Player Scoring Analysis

#导入包与数据#

library(ggplot2)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(DMwR)

## Loading required package: lattice

## Loading required package: grid

## Registered S3 method overwritten by 'xts':  
## method from  
## as.zoo.xts zoo

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo

player\_scoring <- read.csv("/Users/sun/Desktop/NBA/Seasons\_Stats.csv",header=T, na.strings = "NA")  
player <- read.csv("/Users/sun/Desktop/NBA/player\_data.csv",header=T,na.strings = "NA")  
#用每个变量的中位数进行缺失值填充  
for(i in 7:53){  
 player\_scoring[which(is.na(player\_scoring[,i])),i]<-median(player\_scoring[,i],na.rm = T)  
}

summary(player\_scoring)

## X Year Player Pos   
## Min. : 0 Min. :1950 : 67 PF :4966   
## 1st Qu.: 6172 1st Qu.:1981 Eddie Johnson : 33 SG :4811   
## Median :12345 Median :1996 Mike Dunleavy : 32 C :4759   
## Mean :12345 Mean :1993 Gerald Henderson: 29 SF :4699   
## 3rd Qu.:18518 3rd Qu.:2007 Nazr Mohammed : 28 PG :4648   
## Max. :24690 Max. :2017 Kevin Willis : 27 G : 139   
## NA's :67 (Other) :24475 (Other): 669   
## Age Tm G GS MP   
## Min. :18.00 TOT : 2123 Min. : 1.00 Min. : 0.00 Min. : 0   
## 1st Qu.:24.00 NYK : 1043 1st Qu.:27.00 1st Qu.: 1.00 1st Qu.: 354   
## Median :26.00 BOS : 998 Median :58.00 Median : 8.00 Median :1053   
## Mean :26.66 DET : 917 Mean :50.86 Mean :19.51 Mean :1206   
## 3rd Qu.:29.00 PHI : 871 3rd Qu.:75.00 3rd Qu.:26.00 3rd Qu.:1948   
## Max. :44.00 LAL : 834 Max. :88.00 Max. :83.00 Max. :3882   
## NA's :75 (Other):17905   
## PER TS. X3PAr FTr   
## Min. :-90.60 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.: 9.90 1st Qu.:0.4580 1st Qu.:0.0120 1st Qu.:0.2080   
## Median : 12.70 Median :0.5060 Median :0.0640 Median :0.2960   
## Mean : 12.48 Mean :0.4931 Mean :0.1362 Mean :0.3253   
## 3rd Qu.: 15.50 3rd Qu.:0.5430 3rd Qu.:0.2150 3rd Qu.:0.4000   
## Max. :129.10 Max. :1.1360 Max. :1.0000 Max. :6.0000   
##   
## ORB. DRB. TRB. AST.   
## Min. : 0.000 Min. : 0.00 Min. : 0.000 Min. : 0.00   
## 1st Qu.: 3.000 1st Qu.: 9.40 1st Qu.: 6.300 1st Qu.: 6.90   
## Median : 5.400 Median : 12.70 Median : 9.200 Median : 10.50   
## Mean : 6.058 Mean : 13.55 Mean : 9.855 Mean : 12.79   
## 3rd Qu.: 8.300 3rd Qu.: 17.00 3rd Qu.: 12.900 3rd Qu.: 16.50   
## Max. :100.000 Max. :100.00 Max. :100.000 Max. :100.00   
##   
## STL. BLK. TOV. USG.   
## Min. : 0.000 Min. : 0.00 Min. : 0.0 Min. : 0.00   
## 1st Qu.: 1.200 1st Qu.: 0.40 1st Qu.: 12.1 1st Qu.: 16.30   
## Median : 1.500 Median : 0.90 Median : 14.2 Median : 18.60   
## Mean : 1.625 Mean : 1.33 Mean : 14.9 Mean : 18.84   
## 3rd Qu.: 2.000 3rd Qu.: 1.60 3rd Qu.: 16.6 3rd Qu.: 21.10   
## Max. :24.200 Max. :77.80 Max. :100.0 Max. :100.00   
##   
## blanl OWS DWS WS   
## Mode:logical Min. :-5.100 Min. :-1.000 Min. :-2.800   
## NA's:24691 1st Qu.:-0.100 1st Qu.: 0.200 1st Qu.: 0.200   
## Median : 0.400 Median : 0.800 Median : 1.400   
## Mean : 1.254 Mean : 1.226 Mean : 2.481   
## 3rd Qu.: 1.900 3rd Qu.: 1.800 3rd Qu.: 3.800   
## Max. :18.300 Max. :16.000 Max. :25.400   
##   
## WS.48 blank2 OBPM DBPM   
## Min. :-2.51900 Mode:logical Min. :-73.800 Min. :-30.4000   
## 1st Qu.: 0.03200 NA's:24691 1st Qu.: -3.000 1st Qu.: -1.5000   
## Median : 0.07500 Median : -1.500 Median : -0.5000   
## Mean : 0.06524 Mean : -1.734 Mean : -0.5409   
## 3rd Qu.: 0.11300 3rd Qu.: 0.000 3rd Qu.: 0.5000   
## Max. : 2.12300 Max. : 47.800 Max. : 46.8000   
##   
## BPM VORP FG FGA   
## Min. :-86.700 Min. :-2.6000 Min. : 0.0 Min. : 0.0   
## 1st Qu.: -3.600 1st Qu.:-0.1000 1st Qu.: 41.0 1st Qu.: 99.0   
## Median : -1.800 Median : 0.0000 Median : 141.0 Median : 321.0   
## Mean : -2.244 Mean : 0.4716 Mean : 195.2 Mean : 430.3   
## 3rd Qu.: -0.100 3rd Qu.: 0.6000 3rd Qu.: 298.5 3rd Qu.: 660.0   
## Max. : 36.200 Max. :12.4000 Max. :1597.0 Max. :3159.0   
##   
## FG. X3P X3PA X3P.   
## Min. :0.0000 Min. : 0.0 Min. : 0.00 Min. :0.000   
## 1st Qu.:0.3930 1st Qu.: 0.0 1st Qu.: 2.00 1st Qu.:0.250   
## Median :0.4390 Median : 2.0 Median : 11.00 Median :0.292   
## Mean :0.4309 Mean : 17.5 Mean : 51.32 Mean :0.265   
## 3rd Qu.:0.4800 3rd Qu.: 15.0 3rd Qu.: 48.00 3rd Qu.:0.327   
## Max. :1.0000 Max. :402.0 Max. :886.00 Max. :1.000   
##   
## X2P X2PA X2P. eFG.   
## Min. : 0.0 Min. : 0.0 Min. :0.0000 Min. :0.0000   
## 1st Qu.: 35.0 1st Qu.: 82.0 1st Qu.:0.4080 1st Qu.:0.4140   
## Median : 122.0 Median : 270.0 Median :0.4560 Median :0.4630   
## Mean : 178.1 Mean : 381.5 Mean :0.4454 Mean :0.4507   
## 3rd Qu.: 267.5 3rd Qu.: 578.0 3rd Qu.:0.4960 3rd Qu.:0.5010   
## Max. :1597.0 Max. :3159.0 Max. :1.0000 Max. :1.5000   
##   
## FT FTA FT. ORB   
## Min. : 0.0 Min. : 0.0 Min. :0.0000 Min. : 0.00   
## 1st Qu.: 18.0 1st Qu.: 27.0 1st Qu.:0.6630 1st Qu.: 16.00   
## Median : 63.0 Median : 88.0 Median :0.7430 Median : 38.00   
## Mean :102.3 Mean : 136.6 Mean :0.7202 Mean : 58.37   
## 3rd Qu.:148.0 3rd Qu.: 200.0 3rd Qu.:0.8050 3rd Qu.: 77.00   
## Max. :840.0 Max. :1363.0 Max. :1.0000 Max. :587.00   
##   
## DRB TRB AST STL   
## Min. : 0.0 Min. : 0.0 Min. : 0.0 Min. : 0.00   
## 1st Qu.: 44.0 1st Qu.: 52.5 1st Qu.: 19.0 1st Qu.: 12.00   
## Median : 106.0 Median : 159.0 Median : 68.0 Median : 29.00   
## Mean : 140.7 Mean : 223.6 Mean : 114.7 Mean : 38.18   
## 3rd Qu.: 188.0 3rd Qu.: 318.0 3rd Qu.: 160.0 3rd Qu.: 53.00   
## Max. :1111.0 Max. :2149.0 Max. :1164.0 Max. :301.00   
##   
## BLK TOV PF PTS   
## Min. : 0.00 Min. : 0.00 Min. : 0.0 Min. : 0.0   
## 1st Qu.: 4.00 1st Qu.: 26.00 1st Qu.: 39.0 1st Qu.: 107.0   
## Median : 11.00 Median : 55.00 Median :109.0 Median : 364.0   
## Mean : 22.35 Mean : 70.07 Mean :116.3 Mean : 509.7   
## 3rd Qu.: 24.00 3rd Qu.: 95.00 3rd Qu.:182.0 3rd Qu.: 777.0   
## Max. :456.00 Max. :464.00 Max. :386.0 Max. :4029.0   
##

PTS <- as.data.frame(aggregate(player\_scoring$PTS,by=list(player\_scoring$Player),FUN=sum))  
colnames(PTS) <- c("Name","pts")  
PTS <- PTS[order(PTS$pts,decreasing=T),]

#1.变量解释和选择# **变量名与含义** 本数据统计了NBA67个赛季全体个人数据，包括得分，助攻，篮板等数据。下边介绍各项指标的含义即对应的变量名。 Player Efficiency Rating(PER):球员效率值，他是评判球员贡献的指标，PER在综合评判球员的成绩如投篮得分，发球，三分球，助攻，篮板，盖帽和抢断，以及负面结果，如投篮失误和个人犯规后统计该球员的成绩。得到其最重效率值。 Minutes Played(MP):每场比赛的上场时长 Total Rebounds/Percentage(TRB)(TRB.):总篮板球 Defensive Rebound/Rate(DRB)(DRB.):防守篮板，求全转换到防守方手中，进攻方球员投球未进，防守球员拿到球权，抢到球的球员记一个防守篮板球。 Offensive Rebound/Rate(ORB)(ORB.):进攻篮板,投篮不中，但本队队员将篮板抢回，重新获得进攻机会。抢到球的球员记一个进攻篮板球。

Field Goals made/attempted/percentage:(FG)(FGA)(FG.):投篮命中，除了罚球以外的投篮命中得分。 3-Point Field Goals/Attempts/Percentage(X3P)(3PA)(3P.):三分投篮命中 3-Point Attempt Rate(3PAr) 2-Point Field Goals/Attempts/Percentage(X2P)(X2PA)(X2P.):二分投篮命中 Effective Field Goal(eFG.):有效的投篮得分率 Free Throws/Attempts/Percentage(FT)(FTA)(FT.):罚球 True Shooting Rate(TS.) Assists/Assists Percentage(AST)(AST.):助攻 Steals/Steal Percentage(STL)(STL.): Blocks/Blocks Percentage(BLK)(BLK.): Turnovers/Turnover Percentage(TOV)(TOV.):犯规 Personal Fouls(PF): Points(PTS): Box Plus/Minus(BPM): Defensive Box Plus/Minus(DBPM): Offensive Box Plus/Minus(OBPM): Value Over Replacement(VORP): Win Shares Per 48 Minutes(WS/48): Win Shares(WS): Offensive Win Shares(OWS): Usage Percentage(USG.)

#得分最高的前20个球员进攻篮板效率和防守篮板效率的分布图

theme\_set(theme\_bw())   
ORB\_DRB <-aggregate(player\_scoring[,c("PTS","ORB.","DRB.")],by=list(player\_scoring$Player),FUN=mean)  
ORB\_DRB <- ORB\_DRB[order(ORB\_DRB$PTS,decreasing=T),]  
colnames(ORB\_DRB) <- c("Player","PTS","ORBR","DRBR")  
ggplot(ORB\_DRB[1:20,],aes(x=ORBR,y=DRBR))+  
 geom\_point(aes(color=Player,size=PTS))+ theme(legend.position="bottom")+  
 labs(title="DRBR and ORBR",caption = "The size of points represent one's PTS")

大多数得分高的球星的

#构造控球变量： 个人单个赛季控球时长：是当一个队拿到控球权开始，到失去控球权结束的。每场比赛两队的控球机会是接近相等的。衡量两队的控球机会可以有效的评估个人成绩和队伍成绩。每队球员为了赢得比赛，需要在控球的时候拿到更多的分数。我们根据上述变量构造控球变量。 第一种公式为：公式中每次犯规，投篮命中，和罚球结束一次控球。 其中参数是结束控球的罚球所占比率，参数是一个0-1之间的参数。 通常将参数设置为 第二种公式，我们不规定$\alpha和\lambda$，而是直接用64季比赛结果与具体的控球时间做线性回归， 资料表明，根据OLS回归预测的模型参数为：

player\_scoring$POSS <- 3.2258+0.964\*player\_scoring$FGA-0.3452\*(player\_scoring$FGA-player\_scoring$FG)+0.4637\*player\_scoring$FTA-0.2073\*(player\_scoring$FT-player\_scoring$FT)-0.6227\*player\_scoring$ORB+0.3643\*player\_scoring$DRB+0.9767\*player\_scoring$TOV

#数据可视化 1.按照PTS统计出历史上得分最高的球员，

player\_scoring[which(player\_scoring$Player=="Kareem Abdul-Jabbar\*"),c("Year","Player","Pos","Age","Tm")]

## Year Player Pos Age Tm  
## 2869 1970 Kareem Abdul-Jabbar\* C 22 MIL  
## 3071 1971 Kareem Abdul-Jabbar\* C 23 MIL  
## 3317 1972 Kareem Abdul-Jabbar\* C 24 MIL  
## 3583 1973 Kareem Abdul-Jabbar\* C 25 MIL  
## 3853 1974 Kareem Abdul-Jabbar\* C 26 MIL  
## 4099 1975 Kareem Abdul-Jabbar\* C 27 MIL  
## 4376 1976 Kareem Abdul-Jabbar\* C 28 LAL  
## 4651 1977 Kareem Abdul-Jabbar\* C 29 LAL  
## 5011 1978 Kareem Abdul-Jabbar\* C 30 LAL  
## 5383 1979 Kareem Abdul-Jabbar\* C 31 LAL  
## 5728 1980 Kareem Abdul-Jabbar\* C 32 LAL  
## 6086 1981 Kareem Abdul-Jabbar\* C 33 LAL  
## 6450 1982 Kareem Abdul-Jabbar\* C 34 LAL  
## 6824 1983 Kareem Abdul-Jabbar\* C 35 LAL  
## 7216 1984 Kareem Abdul-Jabbar\* C 36 LAL  
## 7560 1985 Kareem Abdul-Jabbar\* C 37 LAL  
## 7923 1986 Kareem Abdul-Jabbar\* C 38 LAL  
## 8303 1987 Kareem Abdul-Jabbar\* C 39 LAL  
## 8682 1988 Kareem Abdul-Jabbar\* C 40 LAL  
## 9109 1989 Kareem Abdul-Jabbar\* C 41 LAL

卡利姆·阿布杜尔·贾巴尔，1947年出生的职业运动员，司职中锋，绰号“天勾”，在1969年NBA选秀中于第1轮第1位以状元秀的身份被密沃尔基雄鹿队选中，新秀赛季当选NBA年度最佳新秀。1970-71赛季贾巴尔场均31.7），获得常规赛MVP以及NBA得分王称号，并率队获得NBA总冠军。1976年转投洛杉矶湖人队，与有“魔术师”之称的埃尔文·约翰逊成为湖人队的两大支柱，率领球队五次夺得NBA总冠军。

player\_scoring[which(player\_scoring$Player=="Karl Malone\*"),c("Year","Player","Pos","Age","Tm")]

## Year Player Pos Age Tm  
## 8123 1986 Karl Malone\* PF 22 UTA  
## 8495 1987 Karl Malone\* PF 23 UTA  
## 8895 1988 Karl Malone\* PF 24 UTA  
## 9340 1989 Karl Malone\* PF 25 UTA  
## 9796 1990 Karl Malone\* PF 26 UTA  
## 10252 1991 Karl Malone\* PF 27 UTA  
## 10705 1992 Karl Malone\* PF 28 UTA  
## 11151 1993 Karl Malone\* PF 29 UTA  
## 11642 1994 Karl Malone\* PF 30 UTA  
## 12090 1995 Karl Malone\* PF 31 UTA  
## 12574 1996 Karl Malone\* PF 32 UTA  
## 13144 1997 Karl Malone\* PF 33 UTA  
## 13690 1998 Karl Malone\* PF 34 UTA  
## 14237 1999 Karl Malone\* PF 35 UTA  
## 14754 2000 Karl Malone\* PF 36 UTA  
## 15253 2001 Karl Malone\* PF 37 UTA  
## 15783 2002 Karl Malone\* PF 38 UTA  
## 16268 2003 Karl Malone\* PF 39 UTA  
## 16801 2004 Karl Malone\* PF 40 LAL

卡尔·马龙（Karl Malone），司职大前锋，绰号“邮差”（The Mailman） 在1985年NBA选秀中，卡尔·马龙在首轮第13顺位被犹他爵士队选中，职业生涯的前18个赛季在犹他爵士队度过，并与约翰·斯托克顿组成的“黑白双煞”让爵士队成为一支劲旅，两人带领爵士队两次打入NBA总决赛，但是都未能夺得NBA总冠军，生涯最后一年加盟洛杉矶湖人队

player\_scoring[which(player\_scoring$Player=="Wilt Chamberlain\*"),c("Year","Player","Pos","Age","Tm")]

## Year Player Pos Age Tm  
## 1474 1960 Wilt Chamberlain\* C 23 PHW  
## 1594 1961 Wilt Chamberlain\* C 24 PHW  
## 1707 1962 Wilt Chamberlain\* C 25 PHW  
## 1828 1963 Wilt Chamberlain\* C 26 SFW  
## 1963 1964 Wilt Chamberlain\* C 27 SFW  
## 2100 1965 Wilt Chamberlain\* C 28 TOT  
## 2101 1965 Wilt Chamberlain\* C 28 SFW  
## 2102 1965 Wilt Chamberlain\* C 28 PHI  
## 2240 1966 Wilt Chamberlain\* C 29 PHI  
## 2367 1967 Wilt Chamberlain\* C 30 PHI  
## 2508 1968 Wilt Chamberlain\* C 31 PHI  
## 2691 1969 Wilt Chamberlain\* C 32 LAL  
## 2894 1970 Wilt Chamberlain\* C 33 LAL  
## 3101 1971 Wilt Chamberlain\* C 34 LAL  
## 3346 1972 Wilt Chamberlain\* C 35 LAL  
## 3616 1973 Wilt Chamberlain\* C 36 LAL

威尔特·张伯伦,司职中锋。曾效力于NBA费城/旧金山勇士队（现金州勇士队）、费城76人队以及洛杉矶湖人队；毕业于堪萨斯大学的他在NBA打球之前曾在哈林篮球队效力，之后被认为是NBA历史上具有统治力的球员之一。2次夺得NBA总冠军），4次荣膺NBA常规赛MVP，13次入选NBA全明星阵容，10次入选NBA最佳阵容，连续2次入选NBA最佳防守阵容一阵，连续7次荣膺NBA得分王，11次荣膺NBA篮板王，1968年荣膺NBA助攻王。

player\_scoring[which(player\_scoring$Player=="Kobe Bryant"),c("Year","Player","Pos","Age","Tm")]

## Year Player Pos Age Tm  
## 12901 1997 Kobe Bryant SG 18 LAL  
## 13480 1998 Kobe Bryant SG 19 LAL  
## 14022 1999 Kobe Bryant SG 20 LAL  
## 14538 2000 Kobe Bryant SG 21 LAL  
## 15029 2001 Kobe Bryant SG 22 LAL  
## 15579 2002 Kobe Bryant SG 23 LAL  
## 16071 2003 Kobe Bryant SG 24 LAL  
## 16577 2004 Kobe Bryant SG 25 LAL  
## 17160 2005 Kobe Bryant SG 26 LAL  
## 17743 2006 Kobe Bryant SG 27 LAL  
## 18296 2007 Kobe Bryant SG 28 LAL  
## 18826 2008 Kobe Bryant SG 29 LAL  
## 19429 2009 Kobe Bryant SG 30 LAL  
## 20007 2010 Kobe Bryant SG 31 LAL  
## 20601 2011 Kobe Bryant SG 32 LAL  
## 21197 2012 Kobe Bryant SG 33 LAL  
## 21758 2013 Kobe Bryant SG 34 LAL  
## 22349 2014 Kobe Bryant SG 35 LAL  
## 22945 2015 Kobe Bryant SG 36 LAL  
## 23587 2016 Kobe Bryant SF 37 LAL

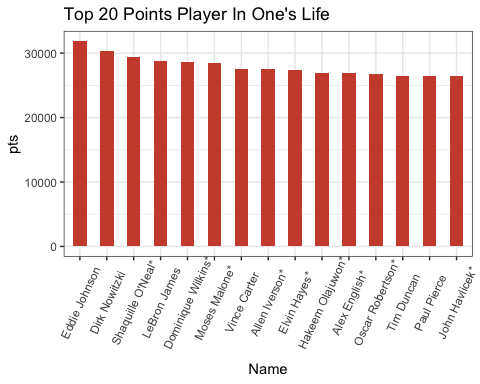
科比·比恩·布莱恩特,司职得分后卫/小前锋（锋卫摇摆人），绰号“黑曼巴”/“小飞侠”。 科比是前NBA球员乔·布莱恩特的儿子。在1996年NBA选秀中，科比于第1轮第13位被夏洛特黄蜂队选中，后被交易至洛杉矶湖人队，整个NBA生涯（1996年-2016年）都效力于洛杉矶湖人队。 科比的职业生涯随湖人队5夺NBA总冠军；荣膺1次常规赛MVP（2007-08赛季），2次总决赛MVP，4次全明星赛MVP；共18次入选NBA全明星阵容，15次入选NBA最佳阵容，12次入选NBA最佳防守阵容。

player\_scoring[which(player\_scoring$Player=="Michael Jordan\*"),c("Year","Player","Pos","Age","Tm")]

## Year Player Pos Age Tm  
## 7717 1985 Michael Jordan\* SG 21 CHI  
## 8095 1986 Michael Jordan\* SG 22 CHI  
## 8470 1987 Michael Jordan\* SG 23 CHI  
## 8870 1988 Michael Jordan\* SG 24 CHI  
## 9298 1989 Michael Jordan\* SG 25 CHI  
## 9748 1990 Michael Jordan\* SG 26 CHI  
## 10213 1991 Michael Jordan\* SG 27 CHI  
## 10667 1992 Michael Jordan\* SG 28 CHI  
## 11107 1993 Michael Jordan\* SG 29 CHI  
## 12049 1995 Michael Jordan\* SG 31 CHI  
## 12529 1996 Michael Jordan\* SG 32 CHI  
## 13097 1997 Michael Jordan\* SG 33 CHI  
## 13655 1998 Michael Jordan\* SG 34 CHI  
## 15752 2002 Michael Jordan\* SF 38 WAS  
## 16241 2003 Michael Jordan\* SF 39 WAS

迈克尔·乔丹（Michael Jordan）司职得分后卫，绰号“飞人”（Air Jordan） [1] 。 迈克尔·乔丹在在1984年NBA选秀中于第1轮第3位被芝加哥公牛队选中，职业生涯曾效力于芝加哥公牛队以及华盛顿奇才队，新秀赛季当选NBA年度最佳新秀。1986-87赛季，乔丹场均得到37.1分，首次获得NBA得分王称号。1991-93赛季，乔丹连续2次荣膺常规赛MVP）和3次总决赛MVP

PTS$Name <- factor(PTS$Name,levels=PTS$Name)  
ggplot(PTS[6:20,],aes(x=Name,y=pts))+  
 geom\_bar(stat="identity", width=.5, fill="tomato3") +  
  
labs(title="Top 20 Points Player In One's Life") +  
  
theme(axis.text.x = element\_text(angle=65, vjust=0.6))

 除上述介绍的5名著名球星，在过往的62个NBA赛季中，其他获得个人最好成绩的球员如上。