Project2

yleung

## R Libraries

library(readxl)  
library(tidyverse)  
library(aod)  
library(caret)  
library(glmnet)  
library(corrplot)  
library(MASS)  
library(regclass)  
library(FactoMineR)  
library(factoextra)  
library(pROC)  
library(psych)  
library(MLmetrics)

## Import Data & ETL

xlsx\_kobe <- read\_excel("C:\\Users\\Yat\\Documents\\MSDS\\MSDS 6372\\Project 2\\project2Data.xlsx")  
  
#Take out an ID column from when data was imported  
kobe<-xlsx\_kobe[,c(2:29)]  
  
#Re-code the character columns into factors  
kobe<-as.data.frame(unclass(kobe))  
kobe$shot\_made\_flag<-as.factor(kobe$shot\_made\_flag)  
kobe$shot\_type<-as.factor(kobe$shot\_type)  
kobe$shot\_zone\_area<-as.factor(kobe$shot\_zone\_area)  
kobe$shot\_zone\_basic<-as.factor(kobe$shot\_zone\_basic)  
kobe$shot\_zone\_range<-as.factor(kobe$shot\_zone\_range)  
kobe$game\_date<-as.factor(kobe$game\_date)  
kobe$season<-as.factor(kobe$season)  
kobe$period<-as.factor(kobe$period) #change period into factor?  
kobe$playoffs<-as.factor(kobe$playoffs)  
  
#subsetting variables (sans IDs)  
kobe<-kobe[,c(1:2,5:20,22:25,27:28)]  
kobe<-na.omit(kobe)  
  
  
#data check  
#head(kobe)  
summary(kobe)

## action\_type combined\_shot\_type lat   
## Jump Shot :15836 Bank Shot: 120 Min. :33.25   
## Layup Shot : 2154 Dunk : 1056 1st Qu.:33.88   
## Driving Layup Shot : 1628 Hook Shot: 127 Median :33.97   
## Turnaround Jump Shot: 891 Jump Shot:19710 Mean :33.95   
## Fadeaway Jump Shot : 872 Layup : 4532 3rd Qu.:34.04   
## Running Jump Shot : 779 Tip Shot : 152 Max. :34.09   
## (Other) : 3537   
## loc\_x loc\_y lon minutes\_remaining  
## Min. :-250.000 Min. :-44.00 Min. :-118.5 Min. : 0.000   
## 1st Qu.: -67.000 1st Qu.: 4.00 1st Qu.:-118.3 1st Qu.: 2.000   
## Median : 0.000 Median : 74.00 Median :-118.3 Median : 5.000   
## Mean : 7.148 Mean : 91.26 Mean :-118.3 Mean : 4.887   
## 3rd Qu.: 94.000 3rd Qu.:160.00 3rd Qu.:-118.2 3rd Qu.: 8.000   
## Max. : 248.000 Max. :791.00 Max. :-118.0 Max. :11.000   
##   
## period playoffs season seconds\_remaining shot\_distance   
## 1:6700 0:21939 2005-06: 1924 Min. : 0.00 Min. : 0.00   
## 2:5635 1: 3758 2002-03: 1852 1st Qu.:13.00 1st Qu.: 5.00   
## 3:7002 2008-09: 1851 Median :28.00 Median :15.00   
## 4:6043 2007-08: 1819 Mean :28.31 Mean :13.46   
## 5: 280 2009-10: 1772 3rd Qu.:43.00 3rd Qu.:21.00   
## 6: 30 2001-02: 1708 Max. :59.00 Max. :79.00   
## 7: 7 (Other):14771   
## shot\_made\_flag shot\_type shot\_zone\_area   
## 0:14232 2PT Field Goal:20285 Back Court(BC) : 72   
## 1:11465 3PT Field Goal: 5412 Center(C) :11289   
## Left Side Center(LC) : 3364   
## Left Side(L) : 3132   
## Right Side Center(RC): 3981   
## Right Side(R) : 3859   
##   
## shot\_zone\_basic shot\_zone\_range team\_id   
## Above the Break 3 : 4720 16-24 ft. :6907 Min. :1.611e+09   
## Backcourt : 60 24+ ft. :5281 1st Qu.:1.611e+09   
## In The Paint (Non-RA): 3880 8-16 ft. :5580 Median :1.611e+09   
## Left Corner 3 : 240 Back Court Shot: 72 Mean :1.611e+09   
## Mid-Range :10532 Less Than 8 ft.:7857 3rd Qu.:1.611e+09   
## Restricted Area : 5932 Max. :1.611e+09   
## Right Corner 3 : 333   
## game\_date matchup opponent shot\_id   
## 2016-04-13: 43 LAL @ SAS : 853 SAS : 1638 Min. : 2   
## 2007-03-30: 41 LAL vs. SAS: 765 PHX : 1535 1st Qu.: 7646   
## 2002-11-07: 39 LAL @ PHX : 751 HOU : 1399 Median :15336   
## 2006-01-22: 39 LAL @ SAC : 751 SAC : 1397 Mean :15328   
## 2008-01-14: 37 LAL vs. HOU: 722 DEN : 1352 3rd Qu.:22976   
## 2010-01-08: 36 LAL vs. PHX: 721 POR : 1292 Max. :30697   
## (Other) :25462 (Other) :21134 (Other):17084   
## arena\_temp avgnoisedb   
## Min. :64.00 Min. : 88.56   
## 1st Qu.:69.00 1st Qu.: 93.40   
## Median :70.00 Median : 94.92   
## Mean :70.11 Mean : 94.95   
## 3rd Qu.:71.00 3rd Qu.: 96.49   
## Max. :79.00 Max. :102.43   
##

str(kobe)

## 'data.frame': 25697 obs. of 24 variables:  
## $ action\_type : Factor w/ 55 levels "Alley Oop Dunk Shot",..: 26 26 26 5 26 27 26 41 26 26 ...  
## $ combined\_shot\_type: Factor w/ 6 levels "Bank Shot","Dunk",..: 4 4 4 2 4 5 4 4 4 4 ...  
## $ lat : num 34 33.9 33.9 34 34.1 ...  
## $ loc\_x : num -157 -101 138 0 -145 0 -65 -33 -94 121 ...  
## $ loc\_y : num 0 135 175 0 -11 0 108 125 238 127 ...  
## $ lon : num -118 -118 -118 -118 -118 ...  
## $ minutes\_remaining : num 10 7 6 6 9 8 6 3 1 11 ...  
## $ period : Factor w/ 7 levels "1","2","3","4",..: 1 1 1 2 3 3 3 3 3 1 ...  
## $ playoffs : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ season : Factor w/ 20 levels "1996-97","1997-98",..: 5 5 5 5 5 5 5 5 5 5 ...  
## $ seconds\_remaining : num 22 45 52 19 32 52 12 36 56 0 ...  
## $ shot\_distance : num 15 16 22 0 14 0 12 12 25 17 ...  
## $ shot\_made\_flag : Factor w/ 2 levels "0","1": 1 2 1 2 1 2 2 1 1 2 ...  
## $ shot\_type : Factor w/ 2 levels "2PT Field Goal",..: 1 1 1 1 1 1 1 1 2 1 ...  
## $ shot\_zone\_area : Factor w/ 6 levels "Back Court(BC)",..: 4 3 5 2 4 2 4 2 3 5 ...  
## $ shot\_zone\_basic : Factor w/ 7 levels "Above the Break 3",..: 5 5 5 6 5 6 3 3 1 5 ...  
## $ shot\_zone\_range : Factor w/ 5 levels "16-24 ft.","24+ ft.",..: 3 1 1 5 3 5 3 3 2 1 ...  
## $ team\_id : num 1.61e+09 1.61e+09 1.61e+09 1.61e+09 1.61e+09 ...  
## $ game\_date : Factor w/ 1558 levels "1996-11-03","1996-11-05",..: 310 310 310 310 310 310 310 310 310 311 ...  
## $ matchup : Factor w/ 74 levels "LAL @ ATL","LAL @ BKN",..: 29 29 29 29 29 29 29 29 29 72 ...  
## $ opponent : Factor w/ 33 levels "ATL","BKN","BOS",..: 26 26 26 26 26 26 26 26 26 31 ...  
## $ shot\_id : num 2 3 4 5 6 7 9 10 11 12 ...  
## $ arena\_temp : num 69 69 69 69 69 69 69 69 69 69 ...  
## $ avgnoisedb : num 94.1 94.1 94.1 94.1 94.1 ...

#numerical variables  
kobe.NV<-kobe[,c(3:7,11:12,23:24)]  
kobe.NV<-na.omit(kobe.NV)  
fit<-prcomp(~., data=kobe.NV, cor=TRUE)

## Warning: In prcomp.default(x, ...) :  
## extra argument 'cor' will be disregarded

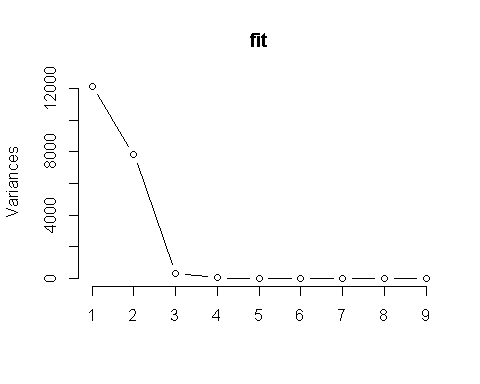
summary(fit) # print variance accounted for

## Importance of components:  
## PC1 PC2 PC3 PC4 PC5 PC6  
## Standard deviation 110.1036 88.4567 17.49319 5.36689 3.44141 2.28796  
## Proportion of Variance 0.5971 0.3854 0.01507 0.00142 0.00058 0.00026  
## Cumulative Proportion 0.5971 0.9825 0.99754 0.99896 0.99954 0.99980  
## PC7 PC8 PC9  
## Standard deviation 2.0224 7.845e-15 4.932e-15  
## Proportion of Variance 0.0002 0.000e+00 0.000e+00  
## Cumulative Proportion 1.0000 1.000e+00 1.000e+00

loadings(fit) # pc loadings

## NULL

plot(fit,type="lines") # scree plot



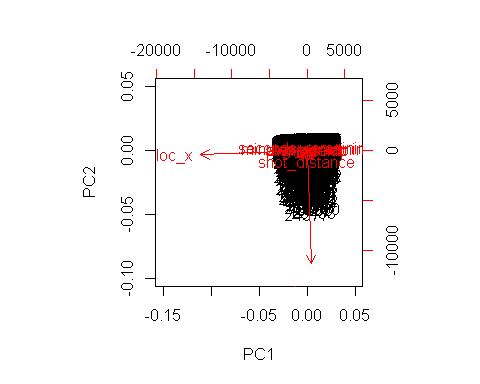
fit$scores # the principal components

## NULL

biplot(fit,expand=10, xlim=c(-0.15, 0.05), ylim=c(-0.1, 0.05))

## Warning in arrows(0, 0, y[, 1L] \* 0.8, y[, 2L] \* 0.8, col = col[2L], length  
## = arrow.len): zero-length arrow is of indeterminate angle and so skipped

## Warning in arrows(0, 0, y[, 1L] \* 0.8, y[, 2L] \* 0.8, col = col[2L], length  
## = arrow.len): zero-length arrow is of indeterminate angle and so skipped



# Varimax Rotated Principal Components  
# Extract, rotate and retain 5 PCs   
component.retained <- principal(kobe.NV, nfactors=5, rotate="varimax")

## Warning in cor.smooth(r): Matrix was not positive definite, smoothing was  
## done

## Warning in principal(kobe.NV, nfactors = 5, rotate = "varimax"): The matrix  
## is not positive semi-definite, scores found from Structure loadings

component.retained

## Principal Components Analysis  
## Call: principal(r = kobe.NV, nfactors = 5, rotate = "varimax")  
## Standardized loadings (pattern matrix) based upon correlation matrix  
## RC1 RC2 RC3 RC5 RC4 h2 u2 com  
## lat -0.98 0.02 0.00 0.03 0.02 0.96 0.03596 1  
## loc\_x 0.00 1.00 0.00 0.00 0.00 1.00 0.00020 1  
## loc\_y 0.98 -0.02 0.00 -0.03 -0.02 0.96 0.03596 1  
## lon 0.00 1.00 0.00 0.00 0.00 1.00 0.00020 1  
## minutes\_remaining -0.05 0.01 -0.01 0.99 0.01 0.98 0.01661 1  
## seconds\_remaining -0.04 0.00 0.00 0.01 1.00 1.00 0.00091 1  
## shot\_distance 0.91 0.03 -0.01 -0.02 -0.02 0.83 0.16620 1  
## arena\_temp -0.01 0.01 0.71 -0.09 0.02 0.52 0.48226 1  
## avgnoisedb 0.00 -0.02 0.73 0.09 -0.02 0.54 0.45957 1  
##   
## RC1 RC2 RC3 RC5 RC4  
## SS loadings 2.76 2.00 1.04 1.00 1.00  
## Proportion Var 0.31 0.22 0.12 0.11 0.11  
## Cumulative Var 0.31 0.53 0.64 0.76 0.87  
## Proportion Explained 0.35 0.26 0.13 0.13 0.13  
## Cumulative Proportion 0.35 0.61 0.74 0.87 1.00  
##   
## Mean item complexity = 1  
## Test of the hypothesis that 5 components are sufficient.  
##   
## The root mean square of the residuals (RMSR) is 0.08   
## with the empirical chi square 12933.02 with prob < 0   
##   
## Fit based upon off diagonal values = 0.93

# Principal Axis Factor Analysis  
axis.fit <- factor.pa(kobe.NV, 5)

## Warning: factor.pa is deprecated. Please use the fa function with fm=pa

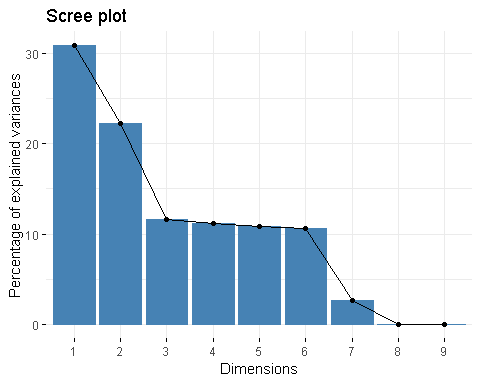
## too many factors requested for this number of variables to use SMC, 1s used instead

## Warning in cor.smooth(r): Matrix was not positive definite, smoothing was  
## done

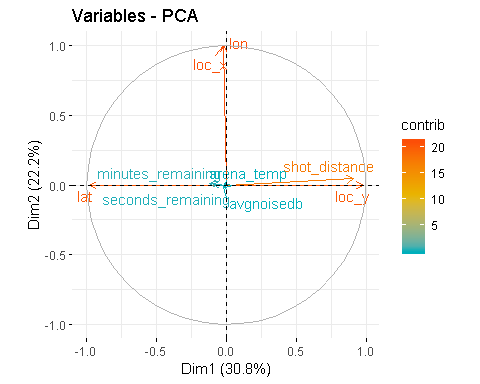
axis.fit

## Factor Analysis using method = pa  
## Call: factor.pa(r = kobe.NV, nfactors = 5)  
## Unstandardized loadings (pattern matrix) based upon covariance matrix  
## PA1 PA2 PA3 PA4 PA5 h2 u2 H2 U2  
## lat -0.99 0.02 0.02 0.03 0.03 0.9904 0.01 0.9900 0.01  
## loc\_x 0.00 1.00 0.00 0.00 0.02 1.0000 0.00 1.0000 0.00  
## loc\_y 0.99 -0.02 -0.02 -0.03 -0.03 0.9904 0.01 0.9900 0.01  
## lon 0.00 1.00 0.00 0.00 0.02 1.0000 0.00 1.0000 0.00  
## minutes\_remaining -0.05 0.00 0.01 0.99 -0.03 0.9833 0.02 0.9801 0.02  
## seconds\_remaining -0.04 0.00 1.00 0.01 0.00 0.9991 0.00 1.0000 0.00  
## shot\_distance 0.83 0.02 -0.02 -0.02 -0.03 0.6848 0.32 0.6815 0.32  
## arena\_temp -0.01 0.01 0.00 -0.01 0.07 0.0047 1.00 0.0047 1.00  
## avgnoisedb 0.01 -0.02 0.00 0.02 0.59 0.3522 0.65 0.3514 0.65  
##   
## PA1 PA2 PA3 PA4 PA5  
## SS loadings 2.66 2.00 1.00 0.98 0.36  
## Proportion Var 0.30 0.22 0.11 0.11 0.04  
## Cumulative Var 0.30 0.52 0.63 0.74 0.78  
## Proportion Explained 0.38 0.29 0.14 0.14 0.05  
## Cumulative Proportion 0.38 0.67 0.81 0.95 1.00  
##   
## Standardized loadings (pattern matrix)  
## item PA1 PA2 PA3 PA4 PA5 h2 u2  
## lat 1 -0.99 0.02 0.02 0.03 0.03 0.9900 0.01  
## loc\_x 2 0.00 1.00 0.00 0.00 0.02 1.0000 0.00  
## loc\_y 3 0.99 -0.02 -0.02 -0.03 -0.03 0.9900 0.01  
## lon 4 0.00 1.00 0.00 0.00 0.02 1.0000 0.00  
## minutes\_remaining 5 -0.05 0.00 0.01 0.99 -0.03 0.9801 0.02  
## seconds\_remaining 6 -0.04 0.00 1.00 0.01 0.00 1.0000 0.00  
## shot\_distance 7 0.82 0.02 -0.02 -0.02 -0.03 0.6815 0.32  
## arena\_temp 8 -0.01 0.01 0.00 -0.01 0.07 0.0047 1.00  
## avgnoisedb 9 0.01 -0.02 0.00 0.02 0.59 0.3514 0.65  
##   
## PA1 PA2 PA3 PA4 PA5  
## SS loadings 2.66 2.00 1.00 0.98 0.36  
## Proportion Var 0.30 0.22 0.11 0.11 0.04  
## Cumulative Var 0.30 0.52 0.63 0.74 0.78  
## Cum. factor Var 0.38 0.67 0.81 0.95 1.00  
##   
## Mean item complexity = 1  
## Test of the hypothesis that 5 factors are sufficient.  
##   
## The degrees of freedom for the null model are 36 and the objective function was 45.79 with Chi Square of 1176407  
## The degrees of freedom for the model are 1 and the objective function was 28.99   
##   
## The root mean square of the residuals (RMSR) is 0   
## The df corrected root mean square of the residuals is 0.01   
##   
## The harmonic number of observations is 25697 with the empirical chi square 8.31 with prob < 0.0039   
## The total number of observations was 25697 with Likelihood Chi Square = 744730.8 with prob < 0   
##   
## Tucker Lewis Index of factoring reliability = -21.794  
## RMSEA index = 5.384 and the 90 % confidence intervals are 5.373 5.394  
## BIC = 744720.7  
## Fit based upon off diagonal values = 1  
## Measures of factor score adequacy   
## PA1 PA2 PA3 PA4 PA5  
## Correlation of (regression) scores with factors 1.00 1 1 0.99 0.60  
## Multiple R square of scores with factors 0.99 1 1 0.98 0.36  
## Minimum correlation of possible factor scores 0.98 1 1 0.96 -0.29

#PCA  
res.pca1 <- prcomp(kobe.NV, scale = TRUE)  
fviz\_eig(res.pca1)



fviz\_pca\_var(res.pca1,  
 col.var = "contrib", # Color by contributions to the PC  
 gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),  
 repel = TRUE # Avoid text overlapping  
 )



#numerical variables with lat removed  
kobe.NV2<-kobe[,c(4:7,11:12,23:24)]  
kobe.NV2<-na.omit(kobe.NV2)  
fit2<-prcomp(~., data=kobe.NV2, cor=TRUE)

## Warning: In prcomp.default(x, ...) :  
## extra argument 'cor' will be disregarded

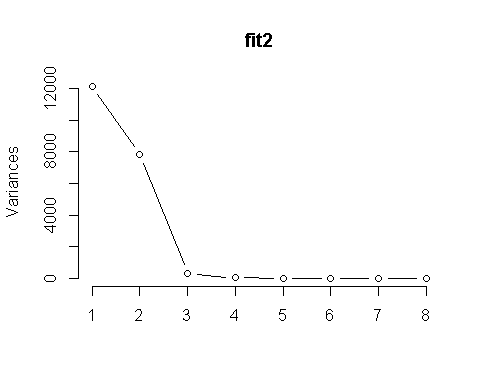
summary(fit2) # print variance accounted for

## Importance of components:  
## PC1 PC2 PC3 PC4 PC5 PC6  
## Standard deviation 110.1036 88.4567 17.49319 5.36689 3.44141 2.28796  
## Proportion of Variance 0.5971 0.3854 0.01507 0.00142 0.00058 0.00026  
## Cumulative Proportion 0.5971 0.9825 0.99754 0.99896 0.99954 0.99980  
## PC7 PC8  
## Standard deviation 2.0224 1.179e-14  
## Proportion of Variance 0.0002 0.000e+00  
## Cumulative Proportion 1.0000 1.000e+00

loadings(fit2) # pc loadings

## NULL

plot(fit2,type="lines") # scree plot



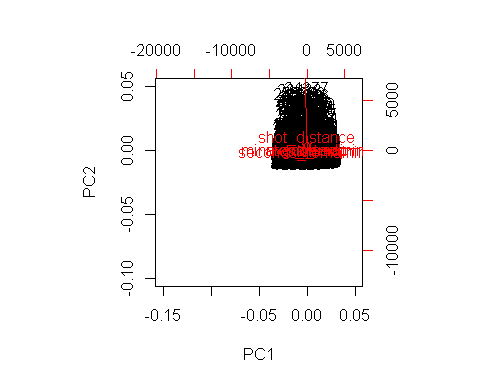
fit2$scores # the principal components

## NULL

biplot(fit2,expand=10, xlim=c(-0.15, 0.05), ylim=c(-0.1, 0.05))

## Warning in arrows(0, 0, y[, 1L] \* 0.8, y[, 2L] \* 0.8, col = col[2L], length  
## = arrow.len): zero-length arrow is of indeterminate angle and so skipped

## Warning in arrows(0, 0, y[, 1L] \* 0.8, y[, 2L] \* 0.8, col = col[2L], length  
## = arrow.len): zero-length arrow is of indeterminate angle and so skipped



# Varimax Rotated Principal Components  
# Extract, rotate and retain 5 PCs   
component.retained2 <- principal(kobe.NV2, nfactors=5, rotate="varimax")

## Warning in log(det(m.inv.r)): NaNs produced

## Warning in log(det(r)): NaNs produced

## In factor.stats, the correlation matrix is singular, an approximation is used

## Warning in principal(kobe.NV2, nfactors = 5, rotate = "varimax"): The  
## matrix is not positive semi-definite, scores found from Structure loadings

component.retained2

## Principal Components Analysis  
## Call: principal(r = kobe.NV2, nfactors = 5, rotate = "varimax")  
## Standardized loadings (pattern matrix) based upon correlation matrix  
## RC1 RC2 RC3 RC5 RC4 h2 u2 com  
## loc\_x 1.00 0.00 0.00 0.00 0.00 1.00 0.00016 1  
## loc\_y -0.02 0.95 0.00 -0.04 -0.02 0.91 0.09054 1  
## lon 1.00 0.00 0.00 0.00 0.00 1.00 0.00016 1  
## minutes\_remaining 0.01 -0.05 -0.01 0.99 0.01 0.98 0.01831 1  
## seconds\_remaining 0.00 -0.04 0.00 0.01 1.00 1.00 0.00074 1  
## shot\_distance 0.02 0.95 -0.01 -0.02 -0.02 0.91 0.09029 1  
## arena\_temp 0.01 -0.01 0.71 -0.10 0.02 0.52 0.48289 1  
## avgnoisedb -0.02 0.00 0.73 0.10 -0.02 0.54 0.45733 1  
##   
## RC1 RC2 RC3 RC5 RC4  
## SS loadings 2.00 1.82 1.04 1.00 1.00  
## Proportion Var 0.25 0.23 0.13 0.13 0.12  
## Cumulative Var 0.25 0.48 0.61 0.73 0.86  
## Proportion Explained 0.29 0.27 0.15 0.15 0.15  
## Cumulative Proportion 0.29 0.56 0.71 0.85 1.00  
##   
## Mean item complexity = 1  
## Test of the hypothesis that 5 components are sufficient.  
##   
## The root mean square of the residuals (RMSR) is 0.09   
## with the empirical chi square 12705.21 with prob < NA   
##   
## Fit based upon off diagonal values = 0.85

# Principal Axis Factor Analysis  
axis.fit2 <- factor.pa(kobe.NV2, 5)

## Warning: factor.pa is deprecated. Please use the fa function with fm=pa

## too many factors requested for this number of variables to use SMC, 1s used instead

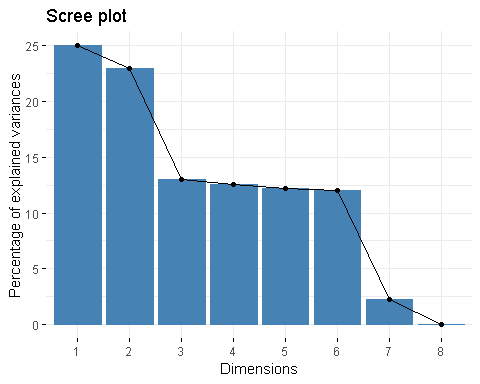
## Warning in log(det(r)): NaNs produced

## In factor.stats, the correlation matrix is singular, an approximation is used

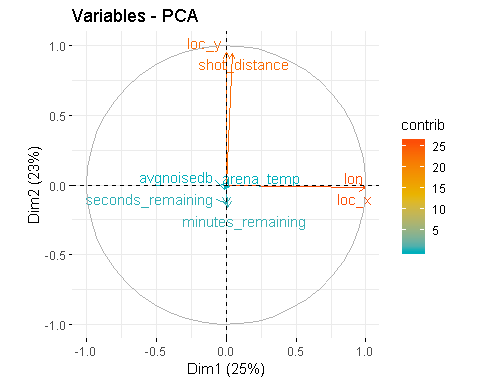
axis.fit2

## Factor Analysis using method = pa  
## Call: factor.pa(r = kobe.NV2, nfactors = 5)  
## Unstandardized loadings (pattern matrix) based upon covariance matrix  
## PA1 PA2 PA3 PA4 PA5 h2 u2 H2 U2  
## loc\_x 1.00 0.00 0.00 0.00 0.02 1.0000 0.00 1.0000 0.00  
## loc\_y -0.02 0.90 -0.02 -0.03 -0.04 0.8190 0.18 0.8198 0.18  
## lon 1.00 0.00 0.00 0.00 0.02 1.0000 0.00 1.0000 0.00  
## minutes\_remaining 0.00 -0.05 0.01 0.99 -0.03 0.9816 0.02 0.9800 0.02  
## seconds\_remaining 0.00 -0.04 1.00 0.01 0.00 0.9993 0.00 1.0000 0.00  
## shot\_distance 0.02 0.90 -0.02 -0.02 -0.04 0.8198 0.18 0.8200 0.18  
## arena\_temp 0.01 -0.01 0.00 -0.01 0.07 0.0045 1.00 0.0045 1.00  
## avgnoisedb -0.02 0.01 0.00 0.02 0.60 0.3646 0.64 0.3629 0.64  
##   
## PA1 PA2 PA3 PA4 PA5  
## SS loadings 2.00 1.64 1.00 0.98 0.37  
## Proportion Var 0.25 0.20 0.12 0.12 0.05  
## Cumulative Var 0.25 0.45 0.58 0.70 0.75  
## Proportion Explained 0.33 0.27 0.17 0.16 0.06  
## Cumulative Proportion 0.33 0.61 0.77 0.94 1.00  
##   
## Standardized loadings (pattern matrix)  
## item PA1 PA2 PA3 PA4 PA5 h2 u2  
## loc\_x 1 1.00 0.00 0.00 0.00 0.02 1.0000 0.00  
## loc\_y 2 -0.02 0.90 -0.02 -0.03 -0.04 0.8198 0.18  
## lon 3 1.00 0.00 0.00 0.00 0.02 1.0000 0.00  
## minutes\_remaining 4 0.00 -0.05 0.01 0.99 -0.03 0.9800 0.02  
## seconds\_remaining 5 0.00 -0.04 1.00 0.01 0.00 1.0000 0.00  
## shot\_distance 6 0.02 0.90 -0.02 -0.02 -0.04 0.8200 0.18  
## arena\_temp 7 0.01 -0.01 0.00 -0.01 0.07 0.0045 1.00  
## avgnoisedb 8 -0.02 0.01 0.00 0.02 0.60 0.3629 0.64  
##   
## PA1 PA2 PA3 PA4 PA5  
## SS loadings 2.00 1.64 1.00 0.98 0.37  
## Proportion Var 0.25 0.20 0.12 0.12 0.05  
## Cumulative Var 0.25 0.45 0.58 0.70 0.75  
## Cum. factor Var 0.33 0.61 0.77 0.94 1.00  
##   
## Mean item complexity = 1  
## Test of the hypothesis that 5 factors are sufficient.  
##   
## The degrees of freedom for the null model are 28 and the objective function was NaN with Chi Square of NaN  
## The degrees of freedom for the model are -2 and the objective function was 37.12   
##   
## The root mean square of the residuals (RMSR) is 0   
## The df corrected root mean square of the residuals is NA   
##   
## The harmonic number of observations is 25697 with the empirical chi square 0.76 with prob < NA   
## The total number of observations was 25697 with Likelihood Chi Square = 953663.5 with prob < NA   
##   
## Tucker Lewis Index of factoring reliability = NaN  
## Fit based upon off diagonal values = 1  
## Measures of factor score adequacy   
## PA1 PA2 PA3 PA4 PA5  
## Correlation of (regression) scores with factors 1 0.95 1 0.99 0.61  
## Multiple R square of scores with factors 1 0.90 1 0.98 0.37  
## Minimum correlation of possible factor scores 1 0.80 1 0.96 -0.26

#PCA  
res.pca2 <- prcomp(kobe.NV2, scale = TRUE)  
fviz\_eig(res.pca2)



fviz\_pca\_var(res.pca2,  
 col.var = "contrib", # Color by contributions to the PC  
 gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),  
 repel = TRUE # Avoid text overlapping  
 )



# Contingency Tables

#contingency table samples. coordination variables not included due to large computation time.  
list.of.xtabs <- lapply(kobe[,c(7,11:12)], function(x) xtabs(~ x + kobe$shot\_made\_flag))  
list.of.xtabs

## $minutes\_remaining  
## kobe$shot\_made\_flag  
## x 0 1  
## 0 2012 1236  
## 1 1257 1017  
## 2 1315 1064  
## 3 1304 1073  
## 4 1319 1089  
## 5 1169 1020  
## 6 1050 931  
## 7 972 904  
## 8 989 760  
## 9 976 900  
## 10 980 818  
## 11 889 653  
##   
## $seconds\_remaining  
## kobe$shot\_made\_flag  
## x 0 1  
## 0 565 268  
## 1 320 221  
## 2 315 193  
## 3 265 183  
## 4 267 213  
## 5 257 211  
## 6 206 183  
## 7 243 181  
## 8 257 193  
## 9 227 163  
## 10 233 204  
## 11 238 204  
## 12 217 196  
## 13 234 204  
## 14 230 191  
## 15 221 177  
## 16 226 191  
## 17 232 174  
## 18 220 202  
## 19 224 197  
## 20 219 195  
## 21 226 177  
## 22 237 182  
## 23 225 199  
## 24 241 184  
## 25 253 184  
## 26 240 179  
## 27 226 192  
## 28 247 228  
## 29 240 177  
## 30 222 175  
## 31 242 168  
## 32 260 209  
## 33 227 205  
## 34 234 182  
## 35 231 182  
## 36 264 205  
## 37 239 206  
## 38 216 229  
## 39 217 194  
## 40 251 195  
## 41 228 212  
## 42 236 183  
## 43 220 187  
## 44 238 191  
## 45 211 209  
## 46 206 175  
## 47 238 180  
## 48 225 173  
## 49 195 173  
## 50 213 171  
## 51 195 190  
## 52 218 187  
## 53 204 165  
## 54 215 170  
## 55 224 143  
## 56 189 194  
## 57 223 212  
## 58 190 177  
## 59 210 177  
##   
## $shot\_distance  
## kobe$shot\_made\_flag  
## x 0 1  
## 0 1683 2925  
## 1 177 345  
## 2 226 257  
## 3 180 139  
## 4 167 150  
## 5 251 193  
## 6 319 238  
## 7 349 254  
## 8 297 271  
## 9 300 228  
## 10 343 245  
## 11 324 278  
## 12 354 290  
## 13 460 326  
## 14 515 379  
## 15 557 415  
## 16 710 455  
## 17 748 477  
## 18 716 479  
## 19 660 471  
## 20 598 395  
## 21 387 284  
## 22 360 211  
## 23 305 188  
## 24 803 460  
## 25 1048 583  
## 26 727 333  
## 27 316 134  
## 28 128 38  
## 29 54 13  
## 30 11 5  
## 31 23 1  
## 32 7 1  
## 33 4 0  
## 34 6 1  
## 35 4 0  
## 36 6 0  
## 37 9 1  
## 38 6 1  
## 39 7 0  
## 40 9 0  
## 41 5 0  
## 42 6 0  
## 43 4 1  
## 44 4 0  
## 45 2 0  
## 46 2 0  
## 47 3 0  
## 48 3 0  
## 49 2 0  
## 50 3 0  
## 51 2 0  
## 52 2 0  
## 53 1 0  
## 54 1 0  
## 55 3 0  
## 56 3 0  
## 57 1 0  
## 58 4 0  
## 59 2 0  
## 60 2 0  
## 61 1 0  
## 62 4 0  
## 63 2 0  
## 64 2 0  
## 65 1 0  
## 67 2 0  
## 68 2 0  
## 69 1 0  
## 70 2 0  
## 71 1 0  
## 74 3 0  
## 77 1 0  
## 79 1 0

# Chi-sq test for variables independence

null: the 2 categorical variables are independent.

#action\_type and combined\_shot\_type  
tbl.1<-table(kobe$action\_type, kobe$combined\_shot\_type)  
chisq.test(tbl.1)

## Warning in chisq.test(tbl.1): Chi-squared approximation may be incorrect

##   
## Pearson's Chi-squared test  
##   
## data: tbl.1  
## X-squared = 128480, df = 270, p-value < 2.2e-16

#shot\_zone\_area and shot\_zone\_range  
tbl.2<-table(kobe$shot\_zone\_area, kobe$shot\_zone\_range)  
chisq.test(tbl.2)

## Warning in chisq.test(tbl.2): Chi-squared approximation may be incorrect

##   
## Pearson's Chi-squared test  
##   
## data: tbl.2  
## X-squared = 49563, df = 20, p-value < 2.2e-16

#shot\_zone\_area and shot\_zone\_basic  
tbl.3<-table(kobe$shot\_zone\_area, kobe$shot\_zone\_basic)  
chisq.test(tbl.3)

## Warning in chisq.test(tbl.3): Chi-squared approximation may be incorrect

##   
## Pearson's Chi-squared test  
##   
## data: tbl.3  
## X-squared = 46657, df = 30, p-value < 2.2e-16

#shot\_zone\_range and shot\_zone\_basic  
tbl.4<-table(kobe$shot\_zone\_range, kobe$shot\_zone\_basic)  
chisq.test(tbl.4)

## Warning in chisq.test(tbl.4): Chi-squared approximation may be incorrect

##   
## Pearson's Chi-squared test  
##   
## data: tbl.4  
## X-squared = 71026, df = 24, p-value < 2.2e-16

## logistic model

#shot\_made\_flag to distance, playoffs and coordinations  
glm.fit<-glm(shot\_made\_flag~loc\_x+loc\_y+lon+playoffs+shot\_distance+shot\_zone\_area+shot\_zone\_basic+shot\_zone\_range,family=binomial(link="logit"),data=kobe)  
  
summary(glm.fit)

##   
## Call:  
## glm(formula = shot\_made\_flag ~ loc\_x + loc\_y + lon + playoffs +   
## shot\_distance + shot\_zone\_area + shot\_zone\_basic + shot\_zone\_range,   
## family = binomial(link = "logit"), data = kobe)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.3968 -1.0375 -0.8924 1.2737 2.6654   
##   
## Coefficients: (3 not defined because of singularities)  
## Estimate Std. Error z value  
## (Intercept) -1.083e+01 9.360e+01 -0.116  
## loc\_x -3.576e-04 3.750e-04 -0.953  
## loc\_y -5.278e-05 4.551e-04 -0.116  
## lon NA NA NA  
## playoffs1 -8.875e-03 3.637e-02 -0.244  
## shot\_distance -4.092e-02 8.808e-03 -4.646  
## shot\_zone\_areaCenter(C) 1.116e+01 9.360e+01 0.119  
## shot\_zone\_areaLeft Side Center(LC) 1.109e+01 9.360e+01 0.119  
## shot\_zone\_areaLeft Side(L) 1.091e+01 9.360e+01 0.117  
## shot\_zone\_areaRight Side Center(RC) 1.126e+01 9.360e+01 0.120  
## shot\_zone\_areaRight Side(R) 1.103e+01 9.360e+01 0.118  
## shot\_zone\_basicBackcourt 9.095e+00 9.360e+01 0.097  
## shot\_zone\_basicIn The Paint (Non-RA) 1.297e-01 1.045e-01 1.241  
## shot\_zone\_basicLeft Corner 3 2.533e-01 1.588e-01 1.595  
## shot\_zone\_basicMid-Range 8.739e-02 6.731e-02 1.298  
## shot\_zone\_basicRestricted Area 6.513e-01 1.480e-01 4.402  
## shot\_zone\_basicRight Corner 3 1.502e-01 1.405e-01 1.069  
## shot\_zone\_range24+ ft. NA NA NA  
## shot\_zone\_range8-16 ft. -6.337e-02 6.541e-02 -0.969  
## shot\_zone\_rangeBack Court Shot NA NA NA  
## shot\_zone\_rangeLess Than 8 ft. -4.863e-01 1.153e-01 -4.216  
## Pr(>|z|)   
## (Intercept) 0.908   
## loc\_x 0.340   
## loc\_y 0.908   
## lon NA   
## playoffs1 0.807   
## shot\_distance 3.38e-06 \*\*\*  
## shot\_zone\_areaCenter(C) 0.905   
## shot\_zone\_areaLeft Side Center(LC) 0.906   
## shot\_zone\_areaLeft Side(L) 0.907   
## shot\_zone\_areaRight Side Center(RC) 0.904   
## shot\_zone\_areaRight Side(R) 0.906   
## shot\_zone\_basicBackcourt 0.923   
## shot\_zone\_basicIn The Paint (Non-RA) 0.215   
## shot\_zone\_basicLeft Corner 3 0.111   
## shot\_zone\_basicMid-Range 0.194   
## shot\_zone\_basicRestricted Area 1.07e-05 \*\*\*  
## shot\_zone\_basicRight Corner 3 0.285   
## shot\_zone\_range24+ ft. NA   
## shot\_zone\_range8-16 ft. 0.333   
## shot\_zone\_rangeBack Court Shot NA   
## shot\_zone\_rangeLess Than 8 ft. 2.49e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 35325 on 25696 degrees of freedom  
## Residual deviance: 34137 on 25679 degrees of freedom  
## AIC: 34173  
##   
## Number of Fisher Scoring iterations: 11

VIF Not very useful in logistic regression, typically notr performed for mixed data set with categorical variables.

## Model Selection(s): Stepwise

Both stepwise models resulted in the same AIC and predictors. Since the question stated that playoffs is a factor in him making the shot, therefore the prediction should use the base.logit (not the two models shown below)

#Stepwise based on the first logit regression  
stepwise<- stepAIC(glm.fit,direction="both",trace = FALSE)  
summary(stepwise)

##   
## Call:  
## glm(formula = shot\_made\_flag ~ shot\_distance + shot\_zone\_area +   
## shot\_zone\_basic + shot\_zone\_range, family = binomial(link = "logit"),   
## data = kobe)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.3950 -1.0354 -0.8903 1.2769 2.6636   
##   
## Coefficients: (2 not defined because of singularities)  
## Estimate Std. Error z value  
## (Intercept) -10.838714 93.659381 -0.116  
## shot\_distance -0.041390 0.007897 -5.241  
## shot\_zone\_areaCenter(C) 11.170506 93.658902 0.119  
## shot\_zone\_areaLeft Side Center(LC) 11.152656 93.658900 0.119  
## shot\_zone\_areaLeft Side(L) 10.968063 93.658922 0.117  
## shot\_zone\_areaRight Side Center(RC) 11.224116 93.658900 0.120  
## shot\_zone\_areaRight Side(R) 11.000033 93.658921 0.117  
## shot\_zone\_basicBackcourt 9.100387 93.664304 0.097  
## shot\_zone\_basicIn The Paint (Non-RA) 0.129282 0.104359 1.239  
## shot\_zone\_basicLeft Corner 3 0.295059 0.147688 1.998  
## shot\_zone\_basicMid-Range 0.087444 0.067297 1.299  
## shot\_zone\_basicRestricted Area 0.651665 0.147753 4.411  
## shot\_zone\_basicRight Corner 3 0.120182 0.130610 0.920  
## shot\_zone\_range24+ ft. NA NA NA  
## shot\_zone\_range8-16 ft. -0.063601 0.064754 -0.982  
## shot\_zone\_rangeBack Court Shot NA NA NA  
## shot\_zone\_rangeLess Than 8 ft. -0.485171 0.114565 -4.235  
## Pr(>|z|)   
## (Intercept) 0.9079   
## shot\_distance 1.60e-07 \*\*\*  
## shot\_zone\_areaCenter(C) 0.9051   
## shot\_zone\_areaLeft Side Center(LC) 0.9052   
## shot\_zone\_areaLeft Side(L) 0.9068   
## shot\_zone\_areaRight Side Center(RC) 0.9046   
## shot\_zone\_areaRight Side(R) 0.9065   
## shot\_zone\_basicBackcourt 0.9226   
## shot\_zone\_basicIn The Paint (Non-RA) 0.2154   
## shot\_zone\_basicLeft Corner 3 0.0457 \*   
## shot\_zone\_basicMid-Range 0.1938   
## shot\_zone\_basicRestricted Area 1.03e-05 \*\*\*  
## shot\_zone\_basicRight Corner 3 0.3575   
## shot\_zone\_range24+ ft. NA   
## shot\_zone\_range8-16 ft. 0.3260   
## shot\_zone\_rangeBack Court Shot NA   
## shot\_zone\_rangeLess Than 8 ft. 2.29e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 35325 on 25696 degrees of freedom  
## Residual deviance: 34138 on 25682 degrees of freedom  
## AIC: 34168  
##   
## Number of Fisher Scoring iterations: 11

## Odds Ratio

#odds ratio calculation using library(epiDisplay)  
#based on the first logit function  
exp(cbind("Odds ratio" = coef(stepwise), confint.default(stepwise, level = 0.95)))

## Odds ratio 2.5 %  
## (Intercept) 1.962484e-05 3.713612e-85  
## shot\_distance 9.594545e-01 9.447178e-01  
## shot\_zone\_areaCenter(C) 7.100505e+04 1.344892e-75  
## shot\_zone\_areaLeft Side Center(LC) 6.974884e+04 1.321102e-75  
## shot\_zone\_areaLeft Side(L) 5.799215e+04 1.098375e-75  
## shot\_zone\_areaRight Side Center(RC) 7.491551e+04 1.418964e-75  
## shot\_zone\_areaRight Side(R) 5.987611e+04 1.134059e-75  
## shot\_zone\_basicBackcourt 8.958763e+03 1.678990e-76  
## shot\_zone\_basicIn The Paint (Non-RA) 1.138011e+00 9.275034e-01  
## shot\_zone\_basicLeft Corner 3 1.343206e+00 1.005611e+00  
## shot\_zone\_basicMid-Range 1.091381e+00 9.565173e-01  
## shot\_zone\_basicRestricted Area 1.918734e+00 1.436307e+00  
## shot\_zone\_basicRight Corner 3 1.127703e+00 8.730093e-01  
## shot\_zone\_range24+ ft. NA NA  
## shot\_zone\_range8-16 ft. 9.383795e-01 8.265318e-01  
## shot\_zone\_rangeBack Court Shot NA NA  
## shot\_zone\_rangeLess Than 8 ft. 6.155917e-01 4.917847e-01  
## 97.5 %  
## (Intercept) 1.037088e+75  
## shot\_distance 9.744210e-01  
## shot\_zone\_areaCenter(C) 3.748791e+84  
## shot\_zone\_areaLeft Side Center(LC) 3.682456e+84  
## shot\_zone\_areaLeft Side(L) 3.061877e+84  
## shot\_zone\_areaRight Side Center(RC) 3.955233e+84  
## shot\_zone\_areaRight Side(R) 3.161341e+84  
## shot\_zone\_basicBackcourt 4.780221e+83  
## shot\_zone\_basicIn The Paint (Non-RA) 1.396295e+00  
## shot\_zone\_basicLeft Corner 3 1.794135e+00  
## shot\_zone\_basicMid-Range 1.245259e+00  
## shot\_zone\_basicRestricted Area 2.563198e+00  
## shot\_zone\_basicRight Corner 3 1.456700e+00  
## shot\_zone\_range24+ ft. NA  
## shot\_zone\_range8-16 ft. 1.065363e+00  
## shot\_zone\_rangeBack Court Shot NA  
## shot\_zone\_rangeLess Than 8 ft. 7.705673e-01

LDA is not performed for logistic regression model.

## Prediction

#load the project2Pred.xlsx  
pred\_kobe <- read\_excel("C:\\Users\\Yat\\Documents\\MSDS\\MSDS 6372\\Project 2\\project2Pred.xlsx")  
#Take out an ID column from when data was imported  
pred\_kobe<-pred\_kobe[,c(2:29)]  
  
#Re-code the character columns into factors  
pred\_kobe<-as.data.frame(unclass(pred\_kobe))  
pred\_kobe$shot\_made\_flag<-as.factor(pred\_kobe$shot\_made\_flag)  
pred\_kobe$shot\_type<-as.factor(pred\_kobe$shot\_type)  
pred\_kobe$shot\_zone\_area<-as.factor(pred\_kobe$shot\_zone\_area)  
pred\_kobe$shot\_zone\_basic<-as.factor(pred\_kobe$shot\_zone\_basic)  
pred\_kobe$shot\_zone\_range<-as.factor(pred\_kobe$shot\_zone\_range)  
pred\_kobe$game\_date<-as.factor(pred\_kobe$game\_date)  
pred\_kobe$season<-as.factor(pred\_kobe$season)  
pred\_kobe$period<-as.factor(pred\_kobe$period) #change period into factor?  
pred\_kobe$playoffs<-as.factor(pred\_kobe$playoffs)  
  
#subsetting variables (sans IDs)  
pred\_kobe<-pred\_kobe[,c(1:2,5:20,22:25,27:28)]  
pred\_kobe<-na.omit(pred\_kobe)  
  
  
#data check  
head(pred\_kobe)

## action\_type combined\_shot\_type lat loc\_x loc\_y lon  
## 1 Jump Shot Jump Shot 33.9723 167 72 -118.1028  
## 2 Jump Shot Jump Shot 34.0163 1 28 -118.2688  
## 3 Driving Layup Shot Layup 34.0443 0 0 -118.2698  
## 4 Driving Layup Shot Layup 34.0443 0 0 -118.2698  
## 5 Jump Shot Jump Shot 33.9683 163 76 -118.1068  
## 6 Jump Shot Jump Shot 33.8503 70 194 -118.1998  
## minutes\_remaining period playoffs season seconds\_remaining  
## 1 10 1 0 2000-01 27  
## 2 8 3 0 2000-01 5  
## 3 0 1 0 2000-01 1  
## 4 10 3 0 2000-01 46  
## 5 11 1 0 2000-01 26  
## 6 10 1 0 2000-01 58  
## shot\_distance shot\_made\_flag shot\_type shot\_zone\_area  
## 1 18 NA 2PT Field Goal Right Side(R)  
## 2 2 NA 2PT Field Goal Center(C)  
## 3 0 NA 2PT Field Goal Center(C)  
## 4 0 NA 2PT Field Goal Center(C)  
## 5 17 NA 2PT Field Goal Right Side(R)  
## 6 20 NA 2PT Field Goal Right Side Center(RC)  
## shot\_zone\_basic shot\_zone\_range team\_id game\_date matchup  
## 1 Mid-Range 16-24 ft. 1610612747 2000-10-31 LAL @ POR  
## 2 Restricted Area Less Than 8 ft. 1610612747 2000-10-31 LAL @ POR  
## 3 Restricted Area Less Than 8 ft. 1610612747 2000-11-01 LAL vs. UTA  
## 4 Restricted Area Less Than 8 ft. 1610612747 2000-11-01 LAL vs. UTA  
## 5 Mid-Range 16-24 ft. 1610612747 2000-11-04 LAL @ VAN  
## 6 Mid-Range 16-24 ft. 1610612747 2000-11-04 LAL @ VAN  
## opponent shot\_id arena\_temp avgnoisedb  
## 1 POR 1 69 94.06  
## 2 POR 8 69 94.06  
## 3 UTA 17 69 95.71  
## 4 UTA 20 69 95.71  
## 5 VAN 33 72 95.09  
## 6 VAN 34 72 95.09

str(pred\_kobe)

## 'data.frame': 5000 obs. of 24 variables:  
## $ action\_type : Factor w/ 49 levels "Alley Oop Dunk Shot",..: 24 24 11 11 24 24 25 25 31 24 ...  
## $ combined\_shot\_type: Factor w/ 6 levels "Bank Shot","Dunk",..: 4 4 5 5 4 4 5 5 5 4 ...  
## $ lat : num 34 34 34 34 34 ...  
## $ loc\_x : num 167 1 0 0 163 70 1 -12 1 -117 ...  
## $ loc\_y : num 72 28 0 0 76 194 19 15 4 116 ...  
## $ lon : num -118 -118 -118 -118 -118 ...  
## $ minutes\_remaining : num 10 8 0 10 11 10 7 5 4 5 ...  
## $ period : Factor w/ 6 levels "1","2","3","4",..: 1 3 1 3 1 1 1 1 1 2 ...  
## $ playoffs : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ season : Factor w/ 20 levels "1996-97","1997-98",..: 5 5 5 5 5 5 5 5 5 5 ...  
## $ seconds\_remaining : num 27 5 1 46 26 58 33 58 9 33 ...  
## $ shot\_distance : num 18 2 0 0 17 20 1 1 0 16 ...  
## $ shot\_made\_flag : Factor w/ 1 level "NA": 1 1 1 1 1 1 1 1 1 1 ...  
## $ shot\_type : Factor w/ 2 levels "2PT Field Goal",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ shot\_zone\_area : Factor w/ 6 levels "Back Court(BC)",..: 6 2 2 2 6 5 2 2 2 3 ...  
## $ shot\_zone\_basic : Factor w/ 7 levels "Above the Break 3",..: 5 6 6 6 5 5 6 6 6 5 ...  
## $ shot\_zone\_range : Factor w/ 5 levels "16-24 ft.","24+ ft.",..: 1 5 5 5 1 1 5 5 5 1 ...  
## $ team\_id : num 1.61e+09 1.61e+09 1.61e+09 1.61e+09 1.61e+09 ...  
## $ game\_date : Factor w/ 1457 levels "1996-11-08","1996-11-17",..: 261 261 262 262 263 263 263 263 263 263 ...  
## $ matchup : Factor w/ 74 levels "LAL @ ATL","LAL @ BKN",..: 29 29 72 72 36 36 36 36 36 36 ...  
## $ opponent : Factor w/ 33 levels "ATL","BKN","BOS",..: 26 26 31 31 32 32 32 32 32 32 ...  
## $ shot\_id : num 1 8 17 20 33 34 35 36 37 38 ...  
## $ arena\_temp : num 69 69 69 69 72 72 72 72 72 72 ...  
## $ avgnoisedb : num 94.1 94.1 95.7 95.7 95.1 ...

#dataframe for the shot\_made\_flag column in the pred\_kobe  
pred\_shot\_made\_flag<-pred\_kobe  
  
#what library to use for AUC, Mis-Classification Rate, Sensitivity, Specificity and objective / loss function  
predictions <- predict(stepwise, pred\_shot\_made\_flag, type="response")

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type =  
## ifelse(type == : prediction from a rank-deficient fit may be misleading

predictions<-as.data.frame(ifelse(predictions>0.5,"1","0"))  
colnames(predictions) <- "Predicted shot\_made\_flag"  
  
  
  
#actual and predicted values from model  
threshold=0.5  
predicted\_values<-ifelse(predict(stepwise,type="response")>threshold,1,0)  
  
actual\_values<-stepwise$y  
  
  
#confusion matrix using the training set  
conf\_matrix<-table(predicted\_values,actual\_values)  
conf\_matrix

## actual\_values  
## predicted\_values 0 1  
## 0 11764 7630  
## 1 2468 3835

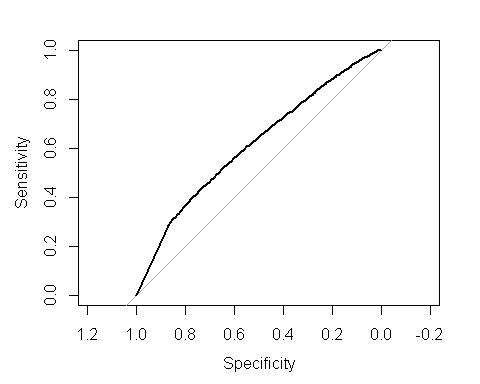
#Sensitivity of the model  
sensitivity(conf\_matrix)

## [1] 0.826588

#specificity of the model  
specificity(conf\_matrix)

## [1] 0.3344963

predicted\_prob<-predict(stepwise,type="response")  
roccurve <- roc(actual\_values, predicted\_prob)  
plot(roccurve)



#AUC   
auc(roccurve)

## Area under the curve: 0.6144

#Log Loss Function  
LogLoss(predicted\_values,actual\_values)

## [1] 13.57258