Club %in% laliga ~ "La Liga",
    Club %in% seriea ~ "Serie A",
    Club %in% superlig ~ "Süper Lig",
    Club %in% ligue1 ~ "Ligue 1",
    Club %in% liganos ~ "Liga Nos",
    Club %in% eredivisie ~ "Eredivisie"
  ),
  Country = case_when(
    League == "Bundesliga" ~ "Germany",
    League == "Premier League" ~ "UK",
    League == "La Liga" ~ "Spain",
    League == "Serie A" ~ "Italy",
    League == "Süper Lig" ~ "Turkey",
    League == "Ligue 1" ~ "France",
    League == "Liga Nos" ~ "Portugal",
    League == "Eredivisie" ~ "Netherlands"
  )
) %>% filter(!is.na(League)) %>% mutate_if(is.factor, as.character)


rm(bundesliga, premierLeague, laliga, seriea, superlig, ligue1, ered
```

3.문자값 수정

```
head(df$Value)
```

```
## [1] "110.5M" "77M"    "118.5M" "72M"    "102M"   "93M"
```

```
# Player Value
df$Values <- str_remove_all(df$Value,"€")
df$Values <- str_replace_all(df$Values,"K", "000")
df$Values <- str_remove_all(df$Values,"M")

df$Values <- as.numeric(df$Values)

# Player Wage
df$Wages <- str_remove_all(df$Wage,"€")
df$Wages <- str_replace_all(df$Wages,"K", "000")

df$Wages <- as.numeric(df$Wages)

df <- df  %>% mutate(Values = if_else(df$Values < 1000 , Values * 10
```

4.포지션 클래스 생성

```
unique(df$Position)
```

```
##  [1] "RF"  "ST"  "LW"  "GK"  "RCM" "LF"  "RS"  "RCB" "LCM" "CB"
## [13] "CDM" "LS"  "LCB" "RM"  "LAM" "LM"  "LB"  "RDM" "RW"  "CM"
## [25] "CF"  "RWB" "LWB" ""
```

```
# 기존에 존재하던 position변수로 파생변수 Class변수 생성
defence <- c("CB", "RB", "LB", "LWB", "RWB", "LCB", "RCB")
midfielder <- c("CM", "CDM","CAM","LM","RM", "LAM", "RAM", "LCM", "R

df %<>% mutate(Class = if_else(Position %in% "GK", "Goal Keeper",
                               if_else(Position %in% defence, "Defen
                                       if_else(Position %in% midfiel

rm(defence, midfielder)
```

5.Height & Weight

```
df %<>%
  mutate(Height = round((as.numeric(str_sub(Height, start=1,end = 1)
         Weight = round(as.numeric(str_sub(Weight, start = 1, end =

levels(factor(df$Country))
```

```
## [1] "France"      "Germany"     "Italy"       "Netherlands" "Port
## [6] "Spain"       "Turkey"      "UK"
```

```
a = factor(df$Country)
levels(a)
```

```
## [1] "France"      "Germany"     "Italy"       "Netherlands" "Port
## [6] "Spain"       "Turkey"      "UK"
```

6.선호 주발의 수정

```
# 선호하는 주발의 교정

df %<>% filter(Preferred.Foot %in% c("Left", "Right"))
df$Preferred.Foot <- as.factor(as.character(df$Preferred.Foot))
```

7.변수명 재설정

```
#3.6 변수명 재설정
df %<>%
```

```
  rename(
    "Heading.Accuracy"= HeadingAccuracy,
    "Short.Passing"= ShortPassing,
    "FK.Accuracy" = FKAccuracy,
    "Long.Passing"= LongPassing,
    "Ball.Control"= BallControl,
    "Sprint.Speed"= SprintSpeed,
    "Shot.Power"= ShotPower,
    "Long.Shots"= LongShots,
    "Standing.Tackle"= StandingTackle,
    "Sliding.Tackle"= SlidingTackle,
    "GK.Diving"= GKDiving,
    "GK.Handling"= GKHandling,
    "GK.Kicking"= GKKicking,
    "GK.Positioning"= GKPositioning,
    "GK.Reflexes"= GKReflexes
  )
```

*8.*불필요 변수 제거

```
#3.7 불필요 변수 제거
df %<>% dplyr::select(-ID, -Body.Type, -Real.Face, -Joined, -Loaned.
```

*9.*데이터 병합

```
b = merge(df,UEFA,key="club") #FIFA와 UEFA 데이터 병합

#glimpse(b)
```

2.2 EDA & Visualization

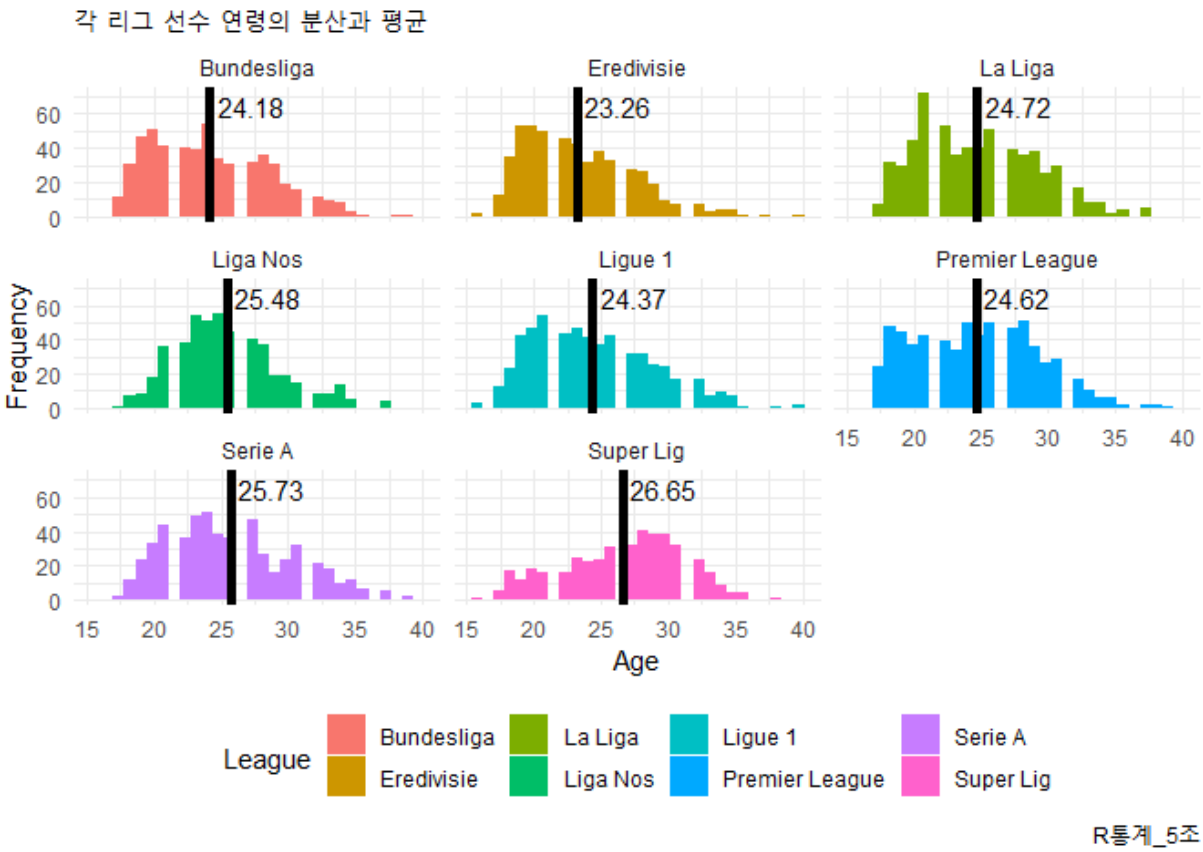
2.2.1 각 리그 선수들의 분산과 평균

```
#5.1. 각 리그 선수들의 연령 분산과 평균

summ <- df %>%
  group_by(League) %>%
  summarise(age = mean(Age))
```

```
## `summarise()` ungrouping output (override with `.groups` argument
```

```
options(repr.plot.width = 15, repr.plot.height = 6)

ggplot()+
  geom_histogram(df, mapping = aes(Age, fill = League))+
  geom_vline(summ, mapping = aes(xintercept = age), color = "black",
  geom_text(summ, mapping = aes(x = age+3, y = 65, label = round(age
  facet_wrap(League~.)+
  theme_minimal()+
  theme(legend.position = "bottom")+
  labs(y = "Frequency", title = "각 리그 선수 연령의 분산과 평균", cap
```

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

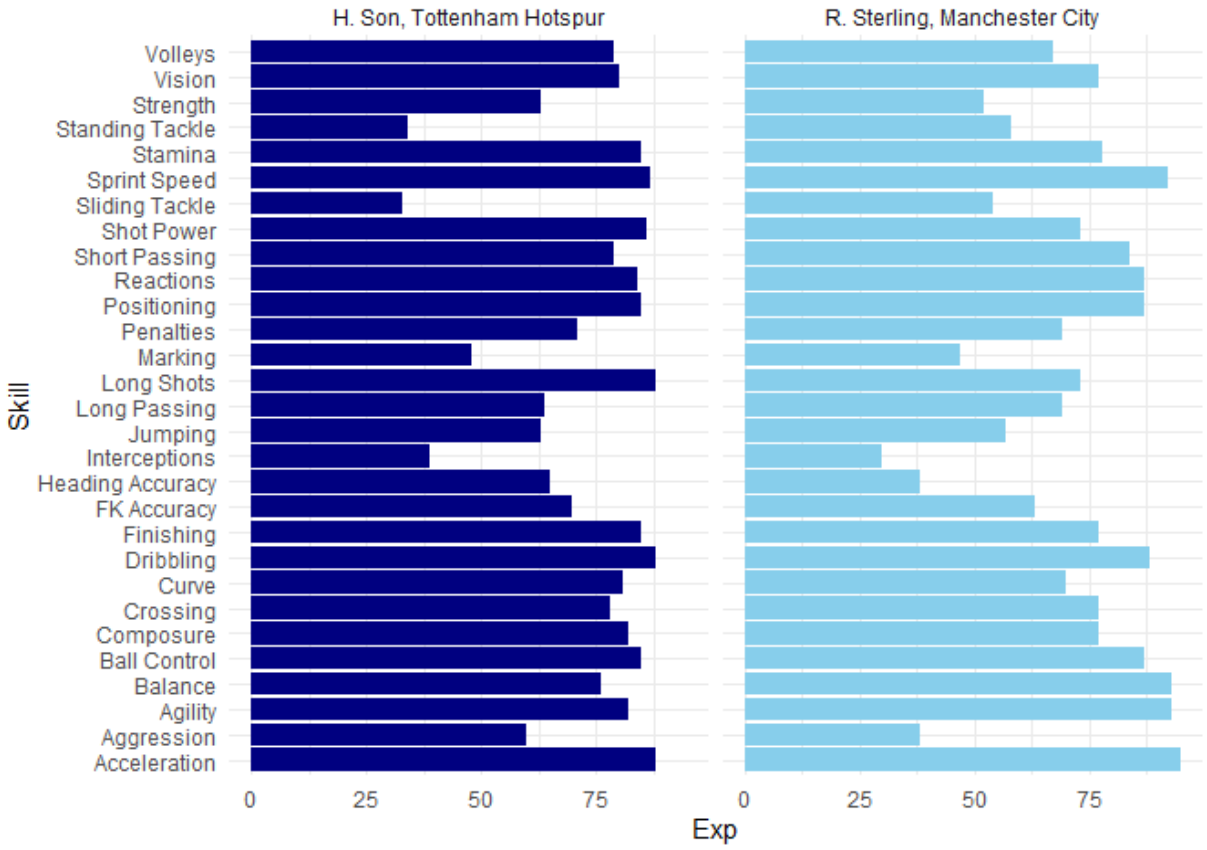


2.2.2 두 선수들 간의 비교

```
#5.4 두 선수들 간의 비교
library(tidyr)
# 비교대상이 될 두 선수 선정
players <- df %>%
  filter(Name %in% c("H. Son", "R. Sterling")) %>%
  # 이름 및 소속 클럽명
  mutate(Name = paste0(Name, ", ", Club)) %>%
  # 선수들의 비교기준 스탯 선정
  dplyr::select(Name,Crossing:Sliding.Tackle) %>%
  # 문자열수정
  rename_all(funs(gsub("[[:punct:]]", " ", .))) %>%
  # 변수에서 관측치로의 변형
  gather(Skill, Exp, Crossing:`Sliding Tackle`, -Name)
head(players  )
```

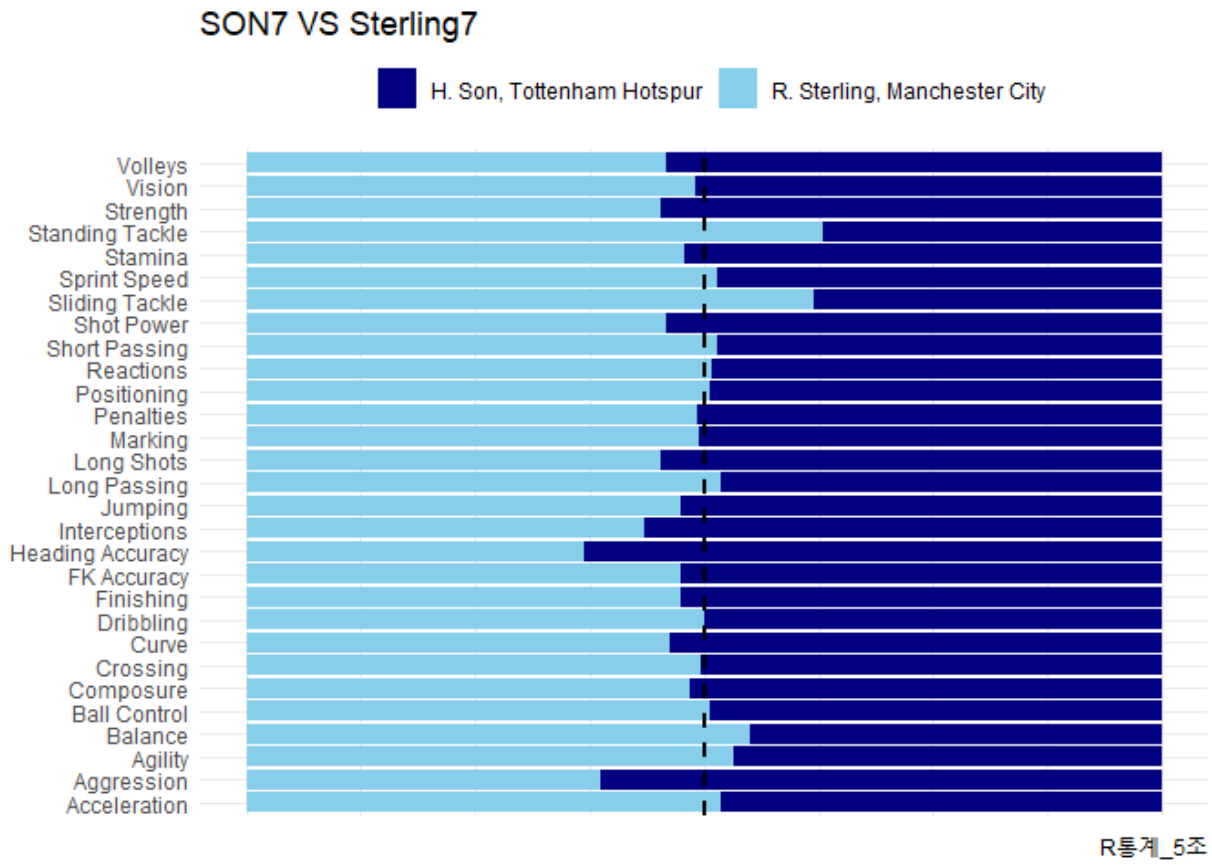
```
##                              Name            Skill Exp
## 1 R. Sterling, Manchester City          Crossing  77
## 2    H. Son, Tottenham Hotspur          Crossing  78
## 3 R. Sterling, Manchester City         Finishing  77
## 4    H. Son, Tottenham Hotspur         Finishing  85
## 5 R. Sterling, Manchester City Heading Accuracy  38
## 6    H. Son, Tottenham Hotspur Heading Accuracy  65
```

```
#두 선수 비교 시각화  <<1>>
options(repr.plot.width = 15, repr.plot.height = 8)

ggplot(players, aes(Skill, Exp, fill = Name))+
  geom_col(show.legend = FALSE)+
  coord_flip()+
  facet_wrap(Name~.)+
  scale_fill_manual(values = c("navy", "skyblue"))+
  theme_minimal()
```




```
#두 선수 비교 시각화 <<2>>
options(repr.plot.width = 15, repr.plot.height = 8)
```

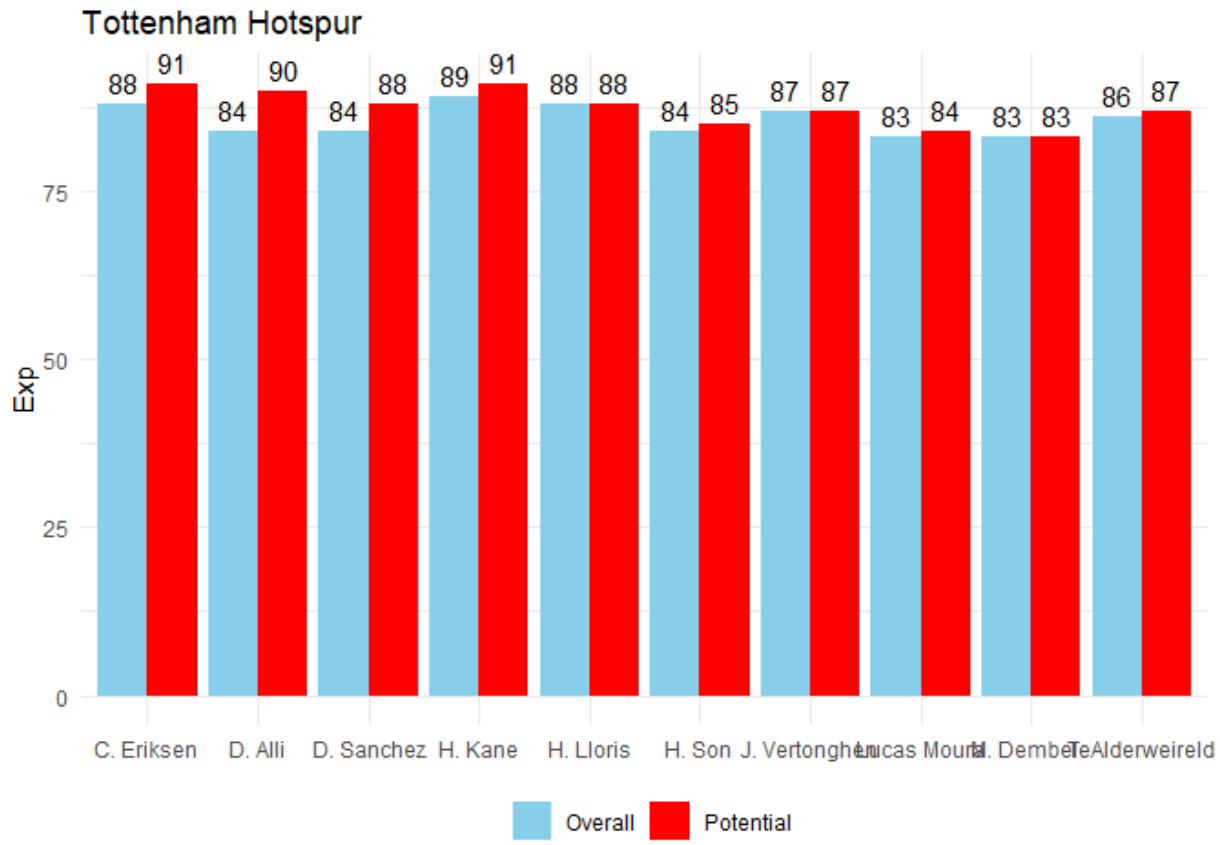
```
ggplot(players, aes(Skill, Exp, fill = Name))+
  geom_col(position = "fill")+
  coord_flip()+
  scale_fill_manual(values = c("navy", "skyblue"))+
  theme_minimal()+
  geom_hline(yintercept = 0.5, color = "black", size = 1, linetype =
  theme(legend.position = "top", axis.text.x=element_blank())+
  labs(title = "SON7 VS Sterling7",
       caption = "R통계_5조",
       fill = NULL,x = NULL, y = NULL)
```



프리미어리그의 각 팀에서 등번호 *7*번을 달고 윙포워드로 활약하고있는 손흥민 선수와 스털링 선수의 *stat*을 비교해본 결과, 두 선수는 매우 유사한 *stat*분포를 가지고 있음을 확인할 수 있었다.

2.2.3 토트넘 선수들의 잠재력&통합스탯

```
options(repr.plot.width = 12, repr.plot.height = 8)

df %>%
  filter(Club == "Tottenham Hotspur") %>%
  dplyr::select(Name, Overall, Potential) %>%
  arrange(-Overall) %>%
  head(10) %>%
  gather(variable, Exp, -Name) %>%
  ggplot(aes(Name, Exp, fill = variable))+
  geom_col(position = "dodge")+
  geom_text(aes(label = Exp),position = position_dodge(width = 0.9),
  scale_fill_manual(values = c("skyblue", "red"))+
  theme_minimal()+
  theme(legend.position = "bottom")+
  labs(fill = NULL, x = NULL, title = "Tottenham Hotspur")
```

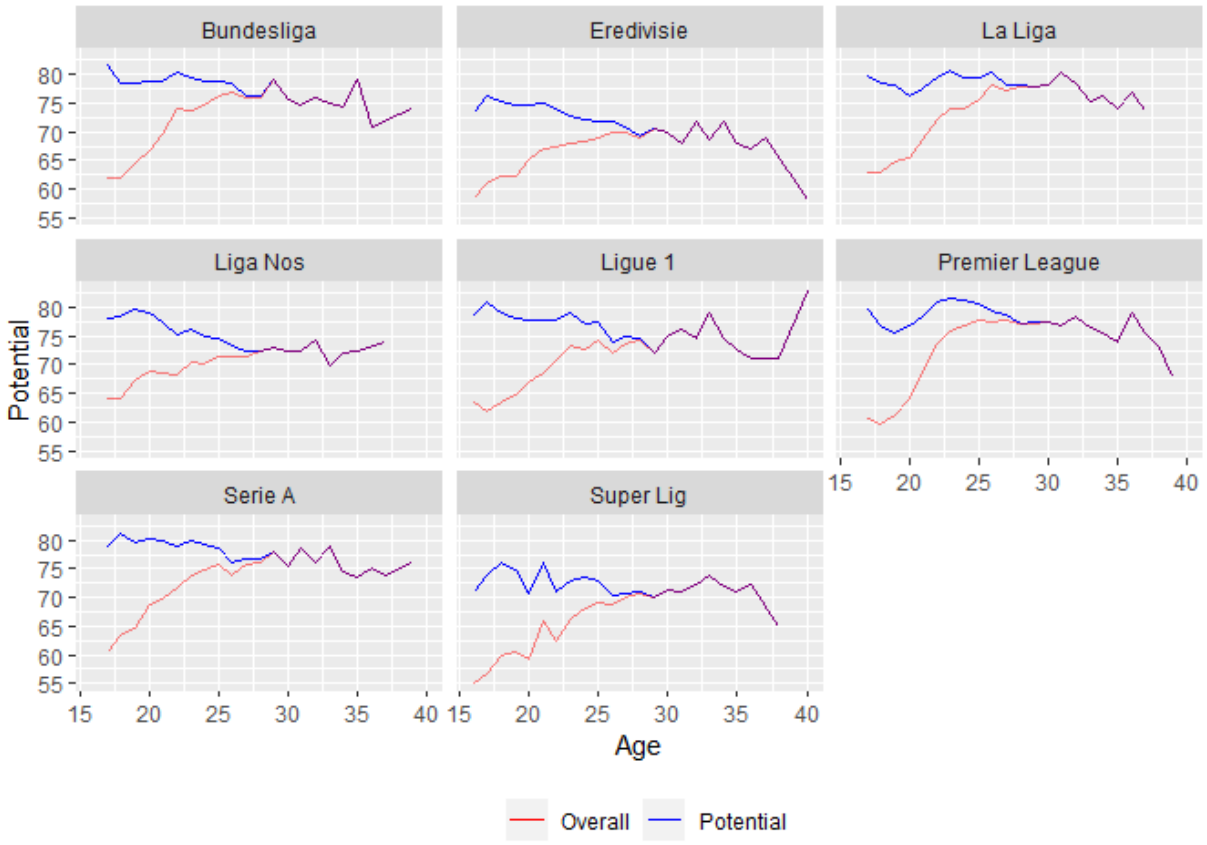


토트넘 내 선수들의 *Overall* 및 *Potential*을 시각화 해본 결과 *Harry Kane*이 가장 높은 *Overall*과 *Potential*을 가진 선수임을 확인할 수 있다

2.2.4 나이에 따른 소속 리그별 평균 잠재력 및 통합스탯

```
options(repr.plot.width = 12, repr.plot.height = 8)

df %>%
  group_by(League, Age) %>%
  summarise(Overall = mean(Overall),
            Potential = mean(Potential)) %>%
  ggplot()+
  geom_line(aes(Age, Potential, color = "Potential")) +
  geom_line(aes(Age, Overall, color = "Overall"), alpha = 0.5) +
  facet_wrap(League~.)+
  scale_color_manual(values = c("red", "blue"))+
  theme(legend.position = "bottom")+
  labs(color = NULL)
```

```
## `summarise()` regrouping output by 'League' (override with `.grou
```



나이에 따른 소속 리그별 평균 잠재력 및 *Overall*의 분포를 확인해 본 결과 유럽의 *4*대 리그라 불리는 분데스리가,프리미어리그,프리메라리가,세리에가 예상처

럼 높은 *Overall* 및 *stat*분포를 보이고 있음을 확인할 수 있다.

한가지 더 주목해보자면, 모든 리그를 통틀어 약 *25*세쯤부터 선수들의 *Potential*과 *Overall*이 일치하여 동일선을 그리는 것을 확인할 수 있다.

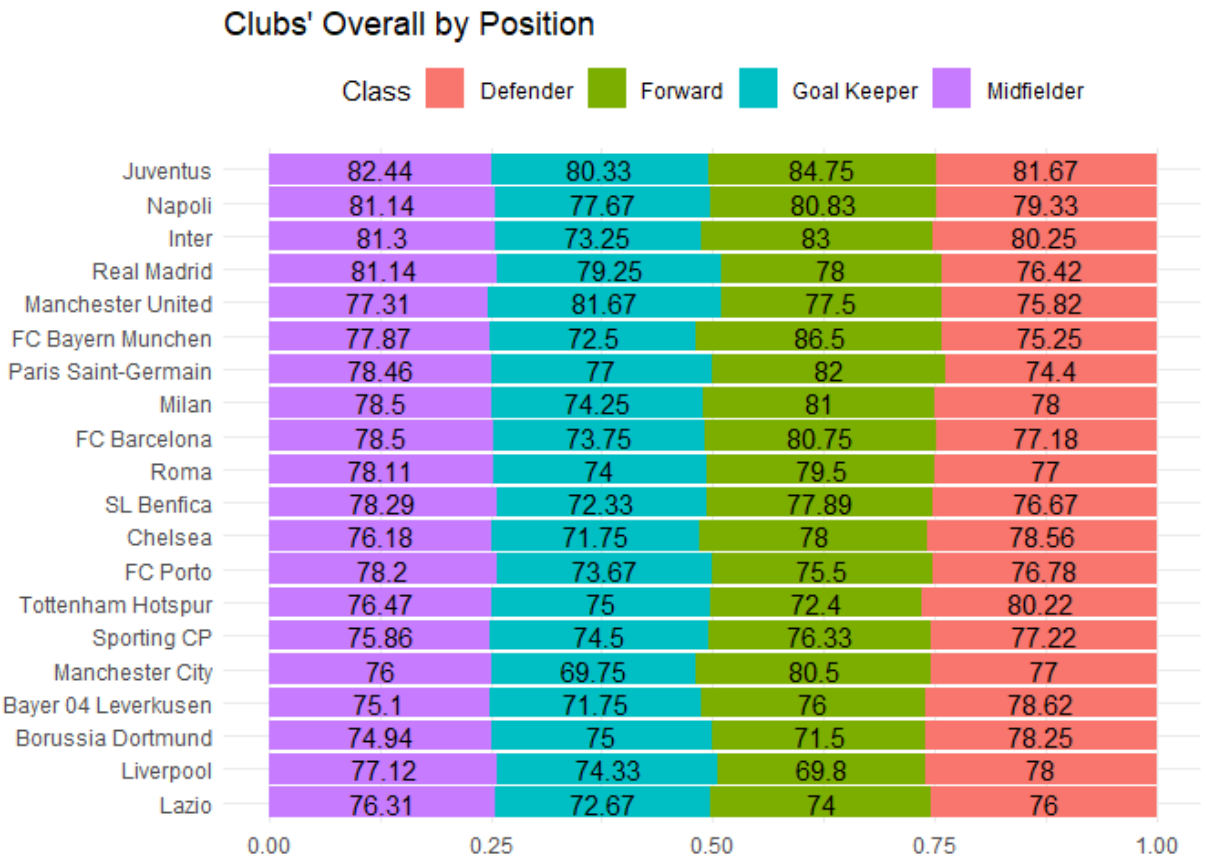
2.2.5 나이에 따른 소속 리그별 평균 잠재력 및 통합스탯

```
#5.11 가장 강한 클럽

options(repr.plot.width = 12, repr.plot.height = 8)

powerful <- df %>%
  group_by(Club) %>%
  summarise(mean = mean(Overall)) %>%
  arrange(-mean) %>%
  head(20)
```

```
## `summarise()` ungrouping output (override with `.groups` argument
```

```
df %>%
  group_by(Club, Class) %>%
  summarise(mean = mean(Overall)) %>%
  ungroup() %>%
  filter(Club %in% powerful$Club) %>%
  ggplot(aes(reorder(Club, mean), mean, fill = Class))+
  geom_col(position = "fill")+
  geom_text(aes(label = round(mean,digits = 2)), position = position
  coord_flip()+
  theme_minimal()+
  theme(legend.position = "top")+
  labs(x = NULL, y = NULL, title = "Clubs' Overall by Position")
```

```
## `summarise()` regrouping output by 'Club' (override with `.groups
```



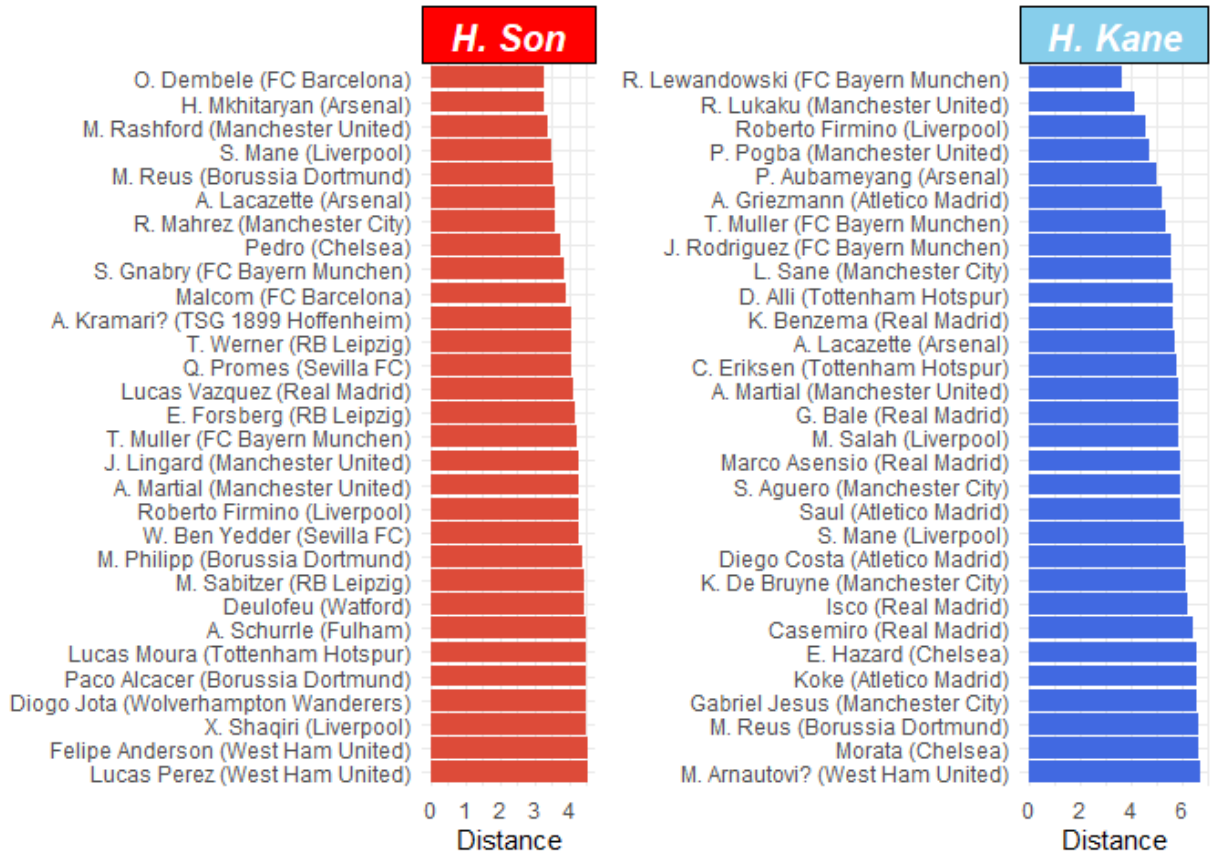
2.3 손흥민 & 케인의 유사선수 분석

유클라디안 거리를 사용한 유사도 분석을 실시하여 알아본 두 선수와 유사한 선수들

```
options(repr.plot.width = 12, repr.plot.height = 8)


similarity <- function(df, player, selectLeague, fill_variable, fill
                       input = c("Eucledian", "Maximum", "Manhattan"
                                 "Minkowski", "Pearson", "Spearman",
                       ){

  res <- NULL

  if(missing("df") | missing("input") | missing("player") | missing(
```

```
if(is.null(df) | is.null(input) | is.null(player) | is.null(fill_v
if(length(selectLeague) < 1) return(res)

if(input == "Euclidean"){
  distance_method <- "euclidean"
  }else if(input == "Maximum"){
    distance_method <- "maximum"
  }else if(input == "Manhattan"){
    distance_method <- "manhattan"
  }else if(input == "Canberra"){
    distance_method <- "canberra"
  }else if(input == "Minkowski"){
    distance_method <- "minkowski"
  }else if(input == "Pearson"){
    distance_method <- "pearson"
  }else if(input == "Spearman"){
    distance_method <- "spearman"
  }else if(input == "Kendall"){
    distance_method <- "kendall"
  }else{
    return(NULL)
  }

plyr <- df %>% filter(Name == player)

smdf <- df %>% filter(League %in% selectLeague, !Name %in% plyr$Na

rwname <- paste0(smdf$Name, " (", smdf$Club, ")")

smdf <- smdf %>% select_if(is.numeric) %>% dplyr::select(-Jersey.N

smdf <- apply(smdf, 2, scale)
rownames(smdf) <- rwname

smdf <- get_dist(smdf, method = distance_method)
smdf <- fviz_dist(smdf)

smdf <- smdf$data %>%
  mutate(Var1 = str_sub(Var1, start = 1, end = str_length(Var1)-1)
```

```
            Var2 = str_sub(Var2, start = 1, end = str_length(Var2)-1)
    arrange(value) %>%
    filter(value > 0, Var1 == rwname[length(rwname)]) %>%
    head(30)

  res <- ggplot(smdf, aes(fct_reorder(Var2, desc(value)), value))+
    geom_col(fill = fill_variable)+
    coord_flip()+
    labs(y = "Distance", x = NULL)+
    theme_minimal()+
    facet_wrap(~paste0(player))+
    theme(strip.background =element_rect(fill=fill_strip,color = "bl
          strip.text.x = element_text(size = 15, colour = "white",fa


  return(res)

}


grid.arrange(
      ncol = 2,
      similarity(df, input = "Euclidean", selectLeague = c("Bundesli
                 player = "H. Son", fill_variable = "#dd4b39", fill_
      similarity(df, input = "Euclidean", selectLeague = c("Bundesli
                 player = "H. Kane", fill_variable = "royalblue", fi
    )
```



유클라디안 거리를 계산하여 손흥민선수와 케인 선수와 유사한 선수들을 확인해 보았다

손흥민 선수는 뎀벨레,므키타리안,래시포드,마네,라카제트 등 상대 수비진을 침투하는 플레이방식의 선수들과 유사하다는 결과를 얻을 수 있었다.

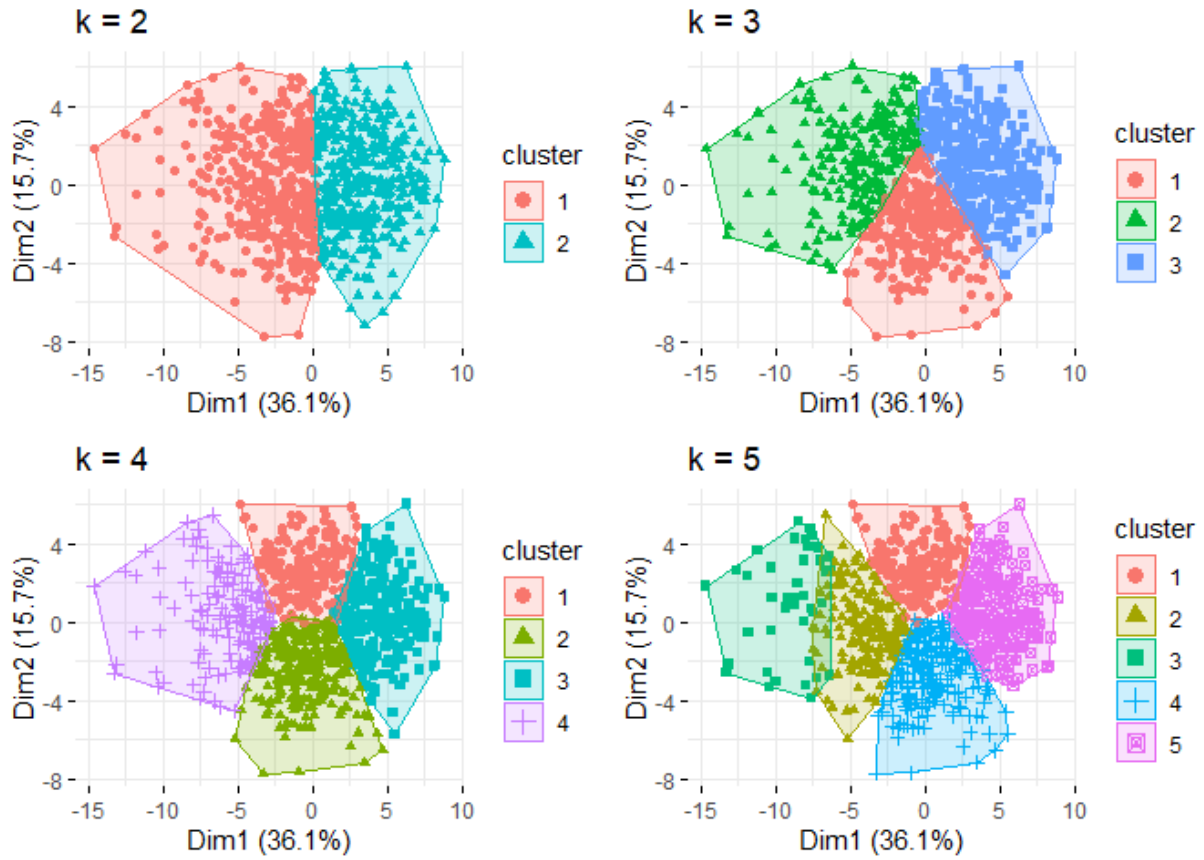
반면 해리케인 선수는 레반도프스키,루카쿠,피르미누,포그바 등 득점능력 뿐 아니라 패스 등 전체적으로 높은 밸런스를 가진 선수들과 유사하다고 판단되어 완성형 공격수라는 별명에 어울리는 결과를 얻을 수 있었다.

2.4. K-means Clustering

전방 공격수들의 특징에 따른 군집을 나누는 모델

```
km <- df %>% filter(Class == "Forward") #포지션이 forward인 class의 선
```

```
x <- km

x$ID <- 1:nrow(x)

row.names(x) <- paste(x$ID, x$Name, sep = "-")
x$Name <- NULL
x$ID <- NULL

x <- x %>% select_if(is.numeric) %>% dplyr::select(-Jersey.Number, -

x <- apply(x, 2, scale)

x <- as.data.frame(x)

k2 <- kmeans(x, center = 2, nstart = 25)   #k값이 2개인 경우
k3 <- kmeans(x, centers = 3, nstart = 25)  #k값이 3개인 경우
k4 <- kmeans(x, centers = 4, nstart = 25)  #k값이 4개인 경우
k5 <- kmeans(x, centers = 5, nstart = 25)  #k값이 5개인 경우

p1 <- fviz_cluster(k2, geom = "point", data = x, ggtheme = theme_min
p2 <- fviz_cluster(k3, geom = "point", data = x, ggtheme = theme_min
p3 <- fviz_cluster(k4, geom = "point", data = x, ggtheme = theme_min
p4 <- fviz_cluster(k5, geom = "point", data = x, ggtheme = theme_min

options(repr.plot.width = 15, repr.plot.height = 8)
grid.arrange(p1,p2,p3,p4)
```



2.5 다중회귀분석

FIFA 선수의 인구통계학적 Feature와 경기중에 관측되는 각종 stat들을 설명변수로 선수의 overall score를 얼마나 설명할 수 있을것인지 다중회귀 모델을 설계하고 개선하는 과정을 통해 규명해보고자 한다.

변수

종속변수 :

"Overall"

설명변수 :

"Age" "Value" "Wage" "Special" "International.Reputation" "Weak.Foot" "Skill.Moves" "Height" "Weight" "각종 경기내STAT"

가설

- Age는 Overall에 유의한 영향을 미칠 것이다.
- Value,Wage는 Overall에 유의한 영향을 미칠 것이다.
- Special은 Overall에 유의한 영향을 미칠 것이다.
- International Reputation은 Overall에 유의한 영향을 미칠 것이다.
- Weak.Foot은 Overall에 유의한 영향을 미칠 것이다.
- Skill.Moves는 Overall에 유의한 영향을 미칠 것이다.
- Height&Weight는 Overall에 유의한 영향을 미칠 것이다.
- 선수의 경기내 관측되는 각종 stat은 Overall에 유의한 영향을 미칠 것이다.

2.5.1 다중회귀분석을 위한 추가적인 전처리

```
#다중회귀분석을 위한 추가적인 전처리 필요
 #불필요한 범주형 특징 제거


data_original = read.csv("complete_data.csv")

#data_original[c(1:5),]

cols.dont.want <- c("X","ID","Name", "Photo", "Nationality", "Flag",
data <- data_original[, ! names(data_original) %in% cols.dont.want,
#data[c(1:5),]
```

```
 # 모든 관측값을 수치형으로 변환

data$Value <- gsub("€", "", data$Value)
data$Value <- as.numeric(gsub("M", "000", data$Value))
data$Release.Clause <- gsub("€", "", data$Release.Clause)
data$Release.Clause <- as.numeric(gsub("M", "000", data$Release.Clau
data$Wage <- gsub("€", "", data$Wage)
data$Wage <- as.numeric(gsub("K", "", data$Wage))
data$Height <- as.numeric(gsub("\'", ".", data$Height))
data$Weight <- as.numeric(gsub("lbs", "", data$Weight))

for (i in c(11:38)){
  data[,i]<- as.numeric(gsub("+", ".", data[,i],fixed=TRUE))
}

for(i in 1:ncol(data)){
  data[is.na(data[,i]), i] <- mean(data[,i], na.rm = TRUE)
}
```

```
#모델에서 가중치가 동일하게 받도록 데이터를 정규화시켜준다


# calculate the pre-process parameters from the dataset
preprocessParams <- preProcess(data, method=c("range"))
```

```
# summarize transform parameters
print(preprocessParams)
```

```
## Created from 18207 samples and 71 variables
##
## Pre-processing:
##   - ignored (0)
##   - re-scaling to [0, 1] (71)
```

```
# transform the dataset using the parameters
data_scaled <- predict(preprocessParams, data) + c(0.0000001)
# summarize the transformed dataset
for(i in 1:ncol(data))
  data_scaled[is.na(data_scaled[,i]), i] <- mean(data_scaled[,i], na
data_scaled[is.infinite(data_scaled[,i]), i] <- mean(data_scaled[,i]
#data_scaled
summary(data_scaled)
```

```
##       Age               Overall             Value
##  Min.   :0.0000001   Min.   :0.0000001   Min.   :0.0000001
##  1st Qu.:0.1724139   1st Qu.:0.3333334   1st Qu.:0.0000736
##  Median :0.3103449   Median :0.4166668   Median :0.0226927
##  Mean   :0.3145589   Mean   :0.4216397   Mean   :0.0226927
##  3rd Qu.:0.4137932   3rd Qu.:0.5208334   3rd Qu.:0.0226927
##  Max.   :1.0000001   Max.   :1.0000001   Max.   :1.0000001
##       Wage              Special          International.Reputation
##  Min.   :0.0000001   Min.   :0.0000001   Min.   :0.0000001
##  1st Qu.:0.0017700   1st Qu.:0.4495357   1st Qu.:0.0000001
##  Median :0.0053098   Median :0.5597524   Median :0.0000001
##  Mean   :0.0172237   Mean   :0.5367245   Mean   :0.0283056
##  3rd Qu.:0.0159293   3rd Qu.:0.6538701   3rd Qu.:0.0000001
##  Max.   :1.0000001   Max.   :1.0000001   Max.   :1.0000001
##    Weak.Foot          Skill.Moves            Height
##  Min.   :0.0000001   Min.   :0.0000001   Min.   :0.0000001
##  1st Qu.:0.5000001   1st Qu.:0.2500001   1st Qu.:0.0055557
##  Median :0.5000001   Median :0.2500001   Median :0.4444445
##  Mean   :0.4868248   Mean   :0.3403272   Mean   :0.3874263
##  3rd Qu.:0.5000001   3rd Qu.:0.5000001   3rd Qu.:0.5555557
```

```
##  Max.   :1.0000001   Max.   :1.0000001   Max.   :1.0000001
##      Weight                LS                  ST
##  Min.   :0.0000001   Min.   :0.0000001   Min.   :0.0000001
##  1st Qu.:0.3308272   1st Qu.:0.3660567   1st Qu.:0.3660567
##  Median :0.4135339   Median :0.4462151   Median :0.4462151
##  Mean   :0.4208958   Mean   :0.4462151   Mean   :0.4462151
##  3rd Qu.:0.4962407   3rd Qu.:0.5490850   3rd Qu.:0.5490850
##  Max.   :1.0000001   Max.   :1.0000001   Max.   :1.0000001
##        RS                  LW                  LF
##  Min.   :0.0000001   Min.   :0.0000001   Min.   :0.0000001
##  1st Qu.:0.3660567   1st Qu.:0.4328359   1st Qu.:0.4090910
##  Median :0.4462151   Median :0.5080554   Median :0.4806280
##  Mean   :0.4462151   Mean   :0.5080554   Mean   :0.4806280
##  3rd Qu.:0.5490850   3rd Qu.:0.5970150   3rd Qu.:0.5757577
##  Max.   :1.0000001   Max.   :1.0000001   Max.   :1.0000001
##        CF                  RF                  RW
##  Min.   :0.0000001   Min.   :0.0000001   Min.   :0.0000001
##  1st Qu.:0.4090910   1st Qu.:0.4090910   1st Qu.:0.4328359
##  Median :0.4806280   Median :0.4806280   Median :0.5080554
##  Mean   :0.4806280   Mean   :0.4806280   Mean   :0.5080554
##  3rd Qu.:0.5757577   3rd Qu.:0.5757577   3rd Qu.:0.5970150
##  Max.   :1.0000001   Max.   :1.0000001   Max.   :1.0000001
##       LAM                 CAM                 RAM
##  Min.   :0.0000001   Min.   :0.0000001   Min.   :0.0000001
##  1st Qu.:0.4090910   1st Qu.:0.4090910   1st Qu.:0.4090910
##  Median :0.4846241   Median :0.4846241   Median :0.4846241
##  Mean   :0.4846241   Mean   :0.4846241   Mean   :0.4846241
##  3rd Qu.:0.5757577   3rd Qu.:0.5757577   3rd Qu.:0.5757577
##  Max.   :1.0000001   Max.   :1.0000001   Max.   :1.0000001
##        LM                  LCM                 CM
##  Min.   :0.0000001   Min.   :0.0000001   Min.   :0.0000001
##  1st Qu.:0.4375001   1st Qu.:0.3958693   1st Qu.:0.3958693
##  Median :0.5115120   Median :0.4857648   Median :0.4857648
##  Mean   :0.5115120   Mean   :0.4857648   Mean   :0.4857648
##  3rd Qu.:0.5937501   3rd Qu.:0.5679863   3rd Qu.:0.5679863
##  Max.   :1.0000001   Max.   :1.0000001   Max.   :1.0000001
##       RCM                 RM                  LWB
##  Min.   :0.0000001   Min.   :0.0000001   Min.   :0.0000001
##  1st Qu.:0.3958693   1st Qu.:0.4375001   1st Qu.:0.3992741
##  Median :0.4857648   Median :0.5115120   Median :0.4993708
##  Mean   :0.4857648   Mean   :0.5115120   Mean   :0.4993708
##  3rd Qu.:0.5679863   3rd Qu.:0.5937501   3rd Qu.:0.5989112
```

```
##  Max.   :1.0000001   Max.   :1.0000001   Max.   :1.0000001
##       LDM                 CDM                 RDM
##  Min.   :0.0000001   Min.   :0.0000001   Min.   :0.0000001
##  1st Qu.:0.3891710   1st Qu.:0.3891710   1st Qu.:0.3891710
##  Median :0.4877700   Median :0.4877700   Median :0.4877700
##  Mean   :0.4877700   Mean   :0.4877700   Mean   :0.4877700
##  3rd Qu.:0.5922167   3rd Qu.:0.5922167   3rd Qu.:0.5922167
##  Max.   :1.0000001   Max.   :1.0000001   Max.   :1.0000001
##       RWB                 LB                  LCB
##  Min.   :0.0000001   Min.   :0.0000001   Min.   :0.0000001
##  1st Qu.:0.3992741   1st Qu.:0.3992741   1st Qu.:0.3542674
##  Median :0.4993708   Median :0.5046977   Median :0.4930223
##  Mean   :0.4993708   Mean   :0.5046977   Mean   :0.4930223
##  3rd Qu.:0.5989112   3rd Qu.:0.6170600   3rd Qu.:0.6280194
##  Max.   :1.0000001   Max.   :1.0000001   Max.   :1.0000001
##       CB                  RCB                 RB
##  Min.   :0.0000001   Min.   :0.0000001   Min.   :0.0000001
##  1st Qu.:0.3542674   1st Qu.:0.3542674   1st Qu.:0.3992741
##  Median :0.4930223   Median :0.4930223   Median :0.5046977
##  Mean   :0.4930223   Mean   :0.4930223   Mean   :0.5046977
##  3rd Qu.:0.6280194   3rd Qu.:0.6280194   3rd Qu.:0.6170600
##  Max.   :1.0000001   Max.   :1.0000001   Max.   :1.0000001
##     Crossing           Finishing         HeadingAccuracy
##  Min.   :0.0000001   Min.   :0.0000001   Min.   :0.0000001
##  1st Qu.:0.3750001   1st Qu.:0.3010754   1st Qu.:0.4555557
##  Median :0.5568183   Median :0.5053764   Median :0.5777779
##  Mean   :0.5083431   Mean   :0.4682895   Mean   :0.5366461
##  3rd Qu.:0.6704546   3rd Qu.:0.6451614   3rd Qu.:0.6666668
##  Max.   :1.0000001   Max.   :1.0000001   Max.   :1.0000001
##   ShortPassing          Volleys           Dribbling
##  Min.   :0.0000001   Min.   :0.0000001   Min.   :0.0000001
##  1st Qu.:0.5465117   1st Qu.:0.3023257   1st Qu.:0.4838711
##  Median :0.6395350   Median :0.4651164   Median :0.6129033
##  Mean   :0.6010084   Mean   :0.4524306   Mean   :0.5523765
##  3rd Qu.:0.7093024   3rd Qu.:0.6162792   3rd Qu.:0.6881721
##  Max.   :1.0000001   Max.   :1.0000001   Max.   :1.0000001
##      Curve            FKAccuracy         LongPassing
##  Min.   :0.0000001   Min.   :0.0000001   Min.   :0.0000001
##  1st Qu.:0.3181819   1st Qu.:0.3076924   1st Qu.:0.4047620
##  Median :0.4772728   Median :0.4175825   Median :0.5595239
##  Mean   :0.4678504   Mean   :0.4380567   Mean   :0.5203803
##  3rd Qu.:0.6363637   3rd Qu.:0.5824177   3rd Qu.:0.6547620
```

```
##  Max.   :1.0000001   Max.   :1.0000001   Max.   :1.0000001
##   BallControl         Acceleration        SprintSpeed
##  Min.   :0.0000001   Min.   :0.0000001   Min.   :0.0000001
##  1st Qu.:0.5384616   1st Qu.:0.5294119   1st Qu.:0.5357144
##  Median :0.6373627   Median :0.6470589   Median :0.6547620
##  Mean   :0.5864777   Mean   :0.6189892   Mean   :0.6277021
##  3rd Qu.:0.7032968   3rd Qu.:0.7411766   3rd Qu.:0.7500001
##  Max.   :1.0000001   Max.   :1.0000001   Max.   :1.0000001
##     Agility           Reactions           Balance
##  Min.   :0.0000001   Min.   :0.0000001   Min.   :0.0000001
##  1st Qu.:0.5000001   1st Qu.:0.4666668   1st Qu.:0.5000001
##  Median :0.6341464   Median :0.5466668   Median :0.6250001
##  Mean   :0.6037026   Mean   :0.5444882   Mean   :0.5995823
##  3rd Qu.:0.7317074   3rd Qu.:0.6266668   3rd Qu.:0.7250001
##  Max.   :1.0000001   Max.   :1.0000001   Max.   :1.0000001
##    ShotPower           Jumping             Stamina
##  Min.   :0.0000001   Min.   :0.0000001   Min.   :0.0000001
##  1st Qu.:0.4623657   1st Qu.:0.5375001   1st Qu.:0.5238096
##  Median :0.6129033   Median :0.6375001   Median :0.6428572
##  Mean   :0.5748393   Mean   :0.6261180   Mean   :0.6097614
##  3rd Qu.:0.7096775   3rd Qu.:0.7250001   3rd Qu.:0.7380953
##  Max.   :1.0000001   Max.   :1.0000001   Max.   :1.0000001
##     Strength          LongShots           Aggression
##  Min.   :0.0000001   Min.   :0.0000001   Min.   :0.0000001
##  1st Qu.:0.5125001   1st Qu.:0.3296704   1st Qu.:0.3928572
##  Median :0.6125001   Median :0.5274726   Median :0.5714287
##  Mean   :0.6038997   Mean   :0.4847251   Mean   :0.5341547
##  3rd Qu.:0.7125001   3rd Qu.:0.6483517   3rd Qu.:0.6904763
##  Max.   :1.0000001   Max.   :1.0000001   Max.   :1.0000001
##  Interceptions        Positioning           Vision
##  Min.   :0.0000001   Min.   :0.0000001   Min.   :0.0000001
##  1st Qu.:0.2584271   1st Qu.:0.3978496   1st Qu.:0.4047620
##  Median :0.5505619   Median :0.5698926   Median :0.5357144
##  Mean   :0.4909920   Mean   :0.5156827   Mean   :0.5166775
##  3rd Qu.:0.6853934   3rd Qu.:0.6666668   3rd Qu.:0.6428572
##  Max.   :1.0000001   Max.   :1.0000001   Max.   :1.0000001
##    Penalties           Composure           Marking
##  Min.   :0.0000001   Min.   :0.0000001   Min.   :0.0000001
##  1st Qu.:0.3908047   1st Qu.:0.5161291   1st Qu.:0.2967034
##  Median :0.5057472   Median :0.6021506   Median :0.5494506
##  Mean   :0.5005587   Mean   :0.5983686   Mean   :0.4866113
##  3rd Qu.:0.6321840   3rd Qu.:0.6881721   3rd Qu.:0.6703298
```

```
##  Max.   :1.0000001   Max.   :1.0000001   Max.   :1.0000001
##  StandingTackle      SlidingTackle          GKDiving
##  Min.   :0.0000001   Min.   :0.0000001   Min.   :0.0000001
##  1st Qu.:0.2747254   1st Qu.:0.2386365   1st Qu.:0.0786518
##  Median :0.5824177   Median :0.5568183   Median :0.1123597
##  Mean   :0.5021741   Mean   :0.4847891   Mean   :0.1754633
##  3rd Qu.:0.7032968   3rd Qu.:0.6931819   3rd Qu.:0.1460675
##  Max.   :1.0000001   Max.   :1.0000001   Max.   :1.0000001
##    GKHandling          GKKicking          GKPositioning
##  Min.   :0.0000001   Min.   :0.0000001   Min.   :0.0000001
##  1st Qu.:0.0769232   1st Qu.:0.0777779   1st Qu.:0.0786518
##  Median :0.1098902   Median :0.1111112   Median :0.1123597
##  Mean   :0.1691385   Mean   :0.1692452   Mean   :0.1729091
##  3rd Qu.:0.1428572   3rd Qu.:0.1444445   3rd Qu.:0.1460675
##  Max.   :1.0000001   Max.   :1.0000001   Max.   :1.0000001
##    GKReflexes        Release.Clause
##  Min.   :0.0000001   Min.   :0.0000001
##  1st Qu.:0.0752689   1st Qu.:0.0000160
##  Median :0.1075270   Median :0.0052538
##  Mean   :0.1689344   Mean   :0.0052538
##  3rd Qu.:0.1397850   3rd Qu.:0.0052538
##  Max.   :1.0000001   Max.   :1.0000001
```

2.5.2 Train set과 Test set 분할

```
# 모델 학습과 평가를 위해 Train set과 Test set을 7:3으로 분할한다.



data_shuffled <- data_scaled[sample(nrow(data_scaled)),]
r1 <- as.integer(nrow(data_shuffled)*0.7) #row number at which split
data_train <- data_shuffled[1:r1,]
data_test <- data_shuffled[(r1+1):nrow(data_shuffled),]

data_shuffled[c(1:2),]
```

```
##            Age   Overall        Value       Wage   Special
## 225  0.6551725 0.7708334 5.882363e-02 0.09911514 0.3517029
## 6525 0.4137932 0.4791668 1.088431e-05 0.01238948 0.6941177
##      International.Reputation Weak.Foot Skill.Moves    Height
## 225                 0.5000001 0.5000001   0.0000001 0.6111112 0.6
## 6525                0.0000001 0.5000001   0.5000001 0.3888890 0.4
##             LS        ST        RS        LW        LF        CF
## 225  0.4462151 0.4462151 0.4462151 0.5080554 0.4806280 0.4806280
## 6525 0.5823628 0.5823628 0.5823628 0.6567165 0.6212122 0.6212122
##             RW       LAM       CAM       RAM        LM       LCM
## 225  0.5080554 0.4846241 0.4846241 0.4846241 0.5115120 0.4857648
## 6525 0.6567165 0.6212122 0.6212122 0.6212122 0.6562501 0.6196214
##            RCM        RM       LWB       LDM       CDM       RDM
## 225  0.4857648 0.5115120 0.4993708 0.4877700 0.4877700 0.4877700
## 6525 0.6196214 0.6562501 0.6170600 0.5583757 0.5583757 0.5583757
##             LB       LCB        CB       RCB        RB   Crossing
## 225  0.5046977 0.4930223 0.4930223 0.4930223 0.5046977 0.06818192
## 6525 0.5989112 0.4830919 0.4830919 0.4830919 0.5989112 0.70454555
##      HeadingAccuracy ShortPassing    Volleys  Dribbling      Curv
## 225        0.1000001    0.1395350 0.08139545 0.06451623 0.0568182
## 6525       0.4777779    0.6976745 0.54651173 0.74193558 0.7840910
##      LongPassing BallControl Acceleration SprintSpeed   Agility R
## 225    0.1547620   0.1978023    0.4470589   0.6071430 0.4390245 0
## 6525   0.7142858   0.6813188    0.8000001   0.7500001 0.6951221 0
##        Balance ShotPower   Jumping   Stamina  Strength LongShots
## 225  0.5500001 0.2365592 0.7250001 0.2619049 0.6875001 0.1318682
## 6525 0.7000001 0.7741936 0.4375001 0.6666668 0.5625001 0.7692309

##      Interceptions Positioning   Vision Penalties Composure   Mar
## 225      0.2134832   0.1182797 0.654762 0.2183909 0.6559141 0.186
## 6525     0.7303372   0.7634410 0.607143 0.6206898 0.7204302 0.373
##      StandingTackle SlidingTackle  GKDiving GKHandling GKKicking
## 225       0.2197803     0.2386365 0.8876405  0.9010990 0.9000001
## 6525      0.5714287     0.6250001 0.1123597  0.0989012 0.1000001
##      GKReflexes Release.Clause
## 225   0.8709678   5.558818e-05
## 6525  0.1075270   2.539041e-06
```

```
data_train[c(1:2),]
```

```
##            Age   Overall        Value       Wage   Special
## 225  0.6551725 0.7708334 5.882363e-02 0.09911514 0.3517029
## 6525 0.4137932 0.4791668 1.088431e-05 0.01238948 0.6941177
##      International.Reputation Weak.Foot Skill.Moves    Height
## 225                 0.5000001 0.5000001   0.0000001 0.6111112 0.6
## 6525                0.0000001 0.5000001   0.5000001 0.3888890 0.4
##             LS        ST        RS        LW        LF        CF
## 225  0.4462151 0.4462151 0.4462151 0.5080554 0.4806280 0.4806280
## 6525 0.5823628 0.5823628 0.5823628 0.6567165 0.6212122 0.6212122
##             RW       LAM       CAM       RAM        LM       LCM
## 225  0.5080554 0.4846241 0.4846241 0.4846241 0.5115120 0.4857648
## 6525 0.6567165 0.6212122 0.6212122 0.6212122 0.6562501 0.6196214
##            RCM        RM       LWB       LDM       CDM       RDM
## 225  0.4857648 0.5115120 0.4993708 0.4877700 0.4877700 0.4877700
## 6525 0.6196214 0.6562501 0.6170600 0.5583757 0.5583757 0.5583757
##             LB       LCB        CB       RCB        RB   Crossing
## 225  0.5046977 0.4930223 0.4930223 0.4930223 0.5046977 0.06818192
## 6525 0.5989112 0.4830919 0.4830919 0.4830919 0.5989112 0.70454555
##      HeadingAccuracy ShortPassing    Volleys  Dribbling      Curv
## 225        0.1000001    0.1395350 0.08139545 0.06451623 0.0568182
## 6525       0.4777779    0.6976745 0.54651173 0.74193558 0.7840910
##      LongPassing BallControl Acceleration SprintSpeed   Agility R
## 225    0.1547620   0.1978023    0.4470589   0.6071430 0.4390245 0
## 6525   0.7142858   0.6813188    0.8000001   0.7500001 0.6951221 0
##        Balance ShotPower   Jumping   Stamina  Strength LongShots
## 225  0.5500001 0.2365592 0.7250001 0.2619049 0.6875001 0.1318682
## 6525 0.7000001 0.7741936 0.4375001 0.6666668 0.5625001 0.7692309
##      Interceptions Positioning   Vision Penalties Composure   Mar
## 225      0.2134832   0.1182797 0.654762 0.2183909 0.6559141 0.186
## 6525     0.7303372   0.7634410 0.607143 0.6206898 0.7204302 0.373
##      StandingTackle SlidingTackle  GKDiving GKHandling GKKicking
## 225       0.2197803     0.2386365 0.8876405  0.9010990 0.9000001
## 6525      0.5714287     0.6250001 0.1123597  0.0989012 0.1000001
##      GKReflexes Release.Clause
## 225   0.8709678   5.558818e-05
## 6525  0.1075270   2.539041e-06
```

```
data_test[c(1:2),]
```

```
##             Age   Overall     Value      Wage   Special
## 14288 0.2413794 0.3125001 0.02269274 0.003539923 0.1962849
## 15642 0.2068967 0.2708334 0.02269274 0.014159392 0.1851394
##       International.Reputation Weak.Foot Skill.Moves    Height
## 14288                    1e-07 0.2500001       1e-07 0.7222223 0.
## 15642                    1e-07 0.2500001       1e-07 0.6666668 0.
##             LS        ST        RS        LW       LF       CF
## 14288 0.4462151 0.4462151 0.4462151 0.5080554 0.480628 0.480628 0
## 15642 0.4462151 0.4462151 0.4462151 0.5080554 0.480628 0.480628 0
##             RW       LAM       CAM       RAM       LM       LCM
## 14288 0.5080554 0.4846241 0.4846241 0.4846241 0.511512 0.4857648
## 15642 0.5080554 0.4846241 0.4846241 0.4846241 0.511512 0.4857648
##             RCM       RM       LWB     LDM     CDM     RDM
## 14288 0.4857648 0.511512 0.4993708 0.48777 0.48777 0.48777 0.4993
## 15642 0.4857648 0.511512 0.4993708 0.48777 0.48777 0.48777 0.4993
##             LCB        CB       RCB        RB   Crossing Finishin
## 14288 0.4930223 0.4930223 0.4930223 0.5046977 0.09090919 0.172043
## 15642 0.4930223 0.4930223 0.4930223 0.5046977 0.07954555 0.129032
##       HeadingAccuracy ShortPassing   Volleys Dribbling      Curve
## 14288       0.1111112    0.2441861 0.1744187 0.1397850 0.10227283
## 15642       0.1222223    0.2209303 0.1046513 0.1290324 0.07954555
##       LongPassing BallControl Acceleration SprintSpeed   Agility
## 14288   0.1309525   0.1538463    0.3294119   0.3571430 0.2073172
## 15642   0.2380953   0.1538463    0.4352942   0.3452382 0.2560977
##         Balance ShotPower   Jumping   Stamina  Strength  LongShot
## 14288 0.3250001 0.1935485 0.4750001 0.2023811 0.7625001 0.1208792
## 15642 0.5125001 0.2258066 0.5500001 0.3452382 0.4250001 0.0879121
##       Interceptions Positioning    Vision  Penalties Composure
## 14288     0.2471911   0.1827958 0.2500001 0.18390815 0.2688173 0.

## 15642     0.2471911   0.1182797 0.3095239 0.09195412 0.4516130 0.
##       StandingTackle SlidingTackle  GKDiving GKHandling GKKicking
## 14288      0.1868133     0.1022728 0.6966293  0.6043957 0.5777779
## 15642      0.1318682     0.1818183 0.6741574  0.5934067 0.5888890
##       GKReflexes Release.Clause
## 14288  0.6666668    0.005253827
## 15642  0.6559141    0.005253827
```

```
not_training_features <- c("Overall")
x_train <- as.matrix(data_train[, ! names(data_train) %in% not_train
y_train <- as.matrix(data_train[,"Overall", drop = F])
```

```
x_test <- as.matrix(data_test[, ! names(data_test) %in% not_training
y_test <- as.matrix(data_test[,"Overall", drop = F])

print("Dimensions :")
```

```
## [1] "Dimensions :"
```

```
print("x_train : ")
```

```
## [1] "x_train : "
```

```
print(dim(x_train))
```

```
## [1] 12744    70
```

```
print("x_test : ")
```

```
## [1] "x_test : "
```

```
print(dim(x_test))
```

```
## [1] 5463   70
```

```
print("y_train : ")
```

```
## [1] "y_train : "
```

```
print(dim(y_train))
```

```
## [1] 12744     1
```

```
print("y_test : ")
```

```
## [1] "y_test : "
```

```
print(dim(y_test))
```

```
## [1] 5463    1
```

2.5.1 다중회귀선형모델 첫 Fitting

```
#다중선형회귀모델을 Fitting한다.

lin_model1<- Overall~Age+Value+Wage+Special+International.Reputation

fit1<-lm(lin_model1,data_train)
summary(fit1)
```

```
##
## Call:
## lm(formula = lin_model1, data = data_train)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.285555 -0.031844  0.000765  0.033002  0.198360
##
## Coefficients: (16 not defined because of singularities)
##                            Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)              -0.343128   0.031778 -10.798  < 2e-16 **
## Age                       0.038337   0.003735  10.266  < 2e-16 **
## Value                    -0.074122   0.013351  -5.552 2.88e-08 **
## Wage                      0.409815   0.017787  23.040  < 2e-16 **
## Special                   0.387188   0.093414   4.145 3.42e-05 **
## International.Reputation  0.047591   0.006422   7.411 1.33e-13 **
## Weak.Foot                 0.004786   0.002935   1.631 0.103020
## Skill.Moves               0.118945   0.004733  25.131  < 2e-16 **
## Height                    0.015126   0.002143   7.058 1.78e-12 **
## Weight                    0.029057   0.006317   4.600 4.27e-06 **
## LS                        0.130226   0.059640   2.184 0.029016 *
## ST                              NA         NA      NA       NA
## RS                              NA         NA      NA       NA
## LW                        0.090495   0.090023   1.005 0.314800
## LF                       -0.273815   0.088417  -3.097 0.001960 **
## CF                              NA         NA      NA       NA
## RF                              NA         NA      NA       NA
## RW                              NA         NA      NA       NA
## LAM                      -0.222606   0.085120  -2.615 0.008928 **
## CAM                             NA         NA      NA       NA
## RAM                             NA         NA      NA       NA
## LM                        0.166823   0.087318   1.911 0.056087 .
## LCM                      -0.127102   0.065327  -1.946 0.051721 .
## CM                              NA         NA      NA       NA
## RCM                             NA         NA      NA       NA
## RM                              NA         NA      NA       NA
## LWB                       0.091453   0.070928   1.289 0.197293
## LDM                       0.386897   0.071778   5.390 7.16e-08 **
## CDM                             NA         NA      NA       NA
## RDM                             NA         NA      NA       NA
## RWB                             NA         NA      NA       NA
## LB                        0.015017   0.068851   0.218 0.827350
## LCB                      -0.299031   0.056771  -5.267 1.41e-07 **
## CB                              NA         NA      NA       NA
## RCB                             NA         NA      NA       NA
## RB                              NA         NA      NA       NA
## Crossing                 -0.065420   0.014224  -4.599 4.28e-06 **
## Finishing                 0.036184   0.014560   2.485 0.012963 *
## HeadingAccuracy           0.167247   0.012356  13.536  < 2e-16 **
## ShortPassing              0.132579   0.011695  11.336  < 2e-16 **
## Volleys                  -0.041118   0.007703  -5.338 9.55e-08 **
## Dribbling                 0.010128   0.015933   0.636 0.525033
```

```
## Curve                -0.007660   0.007642  -1.002 0.316183
## FKAccuracy           -0.018693   0.007405  -2.525 0.011596 *
## LongPassing          -0.072783   0.011394  -6.388 1.74e-10 **
## BallControl           0.222560   0.012677  17.556  < 2e-16 **
## Acceleration          0.029447   0.010060   2.927 0.003427 **
## SprintSpeed           0.022052   0.009340   2.361 0.018246 *
## Agility              -0.013149   0.007978  -1.648 0.099355 .
## Reactions             0.370212   0.008455  43.789  < 2e-16 **
## Balance              -0.049659   0.006787  -7.317 2.69e-13 **
## ShotPower             0.017012   0.011443   1.487 0.137118
## Jumping              -0.001661   0.006352  -0.262 0.793656
## Stamina              -0.033313   0.010397  -3.204 0.001358 **
## Strength              0.027074   0.008716   3.106 0.001899 **
## LongShots            -0.027424   0.009943  -2.758 0.005822 **
## Aggression           -0.018210   0.008102  -2.248 0.024614 *
## Interceptions        -0.068016   0.012747  -5.336 9.66e-08 **
## Positioning          -0.073190   0.012264  -5.968 2.47e-09 **
## Vision               -0.020129   0.009077  -2.217 0.026609 *
## Penalties            -0.021999   0.007414  -2.967 0.003009 **
## Composure             0.173944   0.006737  25.819  < 2e-16 **
## Marking               0.036942   0.010613   3.481 0.000502 **
## StandingTackle        0.015104   0.014950   1.010 0.312363
## SlidingTackle        -0.045329   0.014159  -3.201 0.001371 **
## GKDiving              0.099761   0.013231   7.540 5.02e-14 **
## GKHandling            0.103507   0.013665   7.574 3.86e-14 **
## GKKicking             0.036626   0.012725   2.878 0.004005 **
## GKPositioning         0.111748   0.012975   8.612  < 2e-16 **
## GKReflexes            0.119826   0.013670   8.766  < 2e-16 **
## Release.Clause        0.039520   0.022258   1.776 0.075830 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05032 on 12689 degrees of freedom
## Multiple R-squared:  0.8767, Adjusted R-squared:  0.8761
## F-statistic:  1670 on 54 and 12689 DF,  p-value: < 2.2e-16
```

```
print("Number of Parameters Learnt = ")
```

```
## [1] "Number of Parameters Learnt = "
```

```
print(length(fit1$coefficients))
```

```
## [1] 71
```

***p-value*가 *0.05*보다 작으므로 ,귀무가설이 기각되고 따라서 회귀성이 충족됨을 알 수 있고 이를 모델에 포함시키고자한다.**

다중공선성이 존재함을 확인

상관 행렬을 사용하여 공선형 회귀 변수를 찾는다.

2.5.2 상관행렬을 이용한 공선형 회귀변수 확인

```
corr_x_train <- cor(x_train)
dim(corr_x_train)
```

```
## [1] 70 70
```

```
name_col_rem <- findCorrelation(corr_x_train, cutoff = 0.95, verbose
print("Regressors to Remove: ")
```

```
## [1] "Regressors to Remove: "
```

```
name_col_rem
```

```
##  [1] "LCM"           "CM"            "LAM"           "CAM"
##  [5] "RAM"           "LM"            "RM"            "LW"
##  [9] "RW"            "LF"            "CF"            "LS"
## [13] "ST"            "RS"            "RWB"           "SlidingTack
```

```
## [17] "LDM"          "CDM"          "RDM"          "LB"
## [21] "RB"           "GKDiving"     "GKHandling"   "GKReflexes"
## [25] "GKKicking"    "CB"           "RCB"
```

2.5.3 다른 회귀자와> = 0.95의 상관 관계를 가지고 있기 때문에 제거해야하는 Regressor를 확인 및 이를 제거

```
cols.dont.want <- name_col_rem
data_train <- data_train[, ! names(data_train) %in% cols.dont.want,
data_test <- data_test[, ! names(data_test) %in% cols.dont.want, dro
dim(data_train)
```

```
## [1] 12744    44
```

```
dim(data_test)
```

```
## [1] 5463   44
```

```
not_training_features <- c("Overall")
x_train <- as.matrix(data_train[, ! names(data_train) %in% not_train
y_train <- as.matrix(data_train[,"Overall", drop = F])
x_test <- as.matrix(data_test[, ! names(data_test) %in% not_training
y_test <- as.matrix(data_test[,"Overall", drop = F])
#Regressor의 숫자가 44개로 줄어들었음을 확인할 수 있다.
```

2.5.4 선형모델 재적합

```
new_colnames <- colnames(x_train)
print("New Column Names")
```

```
## [1] "New Column Names"
```

```
new_colnames
```

```
##  [1] "Age"                      "Value"
##  [3] "Wage"                     "Special"
##  [5] "International.Reputation" "Weak.Foot"
##  [7] "Skill.Moves"              "Height"
##  [9] "Weight"                   "RF"
## [11] "RCM"                      "LWB"
## [13] "LCB"                      "Crossing"
## [15] "Finishing"                "HeadingAccuracy"
## [17] "ShortPassing"             "Volleys"
## [19] "Dribbling"                "Curve"
## [21] "FKAccuracy"               "LongPassing"
## [23] "BallControl"              "Acceleration"
## [25] "SprintSpeed"              "Agility"
## [27] "Reactions"                "Balance"
## [29] "ShotPower"                "Jumping"
## [31] "Stamina"                  "Strength"
## [33] "LongShots"                "Aggression"
## [35] "Interceptions"            "Positioning"
## [37] "Vision"                   "Penalties"
## [39] "Composure"                "Marking"
## [41] "StandingTackle"           "GKPositioning"
## [43] "Release.Clause"
```

```
model2_formula <- Overall~Age+Value+Wage+Special+International.Reput

lin_model2<-model2_formula
fit2<-lm(lin_model2,data_train)
summary(fit2)
```

```
##
## Call:
## lm(formula = lin_model2, data = data_train)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.301249 -0.032216  0.000604  0.033521  0.289856
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               -0.045271   0.016688  -2.713  0.00668 **
## Age                        0.040020   0.003785  10.572  < 2e-16 **
## Value                     -0.073563   0.013545  -5.431 5.70e-08 **
## Wage                       0.413992   0.018043  22.944  < 2e-16 **
## Special                    1.285366   0.056090  22.916  < 2e-16 **
## International.Reputation   0.047869   0.006517   7.345 2.19e-13 **
## Weak.Foot                  0.005120   0.002978   1.719  0.08560 .
## Skill.Moves                0.114960   0.004792  23.992  < 2e-16 **
## Height                     0.015637   0.002175   7.189 6.88e-13 **
## Weight                     0.032811   0.006405   5.122 3.06e-07 **
## RCM                        0.065054   0.026367   2.467  0.01363 *
## LWB                        0.069668   0.040485   1.721  0.08530 .
## LCB                       -0.101317   0.042927  -2.360  0.01828 *
## Crossing                  -0.078703   0.009846  -7.994 1.42e-15 **
## Finishing                 -0.020836   0.007955  -2.619  0.00883 **
## HeadingAccuracy            0.090816   0.008674  10.469  < 2e-16 **
## ShortPassing               0.069679   0.010278   6.779 1.26e-11 **
## Volleys                   -0.086557   0.006417 -13.489  < 2e-16 **
## Dribbling                 -0.093233   0.010196  -9.144  < 2e-16 **
## Curve                     -0.058275   0.006526  -8.930  < 2e-16 **
## FKAccuracy                -0.071163   0.006155 -11.561  < 2e-16 **
## LongPassing               -0.108008   0.008972 -12.038  < 2e-16 **
## BallControl                0.144636   0.011487  12.591  < 2e-16 **
## Acceleration              -0.013309   0.008642  -1.540  0.12359
## SprintSpeed               -0.024154   0.007947  -3.040  0.00237 **
## Agility                   -0.058664   0.006334  -9.262  < 2e-16 **
## Reactions                  0.341274   0.007708  44.277  < 2e-16 **
## Balance                   -0.097858   0.005759 -16.992  < 2e-16 **
## ShotPower                 -0.035120   0.007034  -4.993 6.04e-07 **
## Jumping                   -0.051278   0.005017 -10.221  < 2e-16 **
## Stamina                   -0.051299   0.007105  -7.220 5.48e-13 **
## Strength                  -0.009970   0.007325  -1.361  0.17351
```

```
## LongShots                -0.108121   0.007694 -14.052  < 2e-16 **
## Aggression               -0.058856   0.006259  -9.404  < 2e-16 **
## Interceptions            -0.091095   0.007748 -11.757  < 2e-16 **
## Positioning              -0.169971   0.007985 -21.287  < 2e-16 **
## Vision                   -0.103632   0.007586 -13.661  < 2e-16 **
## Penalties                -0.069204   0.006365 -10.872  < 2e-16 **
## Composure                 0.177975   0.006825  26.078  < 2e-16 **
## Marking                  -0.004702   0.007966  -0.590  0.55499
## StandingTackle           -0.091654   0.011294  -8.115 5.29e-16 **
## GKPositioning             0.155367   0.012802  12.136  < 2e-16 **
## Release.Clause            0.039519   0.022593   1.749  0.08030 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05109 on 12701 degrees of freedom
## Multiple R-squared:  0.8727, Adjusted R-squared:  0.8723
## F-statistic:  2074 on 42 and 12701 DF,  p-value: < 2.2e-16
```

```
print("Number of Parameters Learnt = ")
```

```
## [1] "Number of Parameters Learnt = "
```

```
print(length(fit2$coefficients))
```

```
## [1] 43
```

2.5.5 개별 가설 검정 :

이제 p-value <0.05 인 매개 변수만 고려하고 다른 매개 변수는 중요하지 않기 때문에 다음 선형 모델을 선택한 회귀 변수에 맞춰준다.

```
#p-value <0.05를 기준으로 모든 회귀 변수의 포함 / 제외를 위해 참 / 거짓
toselect.x <- summary(fit2)$coeff[-1,4] < 0.05
#p-value가 0.05보다 작은 regressors를 저장
relevant.x <- names(toselect.x)[toselect.x == TRUE]



#유의한 변수들만 남은 선형회귀
sig.formula <- as.formula(paste("Overall ~",paste(relevant.x, collap
print("New Formula ")
```

```
## [1] "New Formula "
```

```
print(sig.formula)
```

```
## Overall ~ Age + Value + Wage + Special + International.Reputation
##     Skill.Moves + Height + Weight + RCM + LCB + Crossing + Finish
##     HeadingAccuracy + ShortPassing + Volleys + Dribbling + Curve
##     FKAccuracy + LongPassing + BallControl + SprintSpeed + Agilit
##     Reactions + Balance + ShotPower + Jumping + Stamina + LongSho
##     Aggression + Interceptions + Positioning + Vision + Penalties
##     Composure + StandingTackle + GKPositioning
```

```
lin_model3 <- sig.formula
fit3<-lm(sig.formula,data_train)
summary(fit3)
```

```
##
## Call:
## lm(formula = sig.formula, data = data_train)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.303756 -0.032115  0.000599  0.033651  0.274724
##
## Coefficients:
```

```
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               -0.061577   0.014808  -4.159 3.22e-05 **

## Age                        0.038905   0.003737  10.410  < 2e-16 **
## Value                     -0.070880   0.013410  -5.285 1.27e-07 **
## Wage                       0.415470   0.018043  23.027  < 2e-16 **
## Special                    1.232565   0.049161  25.072  < 2e-16 **
## International.Reputation   0.049353   0.006499   7.594 3.32e-14 **
## Skill.Moves                0.114459   0.004771  23.991  < 2e-16 **
## Height                     0.015460   0.002171   7.122 1.12e-12 **
## Weight                     0.026868   0.005898   4.556 5.27e-06 **
## RCM                        0.111363   0.017127   6.502 8.21e-11 **
## LCB                       -0.067558   0.021622  -3.125  0.00179 **
## Crossing                  -0.061515   0.006100 -10.084  < 2e-16 **
## Finishing                 -0.020393   0.007774  -2.623  0.00872 **
## HeadingAccuracy            0.086423   0.006461  13.376  < 2e-16 **
## ShortPassing               0.069459   0.010071   6.897 5.58e-12 **
## Volleys                   -0.083690   0.006257 -13.376  < 2e-16 **
## Dribbling                 -0.090559   0.010028  -9.031  < 2e-16 **
## Curve                     -0.055116   0.006348  -8.683  < 2e-16 **
## FKAccuracy                -0.068086   0.005939 -11.464  < 2e-16 **
## LongPassing               -0.113029   0.008193 -13.796  < 2e-16 **
## BallControl                0.143611   0.011245  12.771  < 2e-16 **
## SprintSpeed               -0.021269   0.006603  -3.221  0.00128 **
## Agility                   -0.055708   0.006233  -8.938  < 2e-16 **
## Reactions                  0.345866   0.007568  45.699  < 2e-16 **
## Balance                   -0.093824   0.005520 -16.997  < 2e-16 **
## ShotPower                 -0.032583   0.006854  -4.754 2.01e-06 **
## Jumping                   -0.050282   0.004508 -11.154  < 2e-16 **
## Stamina                   -0.044733   0.005606  -7.980 1.59e-15 **
## LongShots                 -0.108987   0.007439 -14.650  < 2e-16 **
## Aggression                -0.060359   0.004910 -12.292  < 2e-16 **
## Interceptions             -0.084842   0.007343 -11.555  < 2e-16 **
## Positioning               -0.172855   0.007654 -22.585  < 2e-16 **
## Vision                    -0.105717   0.006964 -15.181  < 2e-16 **
## Penalties                 -0.066690   0.006184 -10.785  < 2e-16 **
## Composure                  0.177374   0.006819  26.012  < 2e-16 **
## StandingTackle            -0.083705   0.009648  -8.676  < 2e-16 **
## GKPositioning              0.165692   0.011767  14.081  < 2e-16 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05111 on 12707 degrees of freedom
```

```
## Multiple R-squared:  0.8726, Adjusted R-squared:  0.8722
## F-statistic:  2417 on 36 and 12707 DF,  p-value: < 2.2e-16
```

```
print("Number of Parameters Learnt = ")
```

```
## [1] "Number of Parameters Learnt = "
```

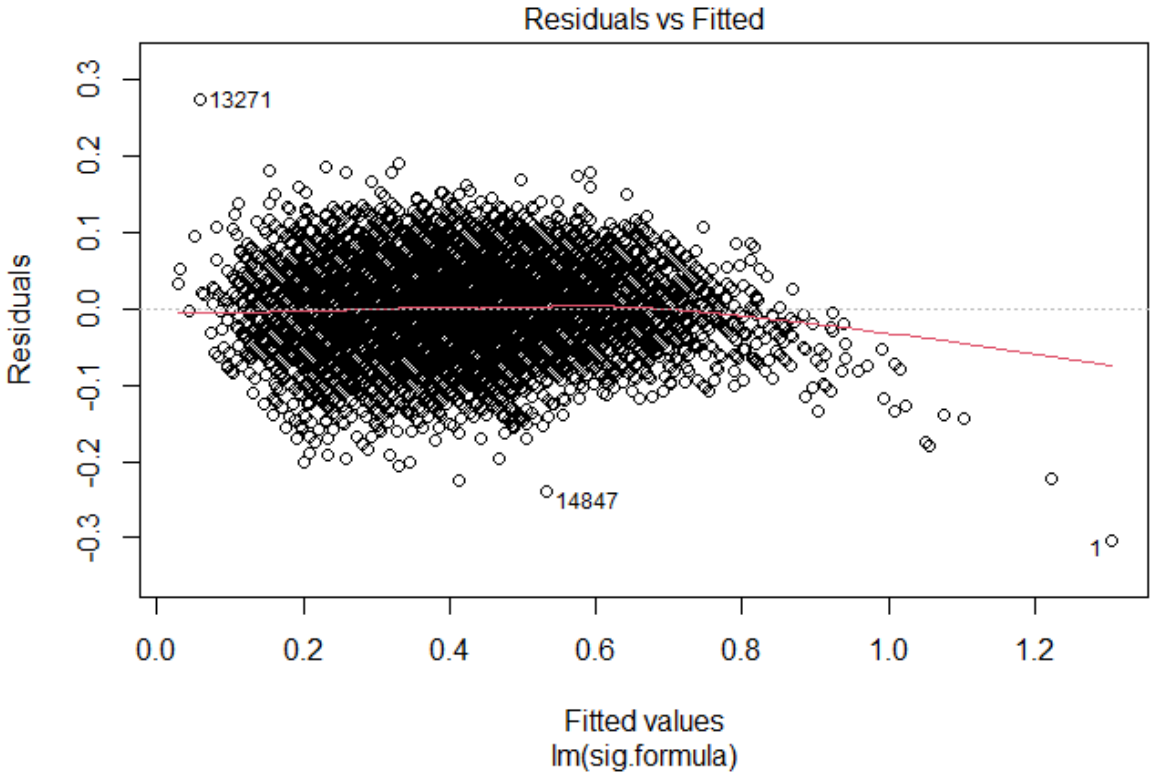
```
print(length(fit3$coefficients))
```

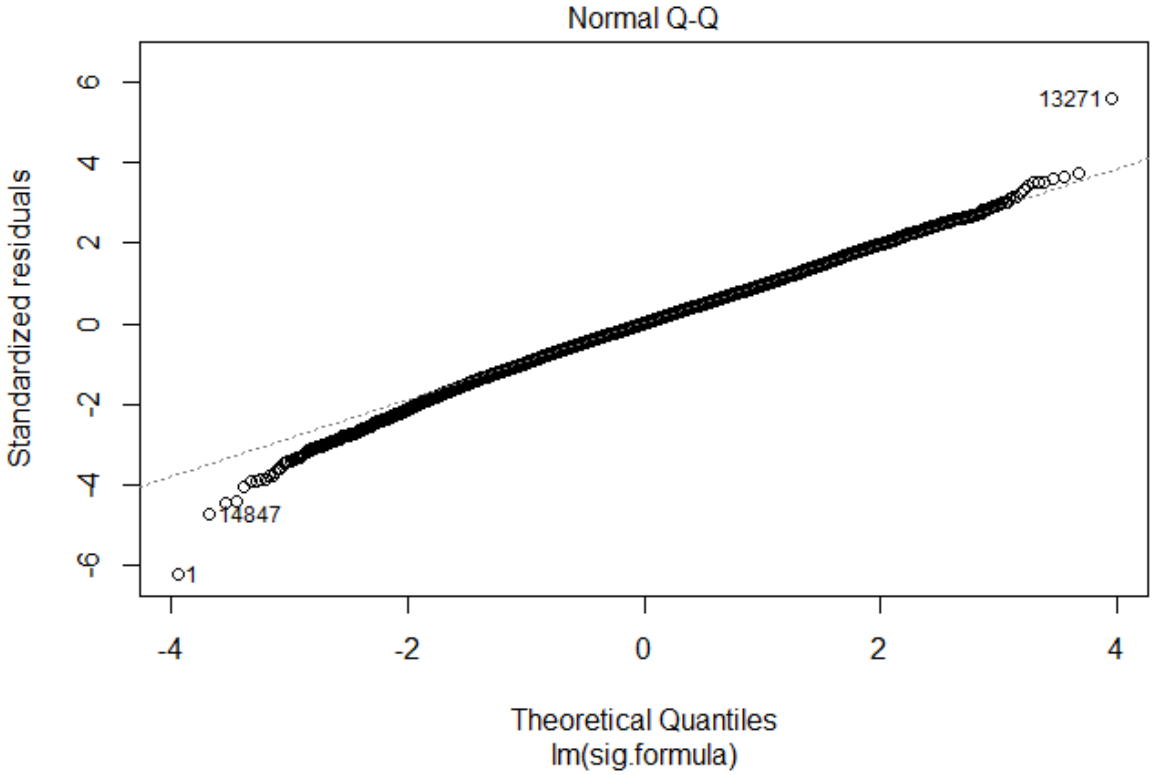
```
## [1] 37
```

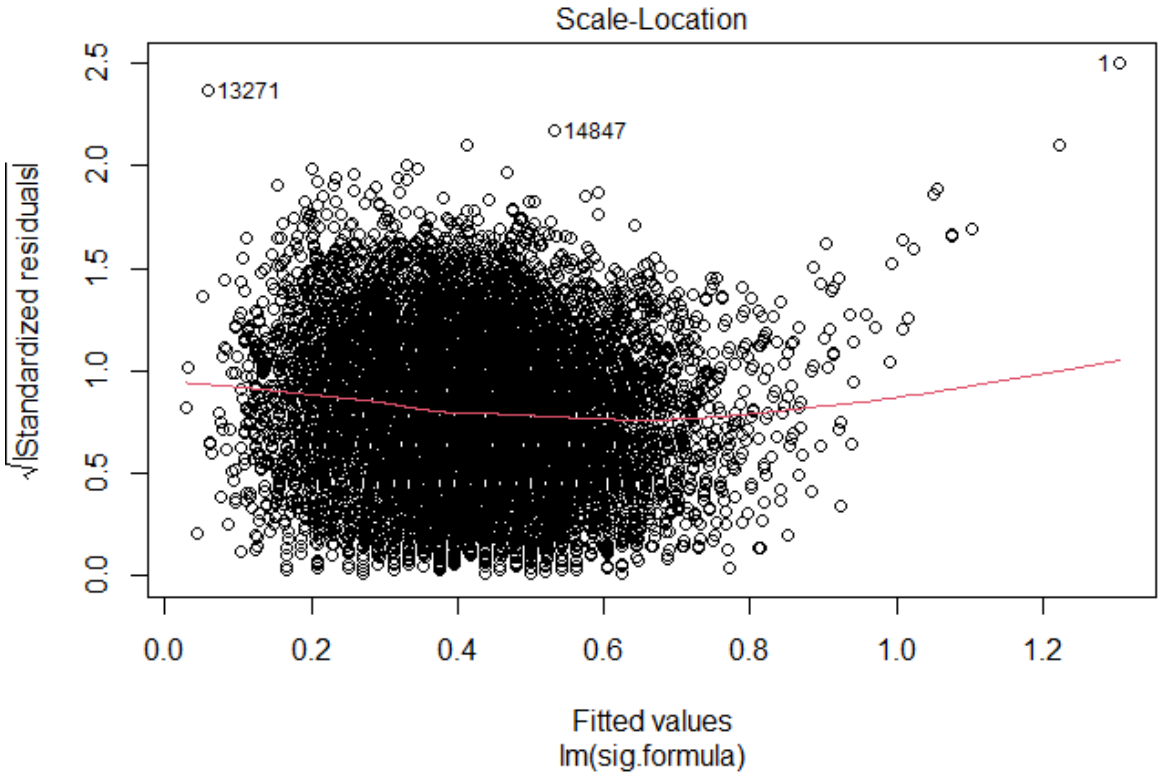
중간결과 *:* 학습 된 매개 변수 수를 초기 값 *71*에서 *36*으로 줄 였지만 *adjusted R^2*는 거의 동일하다*.*

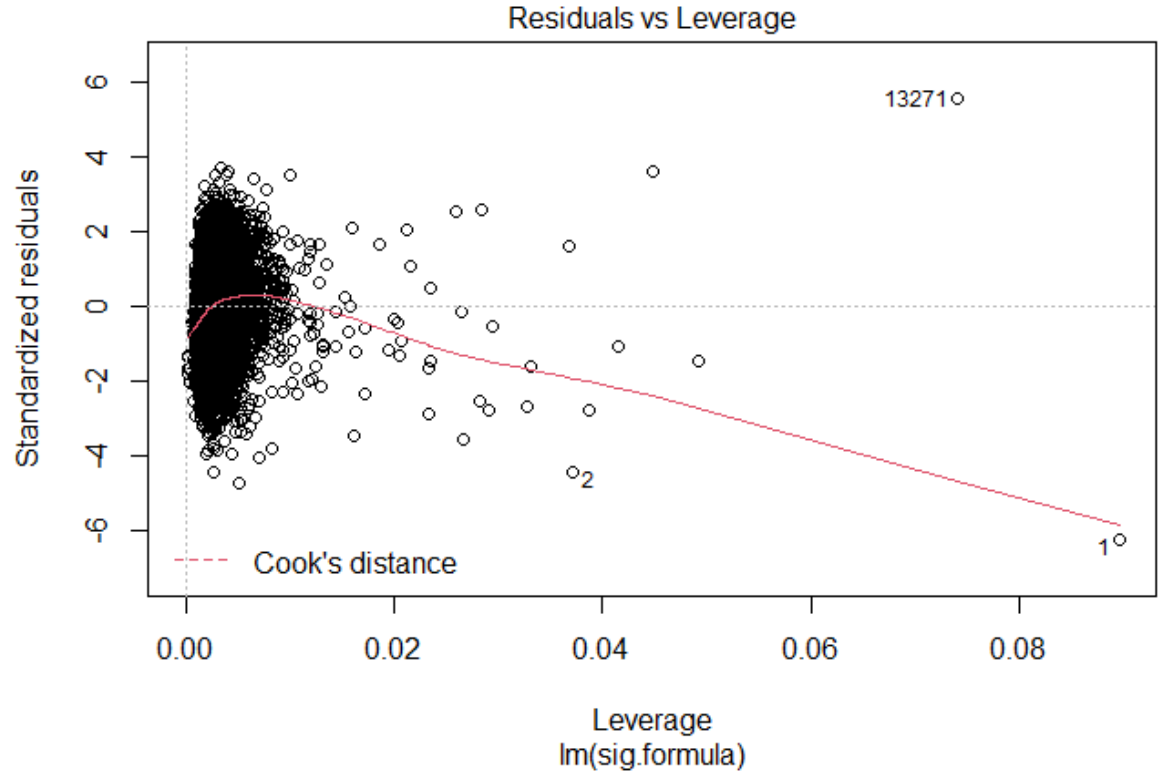
2.5.6 QQ Plot등 잔차확인

```
# 잔차확인 -



plot(fit3)
```


Residuals vs Fitted
Residuals
Fitted values
lm(sig.formula)
13271
14847
1
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
0.0
0.2
0.4
0.6
0.8
1.0
1.2

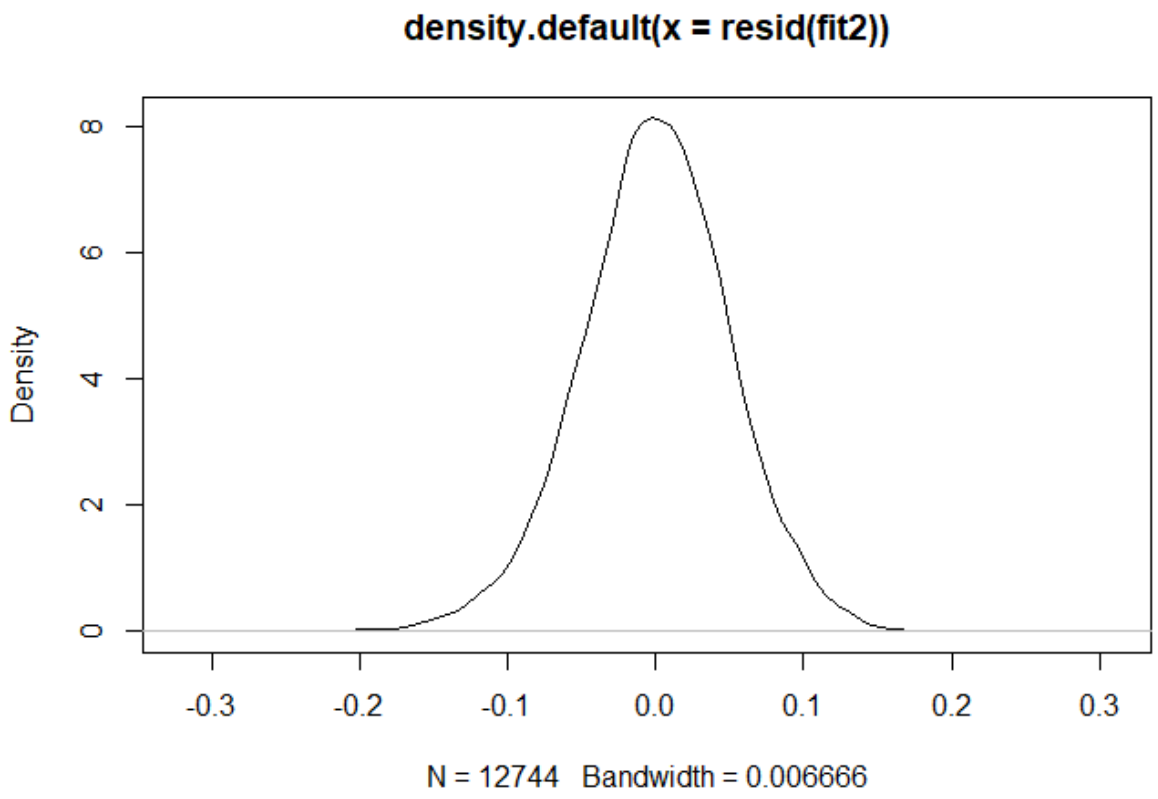

Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(sig.formula)
13271
14847
1
6
4
2
0
-2
-4
-6
-4
-2
0
2
4






```
#
# Observations -
```

```
#
#   Residuals vs Fitted : Most Residuals are equally spread along th
# Normal QQ Plot : Data is normally distributed except few outliers.
# Scale Location Plot: Data is evenly spread except for outliers whe
# Residuals vs Leverage : Cook's Distance Lines are not seen in plot



# Density Plot of Residuals looks like it has a Normal Distribution


plot(density(resid(fit2)))
```



2.5.7 Shapiro-Wilk Normalcy Test

```
#Applying Shapiro-Wilk's Normalcy Test : H_0 : Sample is Normal, H_a
```

```
shapiro.test(resid(fit2)[1:5000])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(fit2)[1:5000]
## W = 0.99811, p-value = 9.219e-06
```

```
shapiro.test(resid(fit2)[5001:10000])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(fit2)[5001:10000]
## W = 0.99607, p-value = 2.691e-10
```

```
shapiro.test(resid(fit2)[10001:length(resid(fit2))])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(fit2)[10001:length(resid(fit2))]
## W = 0.99482, p-value = 3.157e-08
```

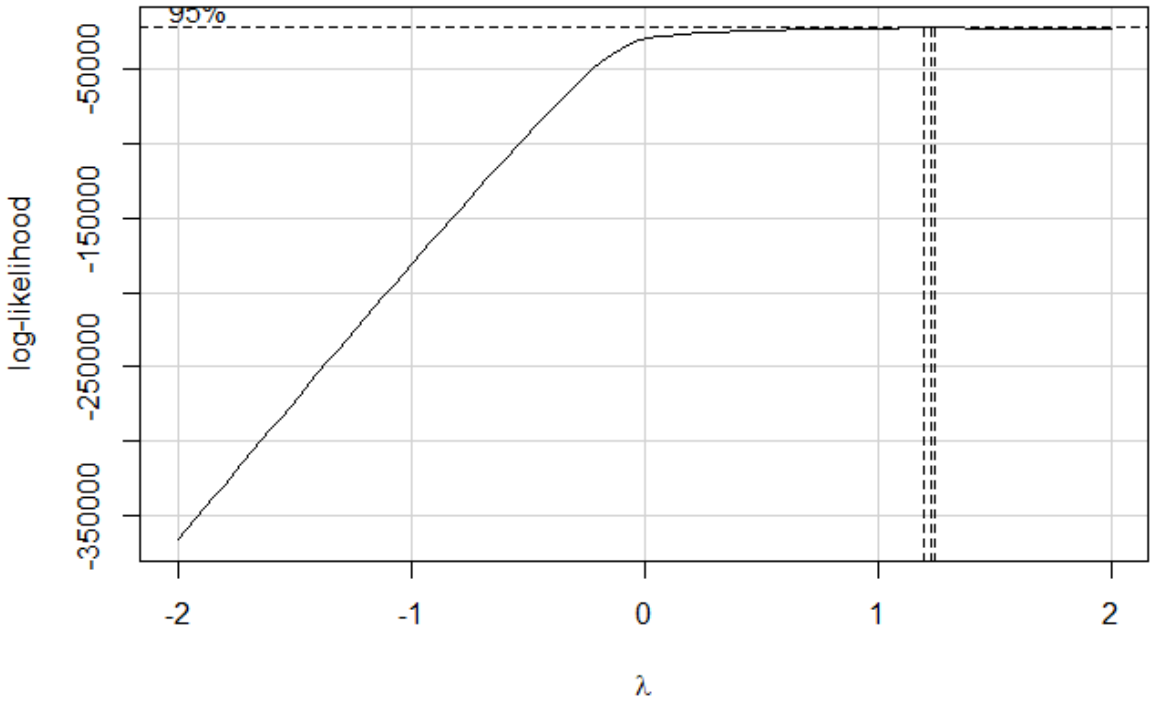
```
x <- replicate(1000, { #각 분포에 1000개의 다른 test를 만들어준다
  c(shapiro.test(rnorm(1000)+c(0,0,1,0,0))$p.value, #$
    shapiro.test(rnorm(5000)+c(0,0,1,0,0))$p.value) #$
} # rnorm은 정규분포로부터 random한 draw를 준다.
)
rownames(x) <- c("n1000","n5000")
rowMeans(x<0.05)
```

```
## n1000 n5000
## 0.132 0.459
```

잔차는 정규적으로 분포함을 확인할 수 있다

###2.5.8 Box-Cox Transformation

```
# BOX-COX TRANSFORMATION을 사용한다.



model4_formula <- as.formula(paste("data_train$Overall ~",paste("dat
lambda<-boxCox(model4_formula,objective.name="Log-Likelihood",plotit
```




```
ind<-which(lambda$y == max(lambda$y))
lambda.max<-lambda$x[ind]
data_train$Overall<-bcPower(data_train$Overall,lambda = lambda.max)
data_test$Overall<-bcPower(data_test$Overall,lambda = lambda.max)
not_training_features <- c("Overall")
x_train <- as.matrix(data_train[, ! names(data_train) %in% not_train
y_train <- as.matrix(data_train[,"Overall", drop = F])
x_test <- as.matrix(data_test[, ! names(data_test) %in% not_training
y_test <- as.matrix(data_test[,"Overall", drop = F])

print("Applied BoxCox. Fitting New Model")
```

```
## [1] "Applied BoxCox. Fitting New Model"
```

```
print(data_train$Overall[c(1,3)])
```

```
## [1] -0.2226667 -0.5184911
```

```
print(data_test$Overall[c(1,3)])
```

```
## [1] -0.6179369 -0.6179369
```

```
model4_formula <- as.formula(paste("Overall~",paste(relevant.x, coll
model4 <-model4_formula
fit4<-lm(model4,data_train)
summary(fit4)
```

```
##
## Call:
## lm(formula = model4, data = data_train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.27131 -0.02601 -0.00020  0.02669  0.20890
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               -0.909600   0.011796 -77.113  < 2e-16 **
## Age                        0.023268   0.002977   7.815 5.90e-15 **

## Value                     -0.041316   0.010683  -3.868  0.00011 **
## Wage                       0.406294   0.014373  28.268  < 2e-16 **
## Special                    0.965508   0.039162  24.654  < 2e-16 **
## International.Reputation   0.063416   0.005177  12.249  < 2e-16 **
## Skill.Moves                0.094076   0.003801  24.753  < 2e-16 **
## Height                     0.011595   0.001729   6.705 2.09e-11 **
## Weight                     0.019403   0.004698   4.130 3.65e-05 **
## RCM                        0.088910   0.013643   6.517 7.46e-11 **
## LCB                       -0.050605   0.017224  -2.938  0.00331 **
## Crossing                  -0.049822   0.004860 -10.252  < 2e-16 **
## Finishing                 -0.015799   0.006193  -2.551  0.01075 *
## HeadingAccuracy            0.068067   0.005147  13.225  < 2e-16 **
## ShortPassing               0.054140   0.008023   6.748 1.56e-11 **
## Volleys                   -0.065520   0.004984 -13.146  < 2e-16 **
## Dribbling                 -0.071669   0.007988  -8.972  < 2e-16 **
## Curve                     -0.042596   0.005056  -8.424  < 2e-16 **
## FKAccuracy                -0.052627   0.004731 -11.124  < 2e-16 **
```

```
## LongPassing          -0.087681   0.006526 -13.435  < 2e-16 **
## BallControl           0.105026   0.008958  11.724  < 2e-16 **
## SprintSpeed          -0.016985   0.005260  -3.229  0.00124 **
## Agility              -0.045491   0.004965  -9.162  < 2e-16 **
## Reactions             0.280890   0.006029  46.590  < 2e-16 **
## Balance              -0.073722   0.004397 -16.765  < 2e-16 **
## ShotPower            -0.024565   0.005460  -4.499 6.87e-06 **
## Jumping              -0.039479   0.003591 -10.993  < 2e-16 **
## Stamina              -0.035939   0.004466  -8.048 9.15e-16 **
## LongShots            -0.086528   0.005926 -14.601  < 2e-16 **
## Aggression           -0.046575   0.003912 -11.907  < 2e-16 **
## Interceptions        -0.066467   0.005849 -11.364  < 2e-16 **
## Positioning          -0.135381   0.006097 -22.205  < 2e-16 **
## Vision               -0.080834   0.005547 -14.571  < 2e-16 **
## Penalties            -0.051351   0.004926 -10.424  < 2e-16 **
## Composure             0.141796   0.005432  26.104  < 2e-16 **
## StandingTackle       -0.067019   0.007685  -8.720  < 2e-16 **
## GKPositioning         0.129315   0.009374  13.795  < 2e-16 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04071 on 12707 degrees of freedom
## Multiple R-squared:  0.8778, Adjusted R-squared:  0.8775
## F-statistic:  2537 on 36 and 12707 DF,  p-value: < 2.2e-16
```

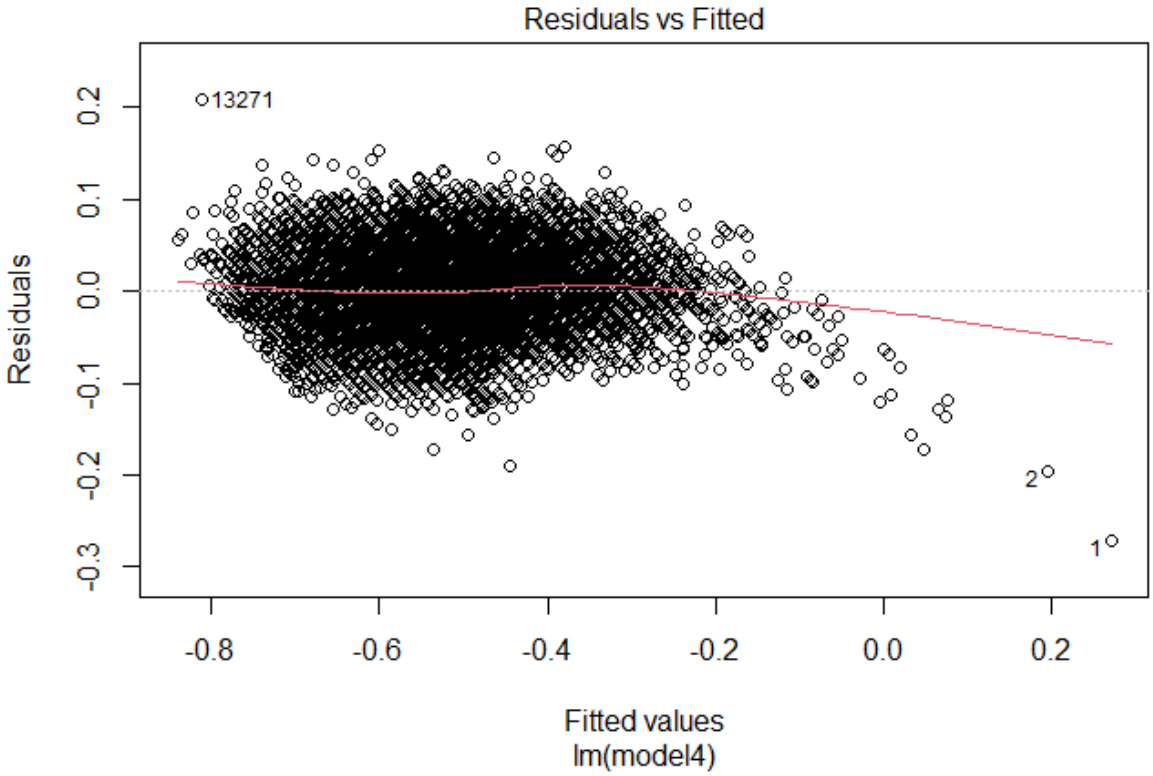
```
print("Number of Coeff learnt =")
```

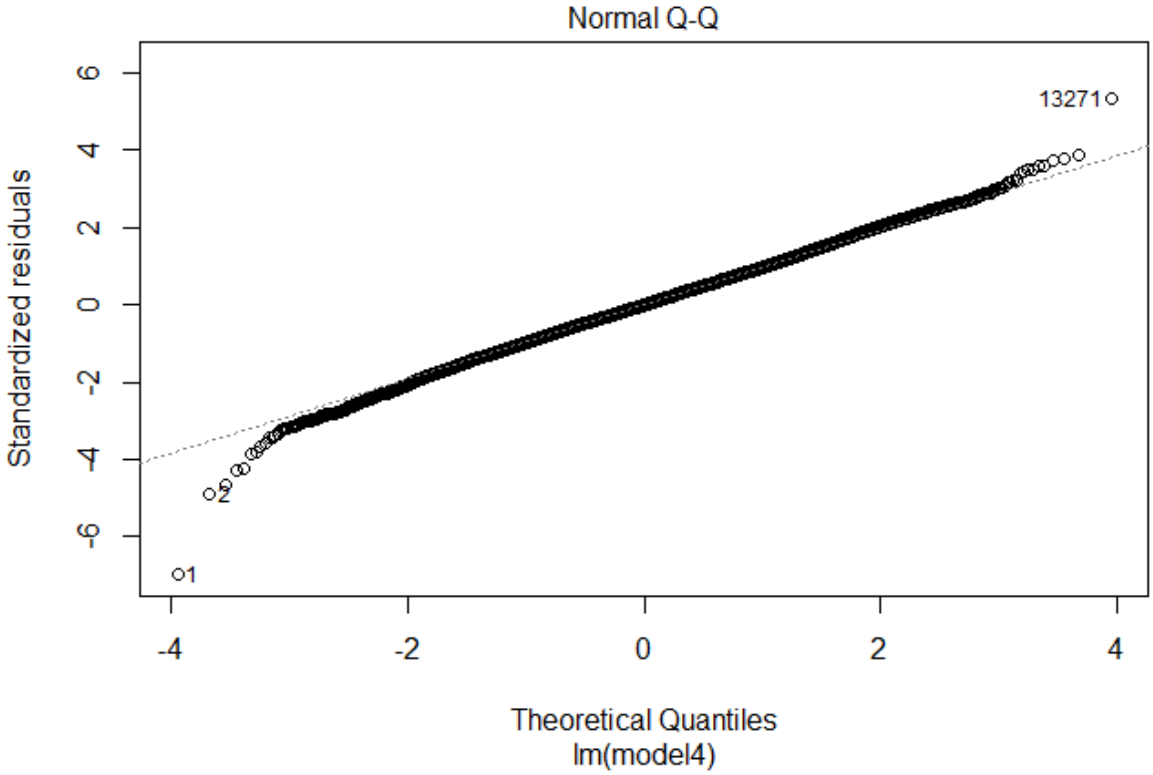
```
## [1] "Number of Coeff learnt ="
```

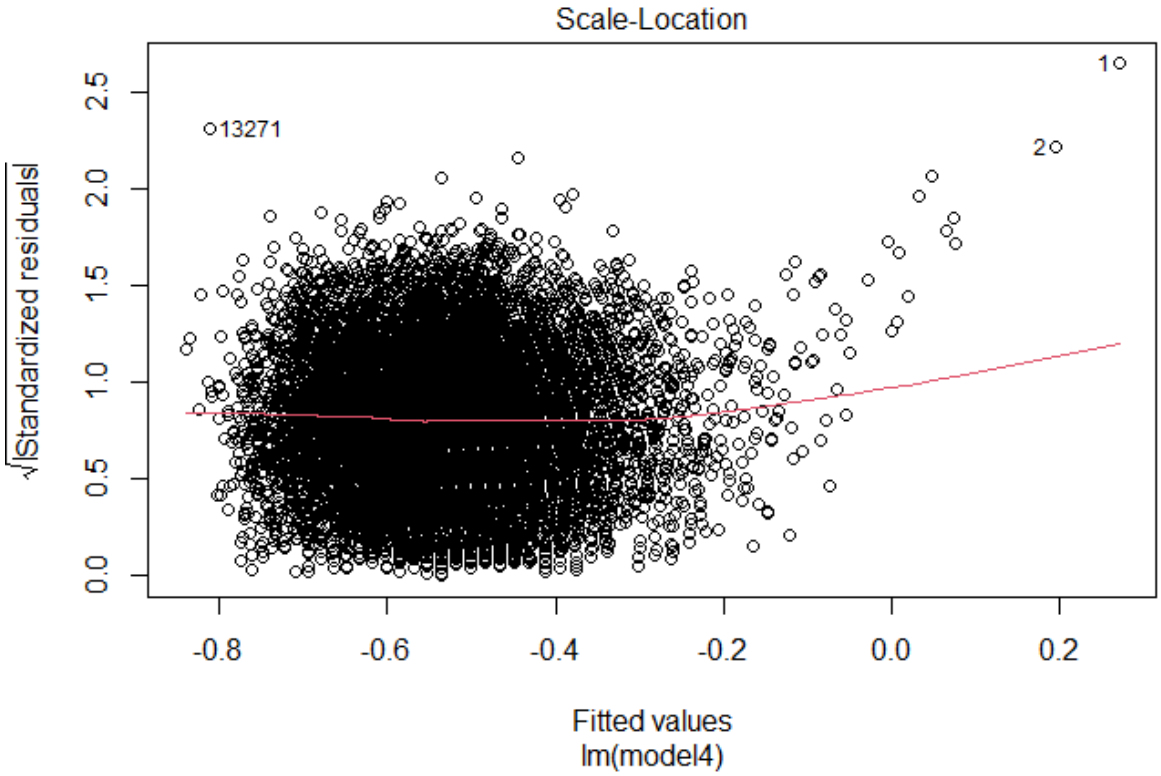
```
print(length(fit4$coefficients))
```

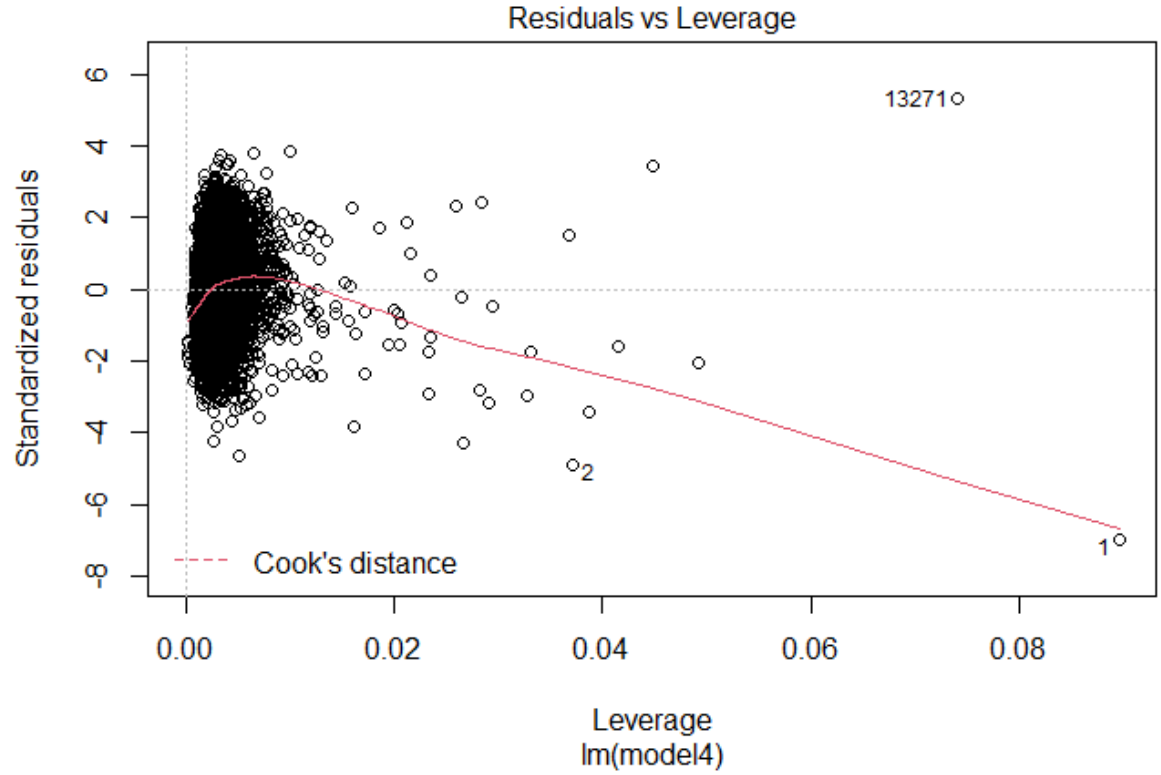
```
## [1] 37
```

```
plot(fit4)
```


Residuals vs Fitted
13271
2
1
Residuals
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.8
-0.6
-0.4
-0.2
0.0
0.2
Fitted values
lm(model4)


Normal Q-Q
13271
2
1
Standardized residuals
6
4
2
0
-2
-4
-6
-4
-2
0
2
4
Theoretical Quantiles
lm(model4)





중간결과 : 모델 개선을 통해 *R^3*값과 *Adjusted R^2* 값이 유의미하게 상승했음을 확인할 수 있다.

<br

2.5.9 모델에서 이상치 확인 및 제거

```
#######   Handling Outliers :



p=dim(data_train)[2]
n=dim(data_train)[1]
print("Original Dim of Training Set = ")
```

```
## [1] "Original Dim of Training Set = "
```

```
print(dim(data_train))
```

```
## [1] 12744    44
```

```
data_train_outrem <- data_train



   #(1)Using Leverage방법론


print("Detecting Outliers using Leverage")
```

```
## [1] "Detecting Outliers using Leverage"
```

```
lev<- lm.influence(fit1)$hat
lev_rem <- lev>3*(p+1)/n
print("Number of Outliers Found = ")
```

```
## [1] "Number of Outliers Found = "
```

```
print(sum(lev_rem == TRUE))
```

```
## [1] 208
```

```
  #(2)Cook's Distance방법론
print("Detecting Outliers using Cook's Distance")
```

```
## [1] "Detecting Outliers using Cook's Distance"
```

```
cd<-cooks.distance(fit1)
cd_rem <- cd>1
print("Number of Outliers Found = ")
```

```
## [1] "Number of Outliers Found = "
```

```
print(sum(cd_rem == TRUE))
```

```
## [1] 0
```

```
 #(3)DFBETAS방법론

print("Detecting Outliers using DFBETAS")
```

```
## [1] "Detecting Outliers using DFBETAS"
```

```
dfb<-dfbeta(fit1)
dim(dfb)
```

```
## [1] 12744    55
```

```
n_dfb = dim(dfb)[1]
p_dfb = dim(dfb)[2]
#Intializing dfb_rem
dfb_rem <- cd_rem
for (i in c(1:n_dfb))
{
  dfb_rem[i]=FALSE
  for (j in c(1:p_dfb))
    if (dfb[i,j]>(2/sqrt(n)))
      dfb_rem[i]<-TRUE
}
print("Number of Outliers Found = ")
```

```
## [1] "Number of Outliers Found = "
```

```
print(sum(dfb_rem == TRUE))
```

```
## [1] 0
```

```
 #(4)DFFITS방법론
print("Detecting Outliers using DFFITS")
```

```
## [1] "Detecting Outliers using DFFITS"
```

```
dfft<-dffits(fit1)
dfft_rem <- dfft>(2*sqrt((p+1)/n))
print("Number of Outliers Found = ")
```

```
## [1] "Number of Outliers Found = "
```

```
print(sum(dfft_rem == TRUE))
```

```
## [1] 507
```

```
 #(5) COVRATIO 방법론

print("Detecting Outliers using COVRATIO")
```

```
## [1] "Detecting Outliers using COVRATIO"
```

```
covr<-covratio(fit1)
covr_rem <- ( covr>(1 + (3*((p+1)/n))) | covr<(1 - (3*((p+1)/n)) ) )
print("Number of Outliers Found = ")
```

```
## [1] "Number of Outliers Found = "
```

```
print(sum(covr_rem == TRUE))
```

```
## [1] 950
```

```
 #위의 모든 방법으로 검출한 이상치들을 제거해준다

data_train_outrem <- data_train
out_rem <- lev_rem | cd_rem | dfb_rem | dfft_rem | covr_rem
data_train_outrem <- data_train_outrem[out_rem==FALSE,]
dim(data_train_outrem)
```

```
## [1] 11462    44
```

```
lin_model5<-sig.formula
fit5<-lm(lin_model5,data_train_outrem)
summary(fit5)
```

```
##
## Call:
## lm(formula = lin_model5, data = data_train_outrem)

##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.146630 -0.023886  0.000155  0.024617  0.101435
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               -0.852792   0.012527 -68.074  < 2e-16 **
## Age                        0.025292   0.002737   9.240  < 2e-16 **
## Value                     -0.080360   0.014871  -5.404 6.66e-08 **
## Wage                       0.664776   0.019077  34.847  < 2e-16 **
## Special                    1.106625   0.043196  25.619  < 2e-16 **
## International.Reputation   0.033516   0.005308   6.314 2.83e-10 **
## Skill.Moves                0.086996   0.003410  25.513  < 2e-16 **
## Height                     0.011040   0.001531   7.211 5.93e-13 **
## Weight                     0.017310   0.004262   4.062 4.90e-05 **
## RCM                        0.122705   0.013778   8.906  < 2e-16 **
## LCB                       -0.095683   0.017746  -5.392 7.11e-08 **
## Crossing                  -0.056562   0.004562 -12.398  < 2e-16 **
## Finishing                 -0.029861   0.005789  -5.158 2.54e-07 **
## HeadingAccuracy            0.067144   0.004702  14.280  < 2e-16 **
## ShortPassing               0.048573   0.007715   6.296 3.17e-10 **
## Volleys                   -0.073092   0.004660 -15.684  < 2e-16 **
## Dribbling                 -0.090467   0.007580 -11.935  < 2e-16 **
## Curve                     -0.050955   0.004701 -10.840  < 2e-16 **
## FKAccuracy                -0.056201   0.004485 -12.530  < 2e-16 **
## LongPassing               -0.102665   0.006335 -16.207  < 2e-16 **
## BallControl                0.100719   0.008547  11.785  < 2e-16 **
## SprintSpeed               -0.027524   0.005234  -5.259 1.48e-07 **
## Agility                   -0.047373   0.004759  -9.954  < 2e-16 **
## Reactions                  0.258430   0.005725  45.139  < 2e-16 **
## Balance                   -0.082239   0.004185 -19.651  < 2e-16 **
## ShotPower                 -0.029838   0.005149  -5.795 7.03e-09 **
## Jumping                   -0.043826   0.003391 -12.923  < 2e-16 **
## Stamina                   -0.049211   0.004373 -11.254  < 2e-16 **
## LongShots                 -0.097242   0.005610 -17.333  < 2e-16 **
## Aggression                -0.053215   0.003574 -14.891  < 2e-16 **
## Interceptions             -0.069126   0.005353 -12.913  < 2e-16 **
```

```
## Positioning               -0.150757   0.005770 -26.128  < 2e-16 **
## Vision                    -0.101010   0.005565 -18.150  < 2e-16 **
## Penalties                 -0.060900   0.004681 -13.011  < 2e-16 **
## Composure                  0.139787   0.005018  27.856  < 2e-16 **

## StandingTackle            -0.071835   0.007126 -10.080  < 2e-16 **
## GKPositioning              0.079294   0.009336   8.494  < 2e-16 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03443 on 11425 degrees of freedom
## Multiple R-squared:  0.8979, Adjusted R-squared:  0.8976
## F-statistic:  2792 on 36 and 11425 DF,  p-value: < 2.2e-16
```

중간점검 : 이상치 제거 후 *adjusted R^2* 값이 상승한것을 확인할 수 있다

2.5.10 Test set을 통한 y값 예측

```
 ##차원확인
dim(data_test)
```

```
## [1] 5463   44
```

```
dim(data_train)
```

```
## [1] 12744    44
```

```
dim(x_train)
```

```
## [1] 12744    43
```

```
dim(x_test)
```

```
## [1] 5463   43
```

```
final_prediction <-predict(fit5, newdata=as.data.frame(x_test)) #최고

#length(final_prediction)
print("See first 10 Prediction ")
```

```
## [1] "See first 10 Prediction "
```

```
print(final_prediction[c(1:10)])
```

```
##      14288      15642      14068      10288      12638       6241
## -0.6251796 -0.6278009 -0.5711662 -0.5215758 -0.5707499 -0.5208227
##       3498      17069       2396
## -0.4974333 -0.7295942 -0.4456081
```

```
print("See first 10 Actual ")
```

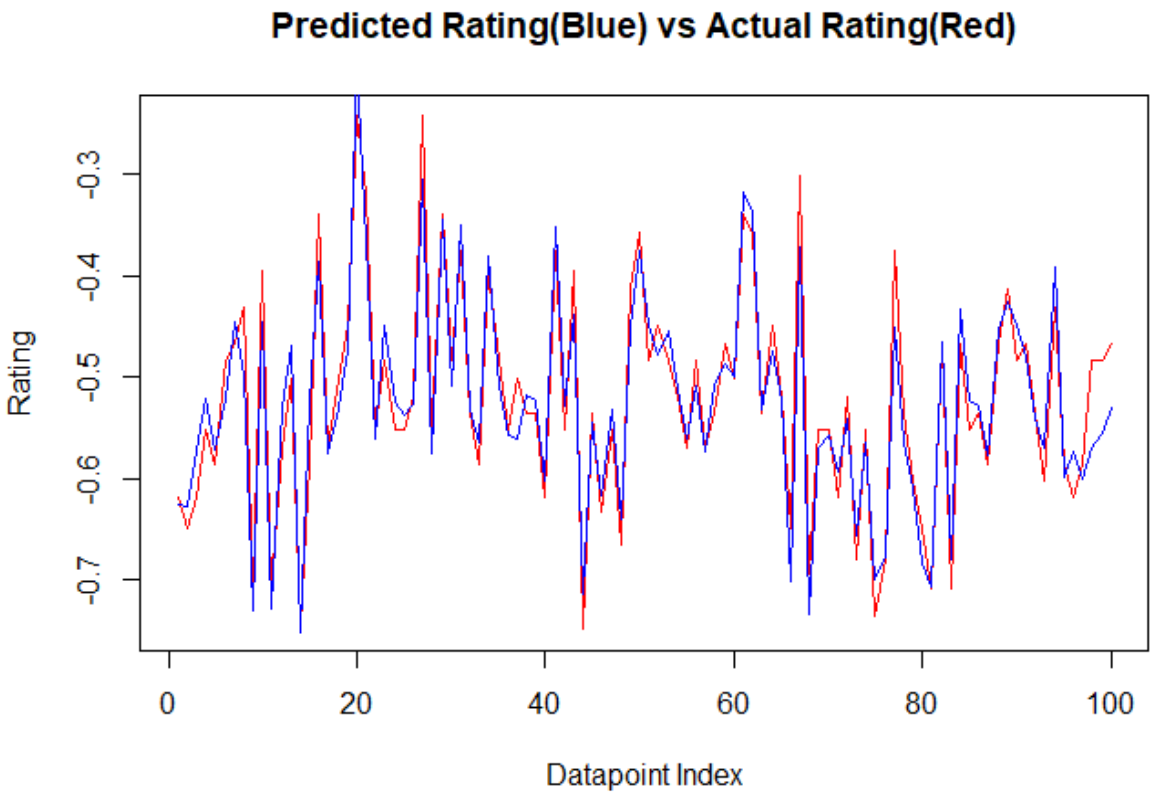
```
## [1] "See first 10 Actual "
```

```
print(y_test[c(1:10)])
```

```
##  [1] -0.6179369 -0.6492267 -0.6179369 -0.5524868 -0.5856600 -0.48
##  [7] -0.4660858 -0.4302803 -0.7083470 -0.3938285
```

```
#Plotting Predicted and Actual Curve for Test Set (Only first 100 Va
```

```
ind = c(1:100)

plot(ind,y_test[c(1:100)],type='l',col="red",xlab="Datapoint Index",
lines(ind,final_prediction[c(1:100)],col="blue")
```



2.5.11 변수선택법 : Step-Forward AIC method

```
#Finding RMSE
rmse_val <- sqrt(norm(y_test-final_prediction)/n)
print("RMSE=")
```

```
## [1] "RMSE="
```

```
print(rmse_val)
```

```
## [1] 0.1169155
```

```
 # - step-forward AIC method

lin_model6<-sig.formula
data_new<-as.data.frame(data_train)
step_model<-lm(lin_model6,data_new)
ols_step_forward_aic(step_model,details = TRUE)
```

```
## Forward Selection Method
## ------------------------
##
## Candidate Terms:
##
## 1 . Age
## 2 . Value
## 3 . Wage
## 4 . Special
## 5 . International.Reputation
## 6 . Skill.Moves
## 7 . Height
## 8 . Weight
## 9 . RCM
## 10 . LCB
## 11 . Crossing
## 12 . Finishing
## 13 . HeadingAccuracy
## 14 . ShortPassing
## 15 . Volleys
## 16 . Dribbling
## 17 . Curve
## 18 . FKAccuracy
## 19 . LongPassing
## 20 . BallControl
## 21 . SprintSpeed
## 22 . Agility
## 23 . Reactions
```

```
## 24 . Balance
## 25 . ShotPower
## 26 . Jumping
## 27 . Stamina
## 28 . LongShots
## 29 . Aggression
## 30 . Interceptions
## 31 . Positioning
## 32 . Vision
## 33 . Penalties
## 34 . Composure
## 35 . StandingTackle
## 36 . GKPositioning
##
##  Step 0: AIC = -18666.29
##  Overall ~ 1
##

## ----------------------------------------------------------------
## Variable                    DF        AIC        Sum Sq       RSS
## ----------------------------------------------------------------
## Reactions                    1    -34899.647    124.188     48.22
## Composure                    1    -28038.310     89.788     82.62
## RCM                          1    -27788.920     88.156     84.26
## Wage                         1    -24297.022     61.595    110.82
## Special                      1    -24251.446     61.198    111.21
## International.Reputation     1    -22869.438     48.459    123.95
## Vision                       1    -22295.368     42.747    129.66
## ShortPassing                 1    -22190.358     41.674    130.74
## LongPassing                  1    -21918.262     38.853    133.56
## BallControl                  1    -21490.024     34.288    138.12
## Age                          1    -21344.312     32.700    139.71
## ShotPower                    1    -21284.354     32.041    140.37
## LCB                          1    -21160.865     30.674    141.74
## Curve                        1    -21016.235     29.056    143.36
## LongShots                    1    -20996.918     28.839    143.57
## Skill.Moves                  1    -20990.183     28.763    143.65
## FKAccuracy                   1    -20777.112     26.341    146.07
## Volleys                      1    -20685.418     25.286    147.13
## Crossing                     1    -20681.029     25.236    147.18
## Aggression                   1    -20676.249     25.180    147.23
## Dribbling                    1    -20455.950     22.613    149.80
## Stamina                      1    -20306.135     20.842    151.57
```

```
## Positioning                1    -20278.232     20.509    151.90
## Penalties                  1    -20140.601     18.860    153.55
## HeadingAccuracy            1    -20056.473     17.843    154.57
## Finishing                  1    -20055.088     17.826    154.59
## Interceptions              1    -19937.451     16.392    156.02
## Agility                    1    -19516.121     11.148    161.26
## Value                      1    -19499.746     10.941    161.47
## Jumping                    1    -19491.274     10.833    161.58
## StandingTackle             1    -19434.382     10.110    162.30
## SprintSpeed                1    -19193.517      7.013    165.40
## Weight                     1    -18963.516      4.001    168.41
## Balance                    1    -18784.888      1.624    170.79
## Height                     1    -18689.677      0.343    172.07
## GKPositioning              1    -18664.337      0.001    172.41
## ---------------------------------------------------------------------
##
##

## - Reactions
##
##
##  Step 1 : AIC = -34899.65
##  Overall ~ Reactions
##
## ---------------------------------------------------------------------
## Variable                   DF       AIC        Sum Sq     RSS
## ---------------------------------------------------------------------
## Wage                        1    -36992.285     7.310    40.919
## Composure                   1    -36773.392     6.601    41.628
## RCM                         1    -36575.118     5.948    42.280
## International.Reputation    1    -36230.456     4.789    43.439
## Special                     1    -35545.687     2.391    45.837
## ShortPassing                1    -35375.764     1.776    46.453
## Skill.Moves                 1    -35374.164     1.770    46.459
## LongPassing                 1    -35362.397     1.727    46.501
## LCB                         1    -35309.168     1.532    46.696
## ShotPower                   1    -35272.574     1.398    46.830
## BallControl                 1    -35245.913     1.300    46.928
## Weight                      1    -35213.479     1.181    47.048
## Vision                      1    -35210.894     1.171    47.058
## Value                       1    -35204.163     1.146    47.082
## Curve                       1    -35138.523     0.903    47.326
## LongShots                   1    -35086.843     0.711    47.518
```

```
## HeadingAccuracy              1    -35072.374     0.657    47.572
## Age                          1    -35069.204     0.645    47.584
## Crossing                     1    -35069.356     0.645    47.583
## FKAccuracy                   1    -35066.820     0.636    47.593
## Volleys                      1    -35063.467     0.623    47.605
## Dribbling                    1    -35056.394     0.597    47.632
## Height                       1    -35050.186     0.574    47.655
## Aggression                   1    -35025.207     0.480    47.748
## Finishing                    1    -35019.524     0.459    47.770
## Stamina                      1    -35003.220     0.398    47.831
## Penalties                    1    -34993.331     0.361    47.868
## SprintSpeed                  1    -34987.204     0.338    47.891
## Jumping                      1    -34975.070     0.292    47.936
## StandingTackle               1    -34950.479     0.200    48.029
## GKPositioning                1    -34947.091     0.187    48.042
## Interceptions                1    -34942.323     0.169    48.060
## Balance                      1    -34937.001     0.149    48.080

## Positioning                  1    -34931.602     0.128    48.100
## Agility                      1    -34927.644     0.113    48.115
## ---------------------------------------------------------------------
##
## - Wage
##
##
##  Step 2 : AIC = -36992.29
##  Overall ~ Reactions + Wage
##
## ---------------------------------------------------------------------
## Variable                   DF       AIC        Sum Sq     RSS
## ---------------------------------------------------------------------
## Composure                   1    -38689.025     5.106    35.812
## RCM                         1    -38400.620     4.287    36.632
## Special                     1    -37573.351     1.830    39.089
## LCB                         1    -37423.171     1.367    39.552
## LongPassing                 1    -37404.383     1.308    39.611
## ShortPassing                1    -37401.437     1.299    39.620
## Age                         1    -37363.428     1.181    39.738
## Skill.Moves                 1    -37355.832     1.157    39.762
## Weight                      1    -37317.973     1.039    39.880
## ShotPower                   1    -37317.231     1.036    39.882
## BallControl                 1    -37274.730     0.903    40.016
## Vision                      1    -37218.233     0.725    40.193
```

```
## Curve                       1    -37174.600     0.588    40.331
## International.Reputation    1    -37173.897     0.585    40.333
## HeadingAccuracy             1    -37159.785     0.541    40.378
## Aggression                  1    -37154.373     0.523    40.395
## LongShots                   1    -37148.741     0.506    40.413
## Crossing                    1    -37132.837     0.455    40.464
## FKAccuracy                  1    -37126.694     0.436    40.483
## Height                      1    -37123.016     0.424    40.495
## Stamina                     1    -37117.410     0.406    40.513
## Dribbling                   1    -37102.690     0.359    40.559
## Volleys                     1    -37100.470     0.352    40.567
## Jumping                     1    -37078.002     0.281    40.638
## Finishing                   1    -37069.890     0.255    40.664
## SprintSpeed                 1    -37056.670     0.213    40.706
## StandingTackle              1    -37052.971     0.201    40.718
## Interceptions               1    -37052.244     0.198    40.720
## Value                       1    -37051.653     0.197    40.722

## Penalties                   1    -37050.045     0.191    40.727
## GKPositioning               1    -37046.903     0.181    40.737
## Balance                     1    -37044.419     0.173    40.745
## Agility                     1    -37013.215     0.074    40.845
## Positioning                 1    -37008.214     0.058    40.861
## ------------------------------------------------------------------
##
## - Composure
##
##
##  Step 3 : AIC = -38689.03
##  Overall ~ Reactions + Wage + Composure
##
## ------------------------------------------------------------------
## Variable                  DF       AIC        Sum Sq     RSS
## ------------------------------------------------------------------
## GKPositioning              1    -39708.277     2.758    33.054
## RCM                        1    -39523.033     2.274    33.538
## Weight                     1    -39326.608     1.753    34.059
## LCB                        1    -39155.495     1.293    34.520
## Balance                    1    -39143.087     1.259    34.553
## Positioning                1    -38991.836     0.846    34.966
## Height                     1    -38989.721     0.841    34.972
## Age                        1    -38911.423     0.625    35.187
## Penalties                  1    -38838.790     0.424    35.388
```

```
## Dribbling                   1    -38831.983     0.405    35.407
## International.Reputation    1    -38821.041     0.375    35.438
## Volleys                     1    -38802.933     0.324    35.488
## Agility                     1    -38795.712     0.304    35.508
## Finishing                   1    -38794.953     0.302    35.510
## LongShots                   1    -38763.828     0.215    35.597
## FKAccuracy                  1    -38751.722     0.181    35.631
## Curve                       1    -38746.829     0.168    35.645
## BallControl                 1    -38744.046     0.160    35.652
## Crossing                    1    -38743.436     0.158    35.654
## Value                       1    -38721.774     0.098    35.715
## Stamina                     1    -38717.136     0.085    35.728
## Jumping                     1    -38715.414     0.080    35.733
## SprintSpeed                 1    -38707.763     0.058    35.754
## HeadingAccuracy             1    -38703.808     0.047    35.765
## Vision                      1    -38701.948     0.042    35.770
## ShotPower                   1    -38699.228     0.034    35.778

## ShortPassing                1    -38693.792     0.019    35.793
## StandingTackle              1    -38692.300     0.015    35.798
## Aggression                  1    -38690.714     0.010    35.802
## Interceptions               1    -38690.643     0.010    35.802
## Special                     1    -38687.754     0.002    35.810
## Skill.Moves                 1    -38687.093     0.000    35.812
## LongPassing                 1    -38687.041     0.000    35.812
## ---------------------------------------------------------------
##
## - GKPositioning
##
##
##  Step 4 : AIC = -39708.28
##  Overall ~ Reactions + Wage + Composure + GKPositioning
##
## ---------------------------------------------------------------
## Variable                  DF       AIC        Sum Sq     RSS
## ---------------------------------------------------------------
## Special                    1    -41097.419     3.418    29.636
## ShortPassing               1    -40751.146     2.602    30.452
## BallControl                1    -40729.875     2.551    30.503
## HeadingAccuracy            1    -40476.788     1.939    31.115
## RCM                        1    -40267.812     1.425    31.630
## Stamina                    1    -40181.607     1.210    31.844
## LCB                        1    -40162.087     1.161    31.893
```

```
## Skill.Moves                  1    -40159.758     1.156    31.899
## LongPassing                  1    -40129.446     1.080    31.975
## ShotPower                    1    -40106.478     1.022    32.033
## Aggression                   1    -40027.100     0.822    32.233
## Dribbling                    1    -39972.977     0.685    32.370
## Weight                       1    -39968.544     0.673    32.381
## Crossing                     1    -39948.819     0.623    32.431
## StandingTackle               1    -39943.207     0.609    32.446
## SprintSpeed                  1    -39933.201     0.583    32.471
## Interceptions                1    -39898.083     0.494    32.561
## Curve                        1    -39831.695     0.324    32.731
## Jumping                      1    -39816.895     0.286    32.769
## LongShots                    1    -39810.524     0.269    32.785
## Height                       1    -39805.696     0.257    32.798
## International.Reputation     1    -39789.512     0.215    32.839
## Balance                      1    -39783.676     0.200    32.854
## FKAccuracy                   1    -39766.438     0.156    32.899

## Age                          1    -39763.058     0.147    32.908
## Finishing                    1    -39746.312     0.104    32.951
## Volleys                      1    -39743.990     0.098    32.957
## Value                        1    -39741.367     0.091    32.964
## Penalties                    1    -39740.965     0.090    32.965
## Vision                       1    -39724.669     0.048    33.007
## Agility                      1    -39724.298     0.047    33.008
## Positioning                  1    -39716.307     0.026    33.028
## ------------------------------------------------------------------
##
## - Special
##
##
##  Step 5 : AIC = -41097.42
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ------------------------------------------------------------------
## Variable                   DF       AIC        Sum Sq     RSS
## ------------------------------------------------------------------
## HeadingAccuracy             1    -42242.820     2.552    27.084
## Balance                     1    -41972.437     1.971    27.665
## Weight                      1    -41877.542     1.764    27.872
## Height                      1    -41553.724     1.047    28.589
## Vision                      1    -41506.067     0.940    28.696
## Positioning                 1    -41473.583     0.866    28.770
```

```
## FKAccuracy                1    -41425.112     0.757    28.879
## LCB                       1    -41423.230     0.753    28.884
## Agility                   1    -41421.402     0.748    28.888
## Curve                     1    -41381.229     0.657    28.979
## LongShots                 1    -41311.919     0.499    29.137
## Crossing                  1    -41279.107     0.424    29.212
## ShortPassing              1    -41250.278     0.358    29.278
## Aggression                1    -41240.329     0.335    29.301
## Volleys                   1    -41223.400     0.296    29.340
## BallControl               1    -41217.143     0.282    29.354
## StandingTackle            1    -41199.930     0.242    29.394
## Dribbling                 1    -41183.807     0.205    29.431
## International.Reputation  1    -41181.413     0.199    29.437
## Finishing                 1    -41176.604     0.188    29.448
## Skill.Moves               1    -41162.639     0.156    29.480
## Interceptions             1    -41162.264     0.155    29.481
## Penalties                 1    -41155.952     0.140    29.496

## Jumping                   1    -41141.920     0.108    29.528
## Stamina                   1    -41131.792     0.084    29.552
## Age                       1    -41130.456     0.081    29.555
## Value                     1    -41119.294     0.055    29.581
## RCM                       1    -41103.623     0.019    29.617
## ShotPower                 1    -41098.092     0.006    29.630
## LongPassing               1    -41095.709     0.001    29.635
## SprintSpeed               1    -41095.623     0.000    29.636
## ------------------------------------------------------------------
##
## - HeadingAccuracy
##
##
##  Step 6 : AIC = -42242.82
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ------------------------------------------------------------------
## Variable                 DF       AIC        Sum Sq     RSS
## ------------------------------------------------------------------
## Skill.Moves               1    -42648.268     0.852    26.232
## BallControl               1    -42556.721     0.663    26.421
## ShortPassing              1    -42505.132     0.556    26.529
## Balance                   1    -42486.135     0.516    26.568
## Weight                    1    -42467.123     0.477    26.608
## Positioning               1    -42447.563     0.436    26.649
```

```
## Height                      1    -42421.384     0.381    26.703
## LongShots                   1    -42353.408     0.238    26.846
## RCM                         1    -42337.694     0.205    26.879
## FKAccuracy                  1    -42332.475     0.194    26.890
## Volleys                     1    -42326.100     0.181    26.904
## Vision                      1    -42324.858     0.178    26.906
## Penalties                   1    -42316.509     0.160    26.924
## International.Reputation    1    -42312.465     0.152    26.933
## LCB                         1    -42297.565     0.120    26.964
## SprintSpeed                 1    -42293.686     0.112    26.972
## Curve                       1    -42276.496     0.076    27.009
## Finishing                   1    -42275.314     0.073    27.011
## LongPassing                 1    -42259.542     0.040    27.045
## Value                       1    -42257.853     0.036    27.048
## Jumping                     1    -42254.592     0.029    27.055
## Agility                     1    -42252.967     0.026    27.059
## Dribbling                   1    -42249.574     0.019    27.066

## StandingTackle              1    -42249.765     0.019    27.065
## Stamina                     1    -42247.017     0.013    27.071
## ShotPower                   1    -42242.049     0.003    27.082
## Crossing                    1    -42241.951     0.002    27.082
## Aggression                  1    -42241.581     0.002    27.083
## Interceptions               1    -42241.654     0.002    27.083
## Age                         1    -42240.924     0.000    27.084
## ---------------------------------------------------------------
##
## - Skill.Moves
##
##
##  Step 7 : AIC = -42648.27
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ---------------------------------------------------------------
## Variable                  DF       AIC        Sum Sq     RSS
## ---------------------------------------------------------------
## Positioning                1    -43187.212     1.090    25.142
## LCB                        1    -42983.772     0.686    25.547
## Balance                    1    -42980.094     0.678    25.554
## Volleys                    1    -42952.651     0.623    25.609
## LongShots                  1    -42946.221     0.610    25.622
## Weight                     1    -42943.519     0.605    25.627
## ShortPassing               1    -42907.647     0.533    25.700
```

```
## Finishing                   1    -42873.199     0.463    25.769
## Penalties                   1    -42865.968     0.448    25.784
## Height                      1    -42858.501     0.433    25.799
## Vision                      1    -42833.844     0.383    25.849
## FKAccuracy                  1    -42815.534     0.346    25.886
## StandingTackle              1    -42808.155     0.331    25.901
## Curve                       1    -42794.386     0.303    25.929
## BallControl                 1    -42770.679     0.255    25.977
## Interceptions               1    -42761.589     0.236    25.996
## International.Reputation    1    -42722.025     0.155    26.077
## RCM                         1    -42719.113     0.150    26.083
## Agility                     1    -42716.288     0.144    26.089
## LongPassing                 1    -42704.718     0.120    26.112
## Dribbling                   1    -42699.297     0.109    26.123
## Aggression                  1    -42685.062     0.080    26.153
## ShotPower                   1    -42677.899     0.065    26.167
## Stamina                     1    -42677.328     0.064    26.168

## SprintSpeed                 1    -42660.773     0.030    26.202
## Value                       1    -42660.299     0.029    26.203
## Crossing                    1    -42657.844     0.024    26.208
## Age                         1    -42652.966     0.014    26.218
## Jumping                     1    -42646.979     0.001    26.231
## ---------------------------------------------------------------
##
## - Positioning
##
##
##  Step 8 : AIC = -43187.21
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ---------------------------------------------------------------
## Variable                  DF       AIC        Sum Sq     RSS
## ---------------------------------------------------------------
## BallControl                1    -43564.880     0.738    24.404
## Balance                    1    -43458.456     0.533    24.609
## Weight                     1    -43433.634     0.485    24.657
## ShortPassing               1    -43387.113     0.395    24.747
## Height                     1    -43357.530     0.338    24.804
## FKAccuracy                 1    -43307.368     0.240    24.902
## RCM                        1    -43280.822     0.188    24.954
## International.Reputation   1    -43265.775     0.158    24.984
## SprintSpeed                1    -43244.236     0.116    25.026
```

```
## LongShots                    1    -43236.256     0.101    25.042
## Curve                        1    -43232.910     0.094    25.048
## Volleys                      1    -43229.098     0.086    25.056
## Vision                       1    -43228.192     0.085    25.057
## Interceptions                1    -43221.012     0.071    25.072
## Penalties                    1    -43213.932     0.057    25.085
## Dribbling                    1    -43211.135     0.051    25.091
## StandingTackle               1    -43203.768     0.037    25.105
## Agility                      1    -43202.121     0.033    25.109
## Jumping                      1    -43201.263     0.032    25.110
## Value                        1    -43196.311     0.022    25.120
## LCB                          1    -43196.298     0.022    25.120
## Stamina                      1    -43195.643     0.021    25.121
## Aggression                   1    -43195.604     0.020    25.122
## Age                          1    -43191.109     0.012    25.130
## ShotPower                    1    -43188.730     0.007    25.135
## Crossing                     1    -43187.202     0.004    25.138

## Finishing                    1    -43187.328     0.004    25.138
## LongPassing                  1    -43185.240     0.000    25.142
## ---------------------------------------------------------------------
##
## - BallControl
##
##
##  Step 9 : AIC = -43564.88
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ---------------------------------------------------------------------
## Variable                   DF       AIC        Sum Sq     RSS
## ---------------------------------------------------------------------
## Weight                       1    -43863.173     0.568    23.836
## Balance                      1    -43854.826     0.553    23.851
## Height                       1    -43738.231     0.333    24.071
## FKAccuracy                   1    -43678.809     0.221    24.183
## International.Reputation     1    -43650.868     0.168    24.236
## LongShots                    1    -43649.045     0.164    24.240
## Vision                       1    -43647.263     0.161    24.243
## LCB                          1    -43632.568     0.133    24.271
## SprintSpeed                  1    -43631.496     0.131    24.273
## Volleys                      1    -43628.799     0.126    24.278
## Curve                        1    -43624.582     0.118    24.286
## ShortPassing                 1    -43612.476     0.095    24.309
```

```
## Penalties                    1    -43610.338     0.091    24.313
## Stamina                      1    -43602.816     0.076    24.328
## RCM                          1    -43593.564     0.059    24.345
## Agility                      1    -43589.352     0.051    24.353
## Dribbling                    1    -43587.804     0.048    24.356
## Age                          1    -43586.928     0.046    24.358
## LongPassing                  1    -43576.473     0.026    24.378
## Value                        1    -43572.032     0.018    24.387
## Crossing                     1    -43570.044     0.014    24.390
## Interceptions                1    -43567.740     0.009    24.395
## Finishing                    1    -43565.649     0.005    24.399
## Aggression                   1    -43563.569     0.001    24.403
## StandingTackle               1    -43563.256     0.001    24.403
## ShotPower                    1    -43562.897     0.000    24.404
## Jumping                      1    -43563.120     0.000    24.404
## ---------------------------------------------------------------------
##

## - Weight
##
##
##  Step 10 : AIC = -43863.17
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ---------------------------------------------------------------------
## Variable                   DF       AIC        Sum Sq     RSS
## ---------------------------------------------------------------------
## FKAccuracy                  1    -43992.277     0.244    23.592
## LongShots                   1    -43989.432     0.239    23.597
## SprintSpeed                 1    -43977.546     0.217    23.619
## Balance                     1    -43960.777     0.186    23.650
## Vision                      1    -43946.193     0.158    23.677
## Volleys                     1    -43945.195     0.157    23.679
## Height                      1    -43944.876     0.156    23.680
## International.Reputation    1    -43944.135     0.155    23.681
## Curve                       1    -43925.225     0.120    23.716
## ShortPassing                1    -43915.029     0.101    23.735
## Penalties                   1    -43911.265     0.094    23.742
## Stamina                     1    -43897.999     0.069    23.767
## LCB                         1    -43897.386     0.068    23.768
## RCM                         1    -43881.313     0.038    23.798
## LongPassing                 1    -43873.967     0.024    23.812
## Dribbling                   1    -43871.642     0.020    23.816
```

```
## Age                          1    -43870.445     0.017    23.818
## ShotPower                    1    -43869.475     0.016    23.820
## Value                        1    -43869.029     0.015    23.821
## Interceptions                1    -43868.449     0.014    23.822
## Finishing                    1    -43867.064     0.011    23.825
## Aggression                   1    -43865.358     0.008    23.828
## Crossing                     1    -43862.933     0.003    23.832
## Jumping                      1    -43862.874     0.003    23.833
## StandingTackle               1    -43862.143     0.002    23.834
## Agility                      1    -43861.208     0.000    23.836
## ---------------------------------------------------------------------
##
## - FKAccuracy
##
##
##  Step 11 : AIC = -43992.28
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special

##
## ---------------------------------------------------------------------
## Variable                    DF       AIC        Sum Sq     RSS
## ---------------------------------------------------------------------
## Balance                      1    -44120.985     0.241    23.351
## International.Reputation     1    -44087.989     0.180    23.412
## Height                       1    -44074.339     0.155    23.437
## LongShots                    1    -44056.587     0.122    23.469
## SprintSpeed                  1    -44049.315     0.109    23.483
## Vision                       1    -44048.464     0.107    23.484
## ShortPassing                 1    -44047.468     0.106    23.486
## Volleys                      1    -44028.948     0.071    23.520
## Interceptions                1    -44020.767     0.056    23.535
## RCM                          1    -44017.538     0.050    23.541
## Age                          1    -44012.557     0.041    23.551
## Aggression                   1    -44008.754     0.034    23.558
## Dribbling                    1    -44007.080     0.031    23.561
## StandingTackle               1    -44006.312     0.030    23.562
## Penalties                    1    -44000.793     0.019    23.572
## LCB                          1    -43999.629     0.017    23.575
## Curve                        1    -43999.133     0.016    23.575
## Stamina                      1    -43998.481     0.015    23.577
## Value                        1    -43997.772     0.014    23.578
## LongPassing                  1    -43997.828     0.014    23.578
## Agility                      1    -43993.846     0.007    23.585
```

```
## Jumping                      1    -43991.699     0.003    23.589
## Crossing                     1    -43990.785     0.001    23.591
## Finishing                    1    -43990.289     0.000    23.592
## ShotPower                    1    -43990.334     0.000    23.592
## ---------------------------------------------------------------
##
## - Balance
##
##
##  Step 12 : AIC = -44120.99
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ---------------------------------------------------------------
## Variable                   DF       AIC        Sum Sq     RSS
## ---------------------------------------------------------------
## SprintSpeed                 1    -44226.069     0.195    23.156
## International.Reputation    1    -44220.406     0.185    23.166

## Vision                      1    -44198.625     0.145    23.206
## LongShots                   1    -44195.240     0.139    23.212
## Interceptions               1    -44191.420     0.132    23.219
## Height                      1    -44171.317     0.096    23.255
## StandingTackle              1    -44170.405     0.094    23.257
## Aggression                  1    -44159.096     0.073    23.278
## ShortPassing                1    -44157.463     0.070    23.281
## Volleys                     1    -44152.927     0.062    23.289
## LongPassing                 1    -44145.896     0.049    23.302
## Age                         1    -44139.291     0.037    23.314
## Dribbling                   1    -44130.267     0.021    23.330
## Curve                       1    -44129.425     0.019    23.332
## Value                       1    -44126.086     0.013    23.338
## Stamina                     1    -44126.245     0.013    23.338
## Penalties                   1    -44125.452     0.012    23.339
## Agility                     1    -44124.337     0.010    23.341
## RCM                         1    -44123.154     0.008    23.343
## Crossing                    1    -44120.889     0.003    23.348
## Jumping                     1    -44120.617     0.003    23.348
## ShotPower                   1    -44120.178     0.002    23.349
## LCB                         1    -44119.748     0.001    23.350
## Finishing                   1    -44119.193     0.000    23.351
## ---------------------------------------------------------------
##
## - SprintSpeed
```

```
##
##
##  Step 13 : AIC = -44226.07
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ----------------------------------------------------------------
## Variable                    DF       AIC        Sum Sq     RSS
## ----------------------------------------------------------------
## International.Reputation     1    -44343.403     0.216    22.940
## ShortPassing                 1    -44304.236     0.145    23.010
## LongShots                    1    -44285.621     0.112    23.044
## Vision                       1    -44273.431     0.090    23.066
## Height                       1    -44272.849     0.088    23.067
## RCM                          1    -44271.872     0.087    23.069
## Age                          1    -44270.512     0.084    23.071
## Volleys                      1    -44259.001     0.063    23.092
## Dribbling                    1    -44252.993     0.052    23.103

## Interceptions                1    -44250.206     0.047    23.108
## Aggression                   1    -44241.277     0.031    23.124
## LCB                          1    -44240.204     0.029    23.126
## StandingTackle               1    -44238.081     0.025    23.130
## Curve                        1    -44232.850     0.016    23.140
## Value                        1    -44231.496     0.013    23.142
## Penalties                    1    -44230.710     0.012    23.144
## LongPassing                  1    -44228.166     0.007    23.148
## Agility                      1    -44227.923     0.007    23.149
## Crossing                     1    -44227.305     0.006    23.150
## Stamina                      1    -44225.322     0.002    23.153
## ShotPower                    1    -44224.583     0.001    23.155
## Finishing                    1    -44224.244     0.000    23.155
## Jumping                      1    -44224.283     0.000    23.155
## ----------------------------------------------------------------
##
## - International.Reputation
##
##
##  Step 14 : AIC = -44343.4
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ----------------------------------------------------------------
## Variable          DF       AIC        Sum Sq     RSS      R-Sq
## ----------------------------------------------------------------
```

```
## ShortPassing       1    -44423.280     0.147    22.793    0.868
## LongShots          1    -44401.268     0.108    22.832    0.868
## Vision             1    -44394.669     0.096    22.844    0.868
## RCM                1    -44393.853     0.094    22.846    0.867
## Height             1    -44389.909     0.087    22.853    0.867
## Volleys            1    -44386.990     0.082    22.858    0.867
## Age                1    -44374.893     0.060    22.880    0.867
## Dribbling          1    -44371.705     0.054    22.885    0.867
## Interceptions      1    -44365.865     0.044    22.896    0.867
## LCB                1    -44360.340     0.034    22.906    0.867
## Value              1    -44359.020     0.032    22.908    0.867
## Aggression         1    -44357.275     0.029    22.911    0.867
## Curve              1    -44353.772     0.022    22.918    0.867
## Penalties          1    -44353.071     0.021    22.919    0.867
## StandingTackle     1    -44353.088     0.021    22.919    0.867
## Stamina            1    -44346.863     0.010    22.930    0.867
## Crossing           1    -44346.063     0.008    22.931    0.867

## LongPassing        1    -44345.658     0.008    22.932    0.867
## Agility            1    -44344.860     0.006    22.934    0.867
## ShotPower          1    -44342.139     0.001    22.939    0.867
## Finishing          1    -44341.596     0.000    22.940    0.867
## Jumping            1    -44341.471     0.000    22.940    0.867
## ---------------------------------------------------------------------
##
## - ShortPassing
##
##
##  Step 15 : AIC = -44423.28
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ---------------------------------------------------------------------
## Variable          DF       AIC        Sum Sq     RSS      R-Sq
## ---------------------------------------------------------------------
## Vision             1    -44498.700     0.138    22.655    0.869
## LongPassing        1    -44489.924     0.122    22.671    0.869
## LongShots          1    -44470.011     0.087    22.706    0.868
## Age                1    -44467.258     0.082    22.711    0.868
## Height             1    -44467.469     0.082    22.710    0.868
## Volleys            1    -44456.406     0.063    22.730    0.868
## Dribbling          1    -44451.519     0.054    22.739    0.868
## Interceptions      1    -44445.452     0.043    22.750    0.868
## RCM                1    -44442.461     0.038    22.755    0.868
```

```
## LCB                1    -44441.367     0.036    22.757    0.868
## Value              1    -44439.552     0.033    22.760    0.868
## StandingTackle     1    -44434.587     0.024    22.769    0.868
## Aggression         1    -44431.039     0.017    22.776    0.868
## Stamina            1    -44430.385     0.016    22.777    0.868
## Curve              1    -44429.520     0.015    22.778    0.868
## Penalties          1    -44428.311     0.013    22.780    0.868
## Crossing           1    -44426.626     0.010    22.783    0.868
## Jumping            1    -44424.475     0.006    22.787    0.868
## Agility            1    -44422.703     0.003    22.790    0.868
## Finishing          1    -44421.372     0.000    22.793    0.868
## ShotPower          1    -44421.407     0.000    22.793    0.868
## ---------------------------------------------------------------------
##
## - Vision
##
##

##  Step 16 : AIC = -44498.7
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ---------------------------------------------------------------------
## Variable          DF       AIC        Sum Sq     RSS      R-Sq
## ---------------------------------------------------------------------
## Interceptions      1    -44549.554     0.094    22.561    0.869
## LongPassing        1    -44544.983     0.086    22.569    0.869
## RCM                1    -44544.539     0.085    22.570    0.869
## Height             1    -44543.426     0.083    22.572    0.869
## Age                1    -44536.666     0.071    22.584    0.869
## StandingTackle     1    -44534.900     0.068    22.587    0.869
## LongShots          1    -44531.774     0.062    22.593    0.869
## Volleys            1    -44523.481     0.048    22.607    0.869
## Dribbling          1    -44519.345     0.040    22.615    0.869
## Aggression         1    -44516.614     0.035    22.620    0.869
## Value              1    -44513.715     0.030    22.625    0.869
## Crossing           1    -44505.657     0.016    22.639    0.869
## Curve              1    -44504.214     0.013    22.642    0.869
## Stamina            1    -44501.853     0.009    22.646    0.869
## LCB                1    -44501.097     0.008    22.647    0.869
## Penalties          1    -44500.825     0.007    22.648    0.869
## Finishing          1    -44500.192     0.006    22.649    0.869
## Jumping            1    -44497.850     0.002    22.653    0.869
## Agility            1    -44497.219     0.001    22.654    0.869
```

```
## ShotPower           1    -44496.710     0.000    22.655    0.869
## ---------------------------------------------------------------------
##
## - Interceptions
##
##
##  Step 17 : AIC = -44549.55
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ---------------------------------------------------------------------
## Variable          DF       AIC        Sum Sq     RSS      R-Sq
## ---------------------------------------------------------------------
## LCB                1    -44659.729     0.198    22.363    0.870
## LongShots          1    -44655.166     0.190    22.371    0.870
## Volleys            1    -44648.409     0.178    22.383    0.870
## RCM                1    -44602.619     0.097    22.464    0.870
## Height             1    -44593.306     0.081    22.480    0.870

## LongPassing        1    -44587.799     0.071    22.490    0.870
## Age                1    -44586.171     0.068    22.493    0.870
## Dribbling          1    -44585.867     0.068    22.493    0.870
## Penalties          1    -44579.384     0.056    22.505    0.869
## Curve              1    -44569.676     0.039    22.522    0.869
## Value              1    -44564.522     0.030    22.531    0.869
## ShotPower          1    -44560.236     0.022    22.539    0.869
## Aggression         1    -44558.853     0.020    22.541    0.869
## Crossing           1    -44555.259     0.014    22.548    0.869
## Finishing          1    -44554.194     0.012    22.549    0.869
## Stamina            1    -44553.742     0.011    22.550    0.869
## Agility            1    -44552.294     0.008    22.553    0.869
## StandingTackle     1    -44549.088     0.003    22.558    0.869
## Jumping            1    -44547.578     0.000    22.561    0.869
## ---------------------------------------------------------------------
##
## - LCB
##
##
##  Step 18 : AIC = -44659.73
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ---------------------------------------------------------------------
## Variable          DF       AIC        Sum Sq     RSS      R-Sq
## ---------------------------------------------------------------------
```

```
## StandingTackle     1    -44727.718     0.122    22.241    0.871
## LongShots          1    -44715.289     0.101    22.263    0.871
## Age                1    -44710.178     0.092    22.272    0.871
## Volleys            1    -44706.872     0.086    22.277    0.871
## Height             1    -44703.113     0.079    22.284    0.871
## LongPassing        1    -44699.800     0.074    22.290    0.871
## Aggression         1    -44686.679     0.051    22.313    0.871
## Dribbling          1    -44682.728     0.044    22.320    0.871
## Value              1    -44673.894     0.028    22.335    0.870
## Curve              1    -44668.892     0.020    22.344    0.870
## Crossing           1    -44666.752     0.016    22.348    0.870
## Stamina            1    -44666.028     0.015    22.349    0.870
## Penalties          1    -44663.278     0.010    22.354    0.870
## RCM                1    -44659.980     0.004    22.359    0.870
## Finishing          1    -44660.022     0.004    22.359    0.870
## ShotPower          1    -44659.748     0.004    22.360    0.870
## Agility            1    -44657.791     0.000    22.363    0.870

## Jumping            1    -44657.744     0.000    22.363    0.870
## ---------------------------------------------------------------------
##
## - StandingTackle
##
##
##  Step 19 : AIC = -44727.72
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ---------------------------------------------------------------------
## Variable       DF       AIC        Sum Sq     RSS      R-Sq     A
## ---------------------------------------------------------------------
## LongShots       1    -44802.620     0.134    22.107    0.872
## Volleys         1    -44788.704     0.110    22.131    0.872
## Height          1    -44771.923     0.080    22.160    0.871
## Age             1    -44769.656     0.077    22.164    0.871
## LongPassing     1    -44764.862     0.068    22.173    0.871
## Aggression      1    -44761.079     0.062    22.179    0.871
## Dribbling       1    -44747.469     0.038    22.203    0.871
## Value           1    -44741.909     0.028    22.213    0.871
## Curve           1    -44739.158     0.023    22.218    0.871
## Penalties       1    -44733.598     0.014    22.227    0.871
## RCM             1    -44732.909     0.013    22.228    0.871
## Stamina         1    -44733.344     0.013    22.228    0.871
## Crossing        1    -44732.009     0.011    22.230    0.871
```

```
## ShotPower       1    -44731.688     0.010    22.231    0.871
## Jumping         1    -44727.828     0.004    22.237    0.871
## Agility         1    -44726.073     0.001    22.240    0.871
## Finishing       1    -44725.843     0.000    22.241    0.871
## ------------------------------------------------------------------
##
## - LongShots
##
##
##  Step 20 : AIC = -44802.62
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ------------------------------------------------------------------
## Variable       DF       AIC        Sum Sq     RSS      R-Sq     A
## ------------------------------------------------------------------
## Volleys         1    -44849.504     0.085    22.023    0.872
## Aggression      1    -44849.583     0.085    22.022    0.872

## Height          1    -44847.520     0.081    22.026    0.872
## LongPassing     1    -44844.227     0.076    22.032    0.872
## Age             1    -44838.735     0.066    22.041    0.872
## Dribbling       1    -44826.018     0.044    22.063    0.872
## Curve           1    -44820.013     0.034    22.074    0.872
## Crossing        1    -44818.841     0.032    22.076    0.872
## Value           1    -44817.981     0.030    22.077    0.872
## Jumping         1    -44809.411     0.015    22.092    0.872
## Finishing       1    -44808.687     0.014    22.093    0.872
## Penalties       1    -44808.693     0.014    22.093    0.872
## Agility         1    -44805.624     0.009    22.098    0.872
## Stamina         1    -44803.550     0.005    22.102    0.872
## RCM             1    -44800.751     0.000    22.107    0.872
## ShotPower       1    -44800.865     0.000    22.107    0.872
## ------------------------------------------------------------------
##
## - Aggression
##
##
##  Step 21 : AIC = -44849.58
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ------------------------------------------------------------------
## Variable       DF       AIC        Sum Sq     RSS      R-Sq     A
## ------------------------------------------------------------------
```

```
## Volleys         1    -44903.459     0.096    21.926    0.873
## LongPassing     1    -44898.006     0.087    21.935    0.873
## Height          1    -44896.735     0.085    21.938    0.873
## Age             1    -44893.195     0.079    21.944    0.873
## Dribbling       1    -44880.403     0.057    21.966    0.873
## Crossing        1    -44878.449     0.053    21.969    0.873
## Curve           1    -44877.291     0.051    21.971    0.873
## Value           1    -44864.372     0.029    21.993    0.872
## Penalties       1    -44859.925     0.021    22.001    0.872
## Jumping         1    -44856.526     0.015    22.007    0.872
## RCM             1    -44855.325     0.013    22.009    0.872
## Agility         1    -44854.864     0.013    22.010    0.872
## Finishing       1    -44852.443     0.008    22.014    0.872
## Stamina         1    -44850.804     0.006    22.017    0.872
## ShotPower       1    -44848.160     0.001    22.021    0.872
## ---------------------------------------------------------------
##

## - Volleys
##
##
##  Step 22 : AIC = -44903.46
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ---------------------------------------------------------------
## Variable       DF       AIC        Sum Sq     RSS      R-Sq     A
## ---------------------------------------------------------------
## LongPassing     1    -44965.191     0.109    21.817    0.873
## Height          1    -44953.136     0.089    21.837    0.873
## Age             1    -44946.334     0.077    21.849    0.873
## Crossing        1    -44942.551     0.071    21.855    0.873
## Dribbling       1    -44939.851     0.066    21.860    0.873
## Curve           1    -44927.841     0.045    21.881    0.873
## Value           1    -44918.323     0.029    21.897    0.873
## Jumping         1    -44918.142     0.029    21.897    0.873
## Agility         1    -44915.773     0.025    21.901    0.873
## Finishing       1    -44913.747     0.021    21.905    0.873
## Penalties       1    -44909.691     0.014    21.912    0.873
## RCM             1    -44907.218     0.010    21.916    0.873
## ShotPower       1    -44902.972     0.003    21.923    0.873
## Stamina         1    -44901.707     0.000    21.926    0.873
## ---------------------------------------------------------------
##
```

```
## - LongPassing
##
##
##  Step 23 : AIC = -44965.19
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ----------------------------------------------------------------
## Variable     DF       AIC        Sum Sq     RSS      R-Sq     Adj
## ----------------------------------------------------------------
## Height        1    -45016.470     0.091    21.726    0.874
## Age           1    -45008.211     0.077    21.740    0.874
## Dribbling     1    -45003.193     0.068    21.748    0.874
## Crossing      1    -45000.727     0.064    21.752    0.874
## Curve         1    -44991.875     0.049    21.768    0.874
## Jumping       1    -44990.250     0.046    21.770    0.874
## Agility       1    -44981.590     0.031    21.785    0.874
## Value         1    -44979.672     0.028    21.788    0.874

## Penalties     1    -44977.782     0.025    21.792    0.874
## Finishing     1    -44971.650     0.014    21.802    0.874
## ShotPower     1    -44964.500     0.002    21.814    0.873
## RCM           1    -44963.860     0.001    21.816    0.873
## Stamina       1    -44963.360     0.000    21.816    0.873
## ----------------------------------------------------------------
##
## - Height
##
##
##  Step 24 : AIC = -45016.47
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ----------------------------------------------------------------
## Variable     DF       AIC        Sum Sq     RSS      R-Sq     Adj
## ----------------------------------------------------------------
## Age           1    -45063.550     0.084    21.642    0.874
## Dribbling     1    -45053.150     0.066    21.660    0.874
## Crossing      1    -45046.943     0.055    21.670    0.874
## Curve         1    -45042.620     0.048    21.678    0.874
## Jumping       1    -45040.766     0.045    21.681    0.874
## Agility       1    -45032.215     0.030    21.695    0.874
## Value         1    -45030.253     0.027    21.699    0.874
## Penalties     1    -45028.806     0.024    21.701    0.874
## Finishing     1    -45022.091     0.013    21.713    0.874
```

```
## ShotPower     1    -45015.377     0.002    21.724    0.874
## RCM           1    -45015.123     0.001    21.725    0.874
## Stamina       1    -45014.778     0.001    21.725    0.874
## ----------------------------------------------------------------
##
## - Age
##
##
##  Step 25 : AIC = -45063.55
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ----------------------------------------------------------------
## Variable     DF       AIC        Sum Sq     RSS      R-Sq     Adj
## ----------------------------------------------------------------
## Crossing      1    -45105.014     0.074    21.568    0.875
## Dribbling     1    -45091.798     0.051    21.591    0.875
## Jumping       1    -45089.544     0.047    21.595    0.875

## Curve         1    -45088.553     0.046    21.596    0.875
## Agility       1    -45081.106     0.033    21.609    0.875
## Penalties     1    -45077.031     0.026    21.616    0.875
## Value         1    -45075.296     0.023    21.619    0.875
## Finishing     1    -45070.795     0.016    21.626    0.875
## RCM           1    -45063.550     0.003    21.639    0.874
## ShotPower     1    -45063.345     0.003    21.639    0.874
## Stamina       1    -45061.749     0.000    21.642    0.874
## ----------------------------------------------------------------
##
## - Crossing
##
##
##  Step 26 : AIC = -45105.01
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ----------------------------------------------------------------
## Variable     DF       AIC        Sum Sq     RSS      R-Sq     Adj
## ----------------------------------------------------------------
## Jumping       1    -45151.238     0.081    21.487    0.875
## Agility       1    -45127.826     0.042    21.526    0.875
## Penalties     1    -45127.630     0.042    21.527    0.875
## Dribbling     1    -45122.389     0.033    21.536    0.875
## Curve         1    -45119.174     0.027    21.541    0.875
## Value         1    -45117.744     0.025    21.544    0.875
```

```
## RCM           1    -45108.354     0.009    21.559    0.875
## Finishing     1    -45106.815     0.006    21.562    0.875
## Stamina       1    -45104.731     0.003    21.566    0.875
## ShotPower     1    -45103.293     0.000    21.568    0.875
## ---------------------------------------------------------------------
##
## - Jumping
##
##
##  Step 27 : AIC = -45151.24
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ---------------------------------------------------------------------
## Variable     DF       AIC        Sum Sq     RSS      R-Sq     Adj
## ---------------------------------------------------------------------
## Penalties     1    -45188.120     0.065    21.422    0.876
## Dribbling     1    -45177.841     0.048    21.439    0.876

## Curve         1    -45174.184     0.042    21.445    0.876
## Agility       1    -45174.465     0.042    21.444    0.876
## Value         1    -45163.195     0.024    21.463    0.876
## Stamina       1    -45152.341     0.005    21.482    0.875
## RCM           1    -45151.141     0.003    21.484    0.875
## Finishing     1    -45150.333     0.002    21.485    0.875
## ShotPower     1    -45149.359     0.000    21.487    0.875
## ---------------------------------------------------------------------
##
## - Penalties
##
##
##  Step 28 : AIC = -45188.12
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ---------------------------------------------------------------------
## Variable     DF       AIC        Sum Sq     RSS      R-Sq     Adj
## ---------------------------------------------------------------------
## Agility       1    -45226.361     0.068    21.354    0.876
## Dribbling     1    -45219.629     0.056    21.365    0.876
## Curve         1    -45216.593     0.051    21.370    0.876
## Value         1    -45200.597     0.024    21.397    0.876
## Stamina       1    -45194.672     0.014    21.407    0.876
## Finishing     1    -45188.502     0.004    21.418    0.876
## RCM           1    -45188.100     0.003    21.418    0.876
```

```
## ShotPower     1    -45186.320     0.000    21.421    0.876
## ----------------------------------------------------------------
##
## - Agility
##
##
##  Step 29 : AIC = -45226.36
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ----------------------------------------------------------------
## Variable     DF       AIC        Sum Sq     RSS      R-Sq     Adj
## ----------------------------------------------------------------
## Curve         1    -45259.807     0.059    21.295    0.876
## Dribbling     1    -45259.154     0.058    21.296    0.876
## Value         1    -45239.165     0.025    21.329    0.876
## Stamina       1    -45237.219     0.022    21.332    0.876
## RCM           1    -45227.931     0.006    21.348    0.876

## ShotPower     1    -45226.790     0.004    21.350    0.876
## Finishing     1    -45225.071     0.001    21.353    0.876
## ----------------------------------------------------------------
##
## - Curve
##
##
##  Step 30 : AIC = -45259.81
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ----------------------------------------------------------------
## Variable     DF       AIC        Sum Sq     RSS      R-Sq     Adj
## ----------------------------------------------------------------
## Dribbling     1    -45293.651     0.060    21.235    0.877
## Stamina       1    -45280.120     0.037    21.257    0.877
## Value         1    -45273.354     0.026    21.269    0.877
## RCM           1    -45263.301     0.009    21.285    0.877
## ShotPower     1    -45261.076     0.005    21.289    0.877
## Finishing     1    -45257.811     0.000    21.295    0.876
## ----------------------------------------------------------------
##
## - Dribbling
##
##
##  Step 31 : AIC = -45293.65
```

```
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ----------------------------------------------------------------
## Variable     DF       AIC        Sum Sq     RSS      R-Sq     Adj
## ----------------------------------------------------------------
## Stamina       1    -45324.234     0.054    21.181    0.877
## Value         1    -45307.394     0.026    21.209    0.877
## RCM           1    -45306.916     0.025    21.209    0.877
## ShotPower     1    -45298.039     0.011    21.224    0.877
## Finishing     1    -45291.652     0.000    21.235    0.877
## ----------------------------------------------------------------
##
## - Stamina
##
##
##  Step 32 : AIC = -45324.23
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special

##
## ----------------------------------------------------------------
## Variable     DF       AIC        Sum Sq     RSS      R-Sq     Adj
## ----------------------------------------------------------------
## RCM           1    -45354.920     0.054    21.126    0.877
## Value         1    -45336.779     0.024    21.156    0.877
## ShotPower     1    -45335.396     0.022    21.159    0.877
## Finishing     1    -45322.815     0.001    21.180    0.877
## ----------------------------------------------------------------
##
## - RCM
##
##
##  Step 33 : AIC = -45354.92
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ----------------------------------------------------------------
## Variable     DF       AIC        Sum Sq     RSS      R-Sq     Adj
## ----------------------------------------------------------------
## ShotPower     1    -45370.333     0.029    21.098    0.878
## Value         1    -45367.481     0.024    21.102    0.878
## Finishing     1    -45357.465     0.008    21.119    0.878
## ----------------------------------------------------------------
##
## - ShotPower
```

```
##
##
##  Step 34 : AIC = -45370.33
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special
##
## ----------------------------------------------------------------
## Variable     DF       AIC        Sum Sq     RSS      R-Sq     Adj
## ----------------------------------------------------------------
## Value         1    -45383.563     0.025    21.072    0.878
## Finishing     1    -45375.096     0.011    21.086    0.878
## ----------------------------------------------------------------
##
## - Value
##
##
##  Step 35 : AIC = -45383.56
##  Overall ~ Reactions + Wage + Composure + GKPositioning + Special

##
## ----------------------------------------------------------------
## Variable     DF       AIC        Sum Sq     RSS      R-Sq     Adj
## ----------------------------------------------------------------
## Finishing     1    -45388.089     0.011    21.062    0.878
## ----------------------------------------------------------------
##
## - Finishing
##
##
## Variables Entered:
##
## - Reactions
## - Wage
## - Composure
## - GKPositioning
## - Special
## - HeadingAccuracy
## - Skill.Moves
## - Positioning
## - BallControl
## - Weight
## - FKAccuracy
## - Balance
## - SprintSpeed
```

```
## - International.Reputation
## - ShortPassing
## - Vision
## - Interceptions
## - LCB
## - StandingTackle
## - LongShots
## - Aggression
## - Volleys
## - LongPassing
## - Height
## - Age
## - Crossing
## - Jumping
## - Penalties
## - Agility
## - Curve

## - Dribbling
## - Stamina
## - RCM
## - ShotPower
## - Value
## - Finishing
##
##
## Final Model Output
## ------------------
##
##                         Model Summary
## ----------------------------------------------------------------
## R                       0.937       RMSE                 0.041
## R-Squared               0.878       Coef. Var           -7.745
## Adj. R-Squared          0.877       MSE                  0.002
## Pred R-Squared          0.877       MAE                  0.032
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                  ANOVA
## -------------------------------------------------------------------
##                Sum of
```

```
##               Squares            DF    Mean Square       F
## ----------------------------------------------------------------------
## Regression    151.355            36          4.204     2536.57
## Residual       21.062         12707          0.002
## Total         172.416         12743
## ----------------------------------------------------------------------
##
##                                          Parameter Estimates
## ----------------------------------------------------------------------
##                    model      Beta    Std. Error    Std. Beta
## ----------------------------------------------------------------------
##              (Intercept)    -0.910         0.012
##                Reactions     0.281         0.006        0.287
##                     Wage     0.406         0.014        0.139
##                Composure     0.142         0.005        0.149
##            GKPositioning     0.129         0.009        0.212
##                  Special     0.966         0.039        1.391

##          HeadingAccuracy     0.068         0.005        0.112
##              Skill.Moves     0.094         0.004        0.152
##              Positioning    -0.135         0.006       -0.243
##              BallControl     0.105         0.009        0.165
##                   Weight     0.019         0.005        0.020
##                FKAccuracy    -0.053         0.005       -0.087
##                   Balance    -0.074         0.004       -0.112
##              SprintSpeed    -0.017         0.005       -0.025
## International.Reputation     0.063         0.005        0.054
##             ShortPassing     0.054         0.008        0.079
##                   Vision    -0.081         0.006       -0.117
##            Interceptions    -0.066         0.006       -0.133
##                      LCB    -0.051         0.017       -0.077
##           StandingTackle    -0.067         0.008       -0.137
##                LongShots    -0.087         0.006       -0.157
##               Aggression    -0.047         0.004       -0.082
##                  Volleys    -0.066         0.005       -0.115
##              LongPassing    -0.088         0.007       -0.137
##                   Height     0.012         0.002        0.025
##                      Age     0.023         0.003        0.032
##                 Crossing    -0.050         0.005       -0.089
##                  Jumping    -0.039         0.004       -0.050
##                Penalties    -0.051         0.005       -0.079
##                  Agility    -0.045         0.005       -0.070
##                    Curve    -0.043         0.005       -0.076
```

```
##                 Dribbling    -0.072          0.008       -0.124
##                   Stamina    -0.036          0.004       -0.058
##                       RCM     0.089          0.014        0.109
##                 ShotPower    -0.025          0.005       -0.039
##                     Value    -0.041          0.011       -0.015
##                 Finishing    -0.016          0.006       -0.028
## ----------------------------------------------------------------
```

```
##
##                                 Selection Summary
## ------------------------------------------------------------------------
## Variable                        AIC        Sum Sq      RSS       R
## ------------------------------------------------------------------------
## Reactions                   -34899.647    124.188    48.229    0.
## Wage                        -36992.285    131.498    40.919    0.
## Composure                   -38689.025    136.604    35.812    0.
## GKPositioning               -39708.277    139.362    33.054    0.
## Special                     -41097.419    142.780    29.636    0.
## HeadingAccuracy             -42242.820    145.332    27.084    0.
## Skill.Moves                 -42648.268    146.184    26.232    0.

## Positioning                 -43187.212    147.274    25.142    0.
## BallControl                 -43564.880    148.012    24.404    0.
## Weight                      -43863.173    148.581    23.836    0.
## FKAccuracy                  -43992.277    148.825    23.592    0.
## Balance                     -44120.985    149.065    23.351    0.
## SprintSpeed                 -44226.069    149.261    23.156    0.
## International.Reputation    -44343.403    149.477    22.940    0.
## ShortPassing                -44423.280    149.624    22.793    0.
## Vision                      -44498.700    149.762    22.655    0.
## Interceptions               -44549.554    149.855    22.561    0.
## LCB                         -44659.729    150.053    22.363    0.
## StandingTackle              -44727.718    150.176    22.241    0.
## LongShots                   -44802.620    150.309    22.107    0.
## Aggression                  -44849.583    150.394    22.022    0.
## Volleys                     -44903.459    150.490    21.926    0.
## LongPassing                 -44965.191    150.600    21.817    0.
## Height                      -45016.470    150.691    21.726    0.
## Age                         -45063.550    150.774    21.642    0.
## Crossing                    -45105.014    150.848    21.568    0.
## Jumping                     -45151.238    150.930    21.487    0.
```

```
## Penalties                -45188.120    150.995    21.422    0.
## Agility                  -45226.361    151.063    21.354    0.
## Curve                    -45259.807    151.122    21.295    0.
## Dribbling                -45293.651    151.182    21.235    0.
## Stamina                  -45324.234    151.236    21.181    0.
## RCM                      -45354.920    151.290    21.126    0.
## ShotPower                -45370.333    151.319    21.098    0.
## Value                    -45383.563    151.344    21.072    0.
## Finishing                -45388.089    151.355    21.062    0.
## ---------------------------------------------------------------
```

2.5.12 Model Output

Model Summary

- R 0.939
- RMSE 0.042
- R-Squared 0.881
- Coef. Var -7.842
- Adj. R-Squared 0.880
- MSE 0.002
- Pred R-Squared 0.880
- MAE 0.033

다중 회귀분석 최종모델 결과 해석

다중회귀분석으로 설계하고 학습 및 개선시킨 모델로 사전에 마련해둔 Test data를 예측해본 결과

Adjusted R-squared 값이 개선 후 모델에서 최종적으로 사용된 44개의 변수가 Overall을 88% 설명할 수 있음을 확인해, 설명력이 높은 모델을 만들어냈음을

알 수 있었다.

,br>

2.6 분류분석

UEFA의 club순위를 참고하여 임의의 파생변수로 생성한 범주형 구간 변수인 Club_grade를 label로 활용하여
선수들이 가진 stat과 feature에 따라 어떤 등급의 팀에 속해야 적절한지 예측하는 분류모델을 만들어 학습시키고, 이를 기반으로 30%의 test 데이터를 예측해 보고자 한다.

너무긴학습시간문제
rpart에 너무 많은 변수와 관측값을 가진 데이터를 학습시켜 불순도 연산에 의해 너무 긴 시간이 걸리는것으로 짐작된다, 첫 학습이 진행되지 않아 스케줄 상 더이상의 연구 진행이 어려울것이라 판단하여 해당 파트의 코드는 주석처리하고 추후 연구에서 개선해보고자 한다

```
#Decision Tree

#
# library(rpart)
#
# # 데이터 샘플링
#
# library(doBy)
# train = sampleBy(~club_grade,frac=0.7,data=b) #70%의 랜덤(train)데이
#
#
#
```

```
#
#
#
# # 70%에 포함된 train데이터의 행 번호를 추출하기 위한 사용자 함수 생성
#
# f1 <- function(x) {
#
#   as.numeric(strsplit(x, '\\.')[[1]][2])                          #
#
# }
#
# rn <- as.vector(sapply(rownames(train), f1))              # 70%에 포
#
# test <- b[-rn,]                                                    #
#
#
#
#
#
#
#
#
#
#
# train[10,]
# ####2. 모델 생성
#
# m <- rpart(club_grade~.,data = train)
#
# # 모델 학습에 필요한 데이터 셋(70%의 train 데이터)
#
#
#                       # 각 컬럼마다의 불순도 연산이 수행되므로 데이터가
#
#
#
#
#
#
#
#
#
#
#
```

```
# ###3. 모델평가
#
# # 3-1) 새로운 데이터셋(Test)에 대한 예측력 확인
#
# val_var <- predict()
#
# sum() / nrow() * 100
#
#
#
#
#
# val_var <- predict(m, newdata = train, type = 'class')
#
# sum(val_var == ) / nrow(train) * 100
#
#
#
#
#
# ##4. 모델을 통한 예측
#
```

3. 결론

의의

해당 보고서에선 손흥민 선수에 의해 점화된 해외축구에 대한 전국적인 관심에 힘입어, 데이터가 생소한 보통의 축구팬들에게도 다소 접근하기 쉬운 기초통계량 기반의 EDA분석과 시각화로 시작하여 유사도분석, 분류분석, 군집분석,다중회귀분석까지 다소 심화된 개념의 분석 및 모델링과 결과들을 통해 저번 중간보

고서의 연구에 이어 데이터를 통해 스포츠를 더욱 깊고 폭넓게 즐기고 유용하게 분석해 볼 수 있다는 시사점을 다시 한번 강조할 수 있었다.

결과정리

2.2 EDA 분석결과

- 각 리그 선수들의 평균 연령은 25세로 형성돼있다.
- 프리미어리그 내 등번호 7번의 윙어인 손흥민과 라힘 스털링 선수를 비교해 본 결과 두 선수는 스탯의 유사한 분포를 가지고 있다.
- 토트넘에서 가장 높은 potential과 Overall을 가진 선수는 Harry Kane 선수이다.
- 나이에 따른 소속 리그별 평균 잠재력 및 통합스탯을 분석해 본 결과 약 25세부터 선수들의 overall과 potential이 일치하여 동일선을 그리기 시작한다.

2.3 선수 유사도 분석결과

***

유클라디안 거리계산법을 활용하여 손흥민,Harry Kane선수와 유사한 선수들을 알아본 결과

- 손흥민 선수는 뎀벨레,므키타리안,래시포드,마네,라카제트 등 상대 수비진을 침투하는 플레이방식의 선수들과 유사하다는 결과를 얻을 수 있었다.

* 반면 해리케인 선수는 레반도프스키,루카쿠,피르미누,포그바 등 득점능력 뿐

아니라 패스 등 전체적으로 높은 밸런스를 가진 선수들과 유사하다고 판단되어 완성형 공격수라는 별명에 어울리는 결과를 얻을 수 있었다.

2.5. 다중회귀 분석 결과

FIFA선수의 인구통계학적 Feature와 경기중에 관측되는 각종 선수stat이 Overall을 얼마나 설명할 수 있을 지 다중회귀분석으로 규명해보았다.

최종 모델 결과

* R 0.939
* RMSE 0.042
* R-Squared 0.881
* Coef. Var -7.842
* Adj. R-Squared 0.880
* MSE 0.002
* Pred R-Squared 0.880
* MAE 0.033

***Adjusted R-squared* 값이 개선 후 모델에서 최종적으로 사용된 *44*개의 변수가 *Overall*을 *88%* 설명할 수 있음을 확인해, 설명력이 높은 모델을 만들어냈음을 알 수 있었다.**

한계점

다중회귀분석의 종속변수로 사용한 Overall은 본래 그 의미가 선수의 경기내 stat을 전체적으로 반영한 점수이므로 당연히 높은 설명력을 가질 수 밖에 없었다.

또한 너무 큰 dataset을 사용하여 분류모델을 학습시키려 할 경우 너무 많은 불순도연산으로 매우 긴 연산기간을 야기하는 것을 고려하지 못하여 원활히 연구를 진행할 수 없었다. 해당 부분의 문제는 추후 머신러닝 및 캡스톤 디자인 과목에서 보완법을 알아보고 추가 연구를 진행해보고자 한다.