Capstone Data Project

Nave Cohen

2022-03-31

library(dplyr)

## Warning: package 'dplyr' was built under R version 4.0.5

##   
## 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(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.0.5

Batting = read.csv("Batting.csv")  
print(head(Batting))

## playerID yearID stint teamID lgID G G\_batting AB R H X2B X3B HR RBI SB CS  
## 1 aardsda01 2004 1 SFN NL 11 11 0 0 0 0 0 0 0 0 0  
## 2 aardsda01 2006 1 CHN NL 45 43 2 0 0 0 0 0 0 0 0  
## 3 aardsda01 2007 1 CHA AL 25 2 0 0 0 0 0 0 0 0 0  
## 4 aardsda01 2008 1 BOS AL 47 5 1 0 0 0 0 0 0 0 0  
## 5 aardsda01 2009 1 SEA AL 73 3 0 0 0 0 0 0 0 0 0  
## 6 aardsda01 2010 1 SEA AL 53 4 0 0 0 0 0 0 0 0 0  
## BB SO IBB HBP SH SF GIDP G\_old  
## 1 0 0 0 0 0 0 0 11  
## 2 0 0 0 0 1 0 0 45  
## 3 0 0 0 0 0 0 0 2  
## 4 0 1 0 0 0 0 0 5  
## 5 0 0 0 0 0 0 0 NA  
## 6 0 0 0 0 0 0 0 NA

print(str(Batting))

## 'data.frame': 97889 obs. of 24 variables:  
## $ playerID : chr "aardsda01" "aardsda01" "aardsda01" "aardsda01" ...  
## $ yearID : int 2004 2006 2007 2008 2009 2010 2012 1954 1955 1956 ...  
## $ stint : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ teamID : chr "SFN" "CHN" "CHA" "BOS" ...  
## $ lgID : chr "NL" "NL" "AL" "AL" ...  
## $ G : int 11 45 25 47 73 53 1 122 153 153 ...  
## $ G\_batting: int 11 43 2 5 3 4 NA 122 153 153 ...  
## $ AB : int 0 2 0 1 0 0 NA 468 602 609 ...  
## $ R : int 0 0 0 0 0 0 NA 58 105 106 ...  
## $ H : int 0 0 0 0 0 0 NA 131 189 200 ...  
## $ X2B : int 0 0 0 0 0 0 NA 27 37 34 ...  
## $ X3B : int 0 0 0 0 0 0 NA 6 9 14 ...  
## $ HR : int 0 0 0 0 0 0 NA 13 27 26 ...  
## $ RBI : int 0 0 0 0 0 0 NA 69 106 92 ...  
## $ SB : int 0 0 0 0 0 0 NA 2 3 2 ...  
## $ CS : int 0 0 0 0 0 0 NA 2 1 4 ...  
## $ BB : int 0 0 0 0 0 0 NA 28 49 37 ...  
## $ SO : int 0 0 0 1 0 0 NA 39 61 54 ...  
## $ IBB : int 0 0 0 0 0 0 NA NA 5 6 ...  
## $ HBP : int 0 0 0 0 0 0 NA 3 3 2 ...  
## $ SH : int 0 1 0 0 0 0 NA 6 7 5 ...  
## $ SF : int 0 0 0 0 0 0 NA 4 4 7 ...  
## $ GIDP : int 0 0 0 0 0 0 NA 13 20 21 ...  
## $ G\_old : int 11 45 2 5 NA NA NA 122 153 153 ...  
## NULL

# Adding statistics that where used in Moneyball  
# Batting Average (BA)  
# AVG = H (Hits) / AB (At Bats)  
Batting$BA =round((Batting$H / Batting$AB),3)   
print(tail(Batting$BA))

## [1] 0.000 0.123 0.275 0.147 0.275 0.214

# On Base Percentage (OBP)  
Batting$OBP = round((Batting$H + Batting$BB + Batting$HBP)/ (Batting$AB + Batting$BB + Batting$HBP + Batting$SF),3)  
print(tail(Batting$OBP))

## [1] 0.000 0.134 0.344 0.147 0.354 0.290

# Slugging Percentage (SLG)  
Batting$B1 = Batting$H - Batting$X2B - Batting$X3B - Batting$HR  
Batting$SLG = round((Batting$B1 + (2\*Batting$X2B) + (3\*Batting$X3B) + (4\*Batting$HR)) / Batting$AB,3)  
print(tail(Batting$SLG))

## [1] 0.000 0.138 0.465 0.147 0.402 0.329

print(str(Batting))

## 'data.frame': 97889 obs. of 28 variables:  
## $ playerID : chr "aardsda01" "aardsda01" "aardsda01" "aardsda01" ...  
## $ yearID : int 2004 2006 2007 2008 2009 2010 2012 1954 1955 1956 ...  
## $ stint : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ teamID : chr "SFN" "CHN" "CHA" "BOS" ...  
## $ lgID : chr "NL" "NL" "AL" "AL" ...  
## $ G : int 11 45 25 47 73 53 1 122 153 153 ...  
## $ G\_batting: int 11 43 2 5 3 4 NA 122 153 153 ...  
## $ AB : int 0 2 0 1 0 0 NA 468 602 609 ...  
## $ R : int 0 0 0 0 0 0 NA 58 105 106 ...  
## $ H : int 0 0 0 0 0 0 NA 131 189 200 ...  
## $ X2B : int 0 0 0 0 0 0 NA 27 37 34 ...  
## $ X3B : int 0 0 0 0 0 0 NA 6 9 14 ...  
## $ HR : int 0 0 0 0 0 0 NA 13 27 26 ...  
## $ RBI : int 0 0 0 0 0 0 NA 69 106 92 ...  
## $ SB : int 0 0 0 0 0 0 NA 2 3 2 ...  
## $ CS : int 0 0 0 0 0 0 NA 2 1 4 ...  
## $ BB : int 0 0 0 0 0 0 NA 28 49 37 ...  
## $ SO : int 0 0 0 1 0 0 NA 39 61 54 ...  
## $ IBB : int 0 0 0 0 0 0 NA NA 5 6 ...  
## $ HBP : int 0 0 0 0 0 0 NA 3 3 2 ...  
## $ SH : int 0 1 0 0 0 0 NA 6 7 5 ...  
## $ SF : int 0 0 0 0 0 0 NA 4 4 7 ...  
## $ GIDP : int 0 0 0 0 0 0 NA 13 20 21 ...  
## $ G\_old : int 11 45 2 5 NA NA NA 122 153 153 ...  
## $ BA : num NaN 0 NaN 0 NaN NaN NA 0.28 0.314 0.328 ...  
## $ OBP : num NaN 0 NaN 0 NaN NaN NA 0.322 0.366 0.365 ...  
## $ B1 : int 0 0 0 0 0 0 NA 85 116 126 ...  
## $ SLG : num NaN 0 NaN 0 NaN NaN NA 0.447 0.54 0.558 ...  
## NULL

# Merging Salary Data with Batting Data  
Salaries = read.csv("Salaries.csv")  
print(summary(Batting))

## playerID yearID stint teamID   
## Length:97889 Min. :1871 Min. :1.000 Length:97889   
## Class :character 1st Qu.:1931 1st Qu.:1.000 Class :character   
## Mode :character Median :1970 Median :1.000 Mode :character   
## Mean :1962 Mean :1.077   
## 3rd Qu.:1995 3rd Qu.:1.000   
## Max. :2013 Max. :5.000   
##   
## lgID G G\_batting AB   
## Length:97889 Min. : 1.00 Min. : 0.00 Min. : 0.0   
## Class :character 1st Qu.: 13.00 1st Qu.: 7.00 1st Qu.: 9.0   
## Mode :character Median : 35.00 Median : 32.00 Median : 61.0   
## Mean : 51.65 Mean : 49.13 Mean :154.1   
## 3rd Qu.: 81.00 3rd Qu.: 81.00 3rd Qu.:260.0   
## Max. :165.00 Max. :165.00 Max. :716.0   
## NA's :1406 NA's :6413   
## R H X2B X3B   
## Min. : 0.00 Min. : 0.00 Min. : 0.0 Min. : 0.000   
## 1st Qu.: 0.00 1st Qu.: 1.00 1st Qu.: 0.0 1st Qu.: 0.000   
## Median : 5.00 Median : 12.00 Median : 2.0 Median : 0.000   
## Mean : 20.47 Mean : 40.37 Mean : 6.8 Mean : 1.424   
## 3rd Qu.: 31.00 3rd Qu.: 66.00 3rd Qu.:10.0 3rd Qu.: 2.000   
## Max. :192.00 Max. :262.00 Max. :67.0 Max. :36.000   
## NA's :6413 NA's :6413 NA's :6413 NA's :6413   
## HR RBI SB CS   
## Min. : 0.000 Min. : 0.00 Min. : 0.000 Min. : 0.000   
## 1st Qu.: 0.000 1st Qu.: 0.00 1st Qu.: 0.000 1st Qu.: 0.000   
## Median : 0.000 Median : 5.00 Median : 0.000 Median : 0.000   
## Mean : 3.002 Mean : 18.47 Mean : 3.265 Mean : 1.385   
## 3rd Qu.: 3.000 3rd Qu.: 28.00 3rd Qu.: 2.000 3rd Qu.: 1.000   
## Max. :73.000 Max. :191.00 Max. :138.000 Max. :42.000   
## NA's :6413 NA's :6837 NA's :7713 NA's :29867   
## BB SO IBB HBP   
## Min. : 0.00 Min. : 0.00 Min. : 0.00 Min. : 0.000   
## 1st Qu.: 0.00 1st Qu.: 2.00 1st Qu.: 0.00 1st Qu.: 0.000   
## Median : 4.00 Median : 11.00 Median : 0.00 Median : 0.000   
## Mean : 14.21 Mean : 21.95 Mean : 1.28 Mean : 1.136   
## 3rd Qu.: 21.00 3rd Qu.: 31.00 3rd Qu.: 1.00 3rd Qu.: 1.000   
## Max. :232.00 Max. :223.00 Max. :120.00 Max. :51.000   
## NA's :6413 NA's :14251 NA's :42977 NA's :9233   
## SH SF GIDP G\_old   
## Min. : 0.000 Min. : 0.0 Min. : 0.00 Min. : 0.00   
## 1st Qu.: 0.000 1st Qu.: 0.0 1st Qu.: 0.00 1st Qu.: 11.00   
## Median : 1.000 Median : 0.0 Median : 1.00 Median : 34.00   
## Mean : 2.564 Mean : 1.2 Mean : 3.33 Mean : 50.99   
## 3rd Qu.: 3.000 3rd Qu.: 2.0 3rd Qu.: 5.00 3rd Qu.: 82.00   
## Max. :67.000 Max. :19.0 Max. :36.00 Max. :165.00   
## NA's :12751 NA's :42446 NA's :32521 NA's :5189   
## BA OBP B1 SLG   
## Min. :0.000 Min. :0.00 Min. : 0.00 Min. :0.000   
## 1st Qu.:0.148 1st Qu.:0.19 1st Qu.: 1.00 1st Qu.:0.179   
## Median :0.231 Median :0.29 Median : 9.00 Median :0.309   
## Mean :0.209 Mean :0.26 Mean : 29.14 Mean :0.291   
## 3rd Qu.:0.275 3rd Qu.:0.34 3rd Qu.: 48.00 3rd Qu.:0.397   
## Max. :1.000 Max. :1.00 Max. :225.00 Max. :4.000   
## NA's :13520 NA's :49115 NA's :6413 NA's :13520

print(summary(Salaries))

## yearID teamID lgID playerID   
## Min. :1985 Length:23956 Length:23956 Length:23956   
## 1st Qu.:1993 Class :character Class :character Class :character   
## Median :1999 Mode :character Mode :character Mode :character   
## Mean :1999   
## 3rd Qu.:2006   
## Max. :2013   
## salary   
## Min. : 0   
## 1st Qu.: 250000   
## Median : 507950   
## Mean : 1864357   
## 3rd Qu.: 2100000   
## Max. :33000000

# notice the minimum year in the batting data goes back to 1871 (column yearID) but in the salaries data it is starts from 1985 (column yearID)   
Batting = subset(Batting, yearID > 1984)  
print(summary(Batting))

## playerID yearID stint teamID   
## Length:35652 Min. :1985 Min. :1.00 Length:35652   
## Class :character 1st Qu.:1993 1st Qu.:1.00 Class :character   
## Mode :character Median :2000 Median :1.00 Mode :character   
## Mean :2000 Mean :1.08   
## 3rd Qu.:2007 3rd Qu.:1.00   
## Max. :2013 Max. :4.00   
##   
## lgID G G\_batting AB   
## Length:35652 Min. : 1.0 Min. : 0.00 Min. : 0.0   
## Class :character 1st Qu.: 14.0 1st Qu.: 4.00 1st Qu.: 3.0   
## Mode :character Median : 34.0 Median : 27.00 Median : 47.0   
## Mean : 51.7 Mean : 46.28 Mean :144.7   
## 3rd Qu.: 77.0 3rd Qu.: 77.00 3rd Qu.:241.0   
## Max. :163.0 Max. :163.00 Max. :716.0   
## NA's :1406 NA's :4377   
## R H X2B X3B   
## Min. : 0.00 Min. : 0.00 Min. : 0.000 Min. : 0.000   
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.000 1st Qu.: 0.000   
## Median : 4.00 Median : 8.00 Median : 1.000 Median : 0.000   
## Mean : 19.44 Mean : 37.95 Mean : 7.293 Mean : 0.824   
## 3rd Qu.: 30.00 3rd Qu.: 61.00 3rd Qu.:11.000 3rd Qu.: 1.000   
## Max. :152.00 Max. :262.00 Max. :59.000 Max. :23.000   
## NA's :4377 NA's :4377 NA's :4377 NA's :4377   
## HR RBI SB CS   
## Min. : 0.000 Min. : 0.00 Min. : 0.000 Min. : 0.000   
## 1st Qu.: 0.000 1st Qu.: 0.00 1st Qu.: 0.000 1st Qu.: 0.000   
## Median : 0.000 Median : 3.00 Median : 0.000 Median : 0.000   
## Mean : 4.169 Mean : 18.41 Mean : 2.811 Mean : 1.219   
## 3rd Qu.: 5.000 3rd Qu.: 27.00 3rd Qu.: 2.000 3rd Qu.: 1.000   
## Max. :73.000 Max. :165.00 Max. :110.000 Max. :29.000   
## NA's :4377 NA's :4377 NA's :4377 NA's :4377   
## BB SO IBB HBP   
## Min. : 0.00 Min. : 0.00 Min. : 0.000 Min. : 0.000   
## 1st Qu.: 0.00 1st Qu.: 1.00 1st Qu.: 0.000 1st Qu.: 0.000   
## Median : 3.00 Median : 12.00 Median : 0.000 Median : 0.000   
## Mean : 14.06 Mean : 27.03 Mean : 1.171 Mean : 1.273   
## 3rd Qu.: 21.00 3rd Qu.: 42.00 3rd Qu.: 1.000 3rd Qu.: 1.000   
## Max. :232.00 Max. :223.00 Max. :120.000 Max. :35.000   
## NA's :4377 NA's :4377 NA's :4378 NA's :4387   
## SH SF GIDP G\_old   
## Min. : 0.000 Min. : 0.000 Min. : 0.00 Min. : 0.0   
## 1st Qu.: 0.000 1st Qu.: 0.000 1st Qu.: 0.00 1st Qu.: 11.0   
## Median : 0.000 Median : 0.000 Median : 1.00 Median : 32.0   
## Mean : 1.465 Mean : 1.212 Mean : 3.25 Mean : 49.7   
## 3rd Qu.: 2.000 3rd Qu.: 2.000 3rd Qu.: 5.00 3rd Qu.: 77.0   
## Max. :39.000 Max. :17.000 Max. :35.00 Max. :163.0   
## NA's :4377 NA's :4378 NA's :4377 NA's :5189   
## BA OBP B1 SLG   
## Min. :0.000 Min. :0.000 Min. : 0.00 Min. :0.000   
## 1st Qu.:0.136 1st Qu.:0.188 1st Qu.: 0.00 1st Qu.:0.167   
## Median :0.233 Median :0.296 Median : 6.00 Median :0.333   
## Mean :0.205 Mean :0.262 Mean : 25.66 Mean :0.304   
## 3rd Qu.:0.274 3rd Qu.:0.342 3rd Qu.: 42.00 3rd Qu.:0.423   
## Max. :1.000 Max. :1.000 Max. :225.00 Max. :4.000   
## NA's :8905 NA's :8821 NA's :4377 NA's :8905

combo\_data = merge(Batting,Salaries,by = c('playerID','yearID'))  
print(summary(combo\_data))

## playerID yearID stint teamID.x   
## Length:25397 Min. :1985 Min. :1.000 Length:25397   
## Class :character 1st Qu.:1993 1st Qu.:1.000 Class :character   
## Mode :character Median :1999 Median :1.000 Mode :character   
## Mean :1999 Mean :1.098   
## 3rd Qu.:2006 3rd Qu.:1.000   
## Max. :2013 Max. :4.000   
##   
## lgID.x G G\_batting AB   
## Length:25397 Min. : 1.00 Min. : 0.00 Min. : 0.0   
## Class :character 1st Qu.: 26.00 1st Qu.: 8.00 1st Qu.: 5.0   
## Mode :character Median : 50.00 Median : 42.00 Median : 85.0   
## Mean : 64.06 Mean : 57.58 Mean :182.4   
## 3rd Qu.:101.00 3rd Qu.:101.00 3rd Qu.:336.0   
## Max. :163.00 Max. :163.00 Max. :716.0   
## NA's :906 NA's :2661   
## R H X2B X3B   
## Min. : 0.00 Min. : 0.00 Min. : 0.000 Min. : 0.000   
## 1st Qu.: 0.00 1st Qu.: 1.00 1st Qu.: 0.000 1st Qu.: 0.000   
## Median : 9.00 Median : 19.00 Median : 3.000 Median : 0.000   
## Mean : 24.71 Mean : 48.18 Mean : 9.276 Mean : 1.033   
## 3rd Qu.: 43.00 3rd Qu.: 87.25 3rd Qu.:16.000 3rd Qu.: 1.000   
## Max. :152.00 Max. :262.00 Max. :59.000 Max. :23.000   
## NA's :2661 NA's :2661 NA's :2661 NA's :2661   
## HR RBI SB CS   
## Min. : 0.000 Min. : 0.00 Min. : 0.000 Min. : 0.00   
## 1st Qu.: 0.000 1st Qu.: 0.00 1st Qu.: 0.000 1st Qu.: 0.00   
## Median : 1.000 Median : 8.00 Median : 0.000 Median : 0.00   
## Mean : 5.369 Mean : 23.56 Mean : 3.568 Mean : 1.54   
## 3rd Qu.: 7.000 3rd Qu.: 39.00 3rd Qu.: 3.000 3rd Qu.: 2.00   
## Max. :73.000 Max. :165.00 Max. :110.000 Max. :29.00   
## NA's :2661 NA's :2661 NA's :2661 NA's :2661   
## BB SO IBB HBP   
## Min. : 0.00 Min. : 0.00 Min. : 0.000 Min. : 0.000   
## 1st Qu.: 0.00 1st Qu.: 2.00 1st Qu.: 0.000 1st Qu.: 0.000   
## Median : 6.00 Median : 20.00 Median : 0.000 Median : 0.000   
## Mean : 17.98 Mean : 33.52 Mean : 1.533 Mean : 1.614   
## 3rd Qu.: 29.00 3rd Qu.: 55.00 3rd Qu.: 2.000 3rd Qu.: 2.000   
## Max. :232.00 Max. :223.00 Max. :120.000 Max. :35.000   
## NA's :2661 NA's :2661 NA's :2662 NA's :2670   
## SH SF GIDP G\_old   
## Min. : 0.000 Min. : 0.000 Min. : 0.000 Min. : 0.00   
## 1st Qu.: 0.000 1st Qu.: 0.000 1st Qu.: 0.000 1st Qu.: 20.00   
## Median : 0.000 Median : 0.000 Median : 2.000 Median : 47.00   
## Mean : 1.786 Mean : 1.554 Mean : 4.127 Mean : 61.43   
## 3rd Qu.: 2.000 3rd Qu.: 2.000 3rd Qu.: 7.000 3rd Qu.:101.00   
## Max. :39.000 Max. :17.000 Max. :35.000 Max. :163.00   
## NA's :2661 NA's :2662 NA's :2661 NA's :3414   
## BA OBP B1 SLG   
## Min. :0.000 Min. :0.000 Min. : 0.0 Min. :0.000   
## 1st Qu.:0.160 1st Qu.:0.208 1st Qu.: 0.0 1st Qu.:0.200   
## Median :0.242 Median :0.305 Median : 13.0 Median :0.351   
## Mean :0.212 Mean :0.270 Mean : 32.5 Mean :0.317   
## 3rd Qu.:0.276 3rd Qu.:0.346 3rd Qu.: 59.0 3rd Qu.:0.432   
## Max. :1.000 Max. :1.000 Max. :225.0 Max. :4.000   
## NA's :5618 NA's :5562 NA's :2661 NA's :5618   
## teamID.y lgID.y salary   
## Length:25397 Length:25397 Min. : 0   
## Class :character Class :character 1st Qu.: 255000   
## Mode :character Mode :character Median : 550000   
## Mean : 1879256   
## 3rd Qu.: 2150000   
## Max. :33000000   
##

# creating a data frame for the lost players (Jason Giambi (giambja01),Johnny Damon (damonjo01),Rainer Gustavo "Ray" Olmedo (saenzol01))  
  
lost\_players = subset(combo\_data,playerID %in% c('giambja01','damonjo01','saenzol01'))  
lost\_players = subset(lost\_players,yearID == 2001)  
lost\_players = select(lost\_players,playerID,H,X2B,X3B,HR,OBP,SLG,BA,AB)  
print(lost\_players)

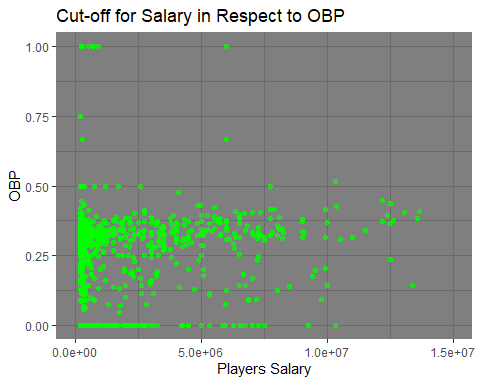
## playerID H X2B X3B HR OBP SLG BA AB  
## 5141 damonjo01 165 34 4 9 0.324 0.363 0.256 644  
## 7878 giambja01 178 47 2 38 0.477 0.660 0.342 520  
## 20114 saenzol01 67 21 1 9 0.291 0.384 0.220 305

library(ggthemes)

## Warning: package 'ggthemes' was built under R version 4.0.5

avl\_players = filter(combo\_data,yearID == 2001)  
pl\_avl = ggplot(avl\_players,aes(salary,OBP)) + geom\_point(color = 'green',alpha = 0.7)  
pl\_avl = pl\_avl + scale\_x\_continuous(name = 'Players Salary',limits = c(0,15000000))+ scale\_y\_continuous(name = 'OBP',limits = c(0,1.0))  
pl\_avl = pl\_avl + ggtitle('Cut-off for Salary in Respect to OBP') + theme\_dark()  
print(pl\_avl)

## Warning: Removed 170 rows containing missing values (geom\_point).



print(sum(lost\_players$AB))

## [1] 1469

# The total AB of the lost players is 1469. This is about 1500, meaning I should probably cut off my avail.players at 1500/3= 500 AB.  
`%!in%` <- Negate(`%in%`)  
avl\_players = filter(avl\_players,salary < 7000000,OBP > 0, AB >= 500,playerID %!in% c(' damonjo01','giambja01','saenzol01'))  
print(head(avl\_players))

## playerID yearID stint teamID.x lgID.x G G\_batting AB R H X2B X3B HR  
## 1 abreubo01 2001 1 PHI NL 162 162 588 118 170 48 4 31  
## 2 aloumo01 2001 1 HOU NL 136 136 513 79 170 31 1 27  
## 3 anderga01 2001 1 ANA AL 161 161 672 83 194 39 2 28  
## 4 anderma02 2001 1 PHI NL 147 147 522 69 153 30 2 11  
## 5 aurilri01 2001 1 SFN NL 156 156 636 114 206 37 5 37  
## 6 bagweje01 2001 1 HOU NL 161 161 600 126 173 43 4 39  
## RBI SB CS BB SO IBB HBP SH SF GIDP G\_old BA OBP B1 SLG teamID.y  
## 1 110 36 14 106 137 11 1 0 9 13 162 0.289 0.393 87 0.543 PHI  
## 2 108 5 1 57 57 14 3 0 8 18 136 0.331 0.396 111 0.554 HOU  
## 3 123 13 6 27 100 4 0 0 5 12 161 0.289 0.314 125 0.478 ANA  
## 4 61 8 5 35 74 5 2 10 5 12 147 0.293 0.337 110 0.421 PHI  
## 5 97 1 3 47 83 2 0 3 3 14 156 0.324 0.369 127 0.572 SFN  
## 6 130 11 3 106 135 5 6 0 5 20 161 0.288 0.397 87 0.568 HOU  
## lgID.y salary  
## 1 NL 4983000  
## 2 NL 5250000  
## 3 AL 4500000  
## 4 NL 280000  
## 5 NL 3250000  
## 6 NL 6500000

possible = head(arrange(avl\_players,desc(OBP)),10)  
possible = select(possible,playerID,OBP,AB,salary)  
possible = head(arrange(possible,salary),3)  
print(possible)

## playerID OBP AB salary  
## 1 pujolal01 0.403 590 200000  
## 2 berkmla01 0.430 577 305000  
## 3 gonzalu01 0.429 609 4833333