Factor Analysis

> head(AustralianOpen\_Finalists\_allstats)

PlayerName Year total\_matchs winpercentage MatchID Round AvgMinsPerGame

1: Andre Agassi 2000 7 1 m\_2000\_A\_114 4th Round 3.84

2: Andre Agassi 2000 7 1 m\_2000\_A\_122 Quarterfinals 3.32

3: Andre Agassi 2000 7 1 m\_2000\_A\_73 2nd Round 3.44

4: Andre Agassi 2000 7 1 m\_2000\_A\_124 Semifinals 3.50

5: Andre Agassi 2000 7 1 m\_2000\_A\_44 1st Round 3.48

6: Andre Agassi 2000 7 1 m\_2000\_A\_97 3rd Round 3.39

> str(AustralianOpen\_Finalists\_allstats)

Classes ‘data.table’ and 'data.frame': 277 obs. of 27 variables:

$ PlayerName : chr "Andre Agassi" "Andre Agassi" "Andre Agassi" "Andre Agassi" ...

$ Year : num 2000 2000 2000 2000 2000 ...

$ total\_matchs : int 7 7 7 7 7 7 7 7 7 7 ...

$ winpercentage : num 1 1 1 1 1 1 1 1 1 1 ...

$ MatchID : chr "m\_2000\_A\_114" "m\_2000\_A\_122" "m\_2000\_A\_73" "m\_2000\_A\_124" ...

$ Round : chr "4th Round" "Quarterfinals" "2nd Round" "Semifinals" ...

$ AvgMinsPerGame : num 3.84 3.32 3.44 3.5 3.48 3.39 3.86 3.81 4 3.75 ...

$ AvgSecsPerPoint : num 37.9 35.1 37.2 34.5 37.3 37 35 38.3 32.6 33.3 ...

$ AvgMinsPerSet : num 41.3 31 31 35 29 31.7 34.8 39.3 68 33.8 ...

$ Tournament : chr "Australian Open" "Australian Open" "Australian Open" "Australian Open" ...

$ TotalMatchMins : num 165 93 93 175 87 95 139 118 68 135 ...

$ Points : num 0 0 0 0 0 0 0 0 0 0 ...

$ Age : num 30 30 30 30 30 30 30 31 31 31 ...

$ Rank : num 1 1 1 1 1 1 1 6 6 6 ...

$ Winner : logi TRUE TRUE TRUE TRUE TRUE TRUE ...

$ TotalSets : num 3 3 3 3 3 3 3 3 1 3 ...

$ avgOdds : num 0 0 0 0 0 0 0 0 0 0 ...

$ maxOdds : num 0 0 0 0 0 0 0 0 0 0 ...

$ SP\_Percent : num 0.709 0.574 0.581 0.69 0.551 ...

$ RP\_Percent : num 0.291 0.426 0.419 0.31 0.449 ...

$ BP\_Win\_Percentage : num 0.778 0.5 0 0.889 1 ...

$ Aces : num 8 6 8 13 6 8 9 6 8 5 ...

$ firstServeReturnsWon : num 11 13 12 19 18 14 23 30 19 33 ...

$ SecondServeReturnsWon: num 28 27 27 29 22 25 27 18 16 32 ...

$ FirstServesIn : num 96 45 50 101 40 35 77 55 40 77 ...

$ DoubleFaults : num 4 1 1 3 1 3 5 0 2 2 ...

$ FirstServePercentage : num 0.691 0.662 0.658 0.682 0.656 ...

- attr(\*, ".internal.selfref")=<externalptr>

- attr(\*, "sorted")= chr "PlayerName" "Year"

> summary(AustralianOpen\_Finalists\_allstats)

PlayerName Year total\_matchs winpercentage MatchID

Length:277 Min. :2000 Min. :6.000 Min. :0.8333 Length:277

Class :character 1st Qu.:2005 1st Qu.:7.000 1st Qu.:0.8571 Class :character

Mode :character Median :2009 Median :7.000 Median :0.8571 Mode :character

Mean :2009 Mean :6.935 Mean :0.9278

3rd Qu.:2014 3rd Qu.:7.000 3rd Qu.:1.0000

Max. :2019 Max. :7.000 Max. :1.0000

Round AvgMinsPerGame AvgSecsPerPoint AvgMinsPerSet Tournament

Length:277 Min. :2.930 Min. :30.20 Min. : 0.00 Length:277

Class :character 1st Qu.:3.860 1st Qu.:37.60 1st Qu.:34.70 Class :character

Mode :character Median :4.280 Median :40.70 Median :40.60 Mode :character

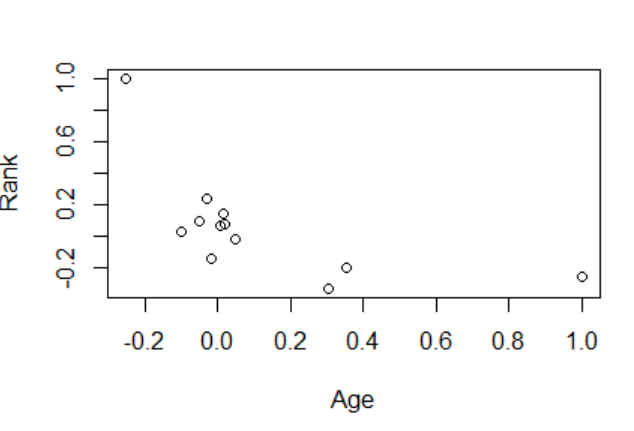
Mean :4.361 Mean :41.25 Mean :41.29

3rd Qu.:4.700 3rd Qu.:44.30 3rd Qu.:47.30

Max. :9.030 Max. :75.00 Max. :93.30

cor\_tennis<-cor(AustralianOpen\_Finalists\_allstats\_Numeric)

> plot(cor\_tennis)



> AustralianOpen\_Finalists\_allstats\_pca <- prcomp(AustralianOpen\_Finalists\_allstats\_Numeric, scale=TRUE)

> plot(AustralianOpen\_Finalists\_allstats\_pca)

> summary(AustralianOpen\_Finalists\_allstats\_pca)

Importance of components:

PC1 PC2 PC3 PC4 PC5 PC6 PC7 PC8 PC9

Standard deviation 1.7420 1.4031 1.3450 1.06024 0.94928 0.87265 0.82817 0.81273 0.75323

Proportion of Variance 0.2529 0.1641 0.1507 0.09368 0.07509 0.06346 0.05715 0.05504 0.04728

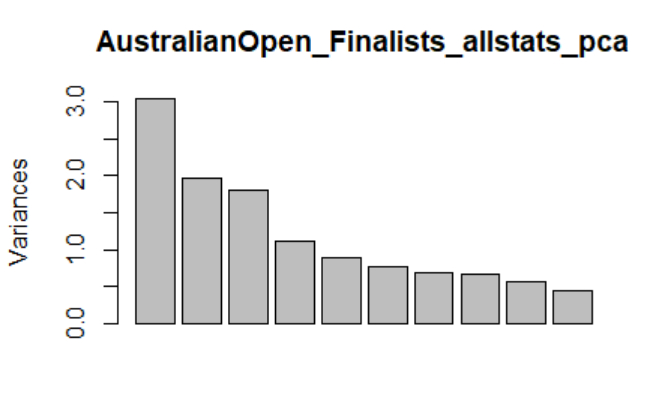
Cumulative Proportion 0.2529 0.4169 0.5677 0.66136 0.73645 0.79991 0.85707 0.91211 0.95939

PC10 PC11 PC12

Standard deviation 0.66214 0.22105 8.502e-16

Proportion of Variance 0.03654 0.00407 0.000e+00

Cumulative Proportion 0.99593 1.00000 1.000e+00



> eigen\_AO\_Finalists

PC1 PC2 PC3 PC4 PC5 PC6 PC7

3.034501e+00 1.968804e+00 1.808901e+00 1.124111e+00 9.011337e-01 7.615134e-01 6.858578e-01

PC8 PC9 PC10 PC11 PC12

6.605265e-01 5.673569e-01 4.384332e-01 4.886229e-02 7.227796e-31

> sumlambdas<-sum(eigen\_AO\_Finalists)

> sumlambdas

[1] 12

> propvar<-eigen\_AO\_Finalists/sumlambdas

> propvar

PC1 PC2 PC3 PC4 PC5 PC6 PC7

2.528751e-01 1.640670e-01 1.507418e-01 9.367588e-02 7.509447e-02 6.345945e-02 5.715482e-02

PC8 PC9 PC10 PC11 PC12

5.504387e-02 4.727975e-02 3.653610e-02 4.071857e-03 6.023163e-32

> cumvar\_AO\_Finalists<-cumsum(propvar)

> cumvar\_AO\_Finalists

PC1 PC2 PC3 PC4 PC5 PC6 PC7 PC8 PC9

0.2528751 0.4169420 0.5676838 0.6613597 0.7364542 0.7999136 0.8570684 0.9121123 0.9593920

PC10 PC11 PC12

0.9959281 1.0000000 1.0000000

> matlambdas<-rbind(eigen\_AO\_Finalists,propvar,cumvar\_AO\_Finalists)

> rownames(matlambdas)<-c("Eigenvalues","Prop.variance","Cum.propvariance")

> matlambdas

PC1 PC2 PC3 PC4 PC5 PC6 PC7

Eigenvalues 3.0345008 1.968804 1.8089011 1.12411062 0.90113368 0.76151336 0.68585783

Prop.variance 0.2528751 0.164067 0.1507418 0.09367588 0.07509447 0.06345945 0.05715482

Cum.propvariance 0.2528751 0.416942 0.5676838 0.66135968 0.73645416 0.79991360 0.85706842

PC8 PC9 PC10 PC11 PC12

Eigenvalues 0.66052646 0.56735695 0.4384332 0.048862288 7.227796e-31

Prop.variance 0.05504387 0.04727975 0.0365361 0.004071857 6.023163e-32

Cum.propvariance 0.91211229 0.95939204 0.9959281 1.000000000 1.000000e+00

> print(AustralianOpen\_Finalists\_allstats\_pca)

Standard deviations (1, .., p=12):

[1] 1.741982e+00 1.403141e+00 1.344954e+00 1.060241e+00 9.492806e-01 8.726473e-01

[7] 8.281653e-01 8.127278e-01 7.532310e-01 6.621429e-01 2.210482e-01 8.501644e-16

Rotation (n x k) = (12 x 12):

PC1 PC2 PC3 PC4 PC5

Age -0.00415178 0.42830942 -0.232554474 0.314609668 -0.04632629

Rank 0.13630501 -0.38197446 0.243595669 0.155808021 0.34813788

avgOdds 0.11807701 0.34188696 -0.274509489 0.476805302 -0.14611186

SP\_Percent 0.48391843 0.19453812 0.308779847 -0.007946174 -0.03292276

RP\_Percent -0.48391843 -0.19453812 -0.308779847 0.007946174 0.03292276

BP\_Win\_Percentage 0.26061584 0.18226722 -0.006836468 -0.395263686 0.11221296

Aces 0.31732681 -0.19283950 -0.016447010 0.370417984 0.37960882

firstServeReturnsWon 0.11056995 -0.24004315 -0.506991061 0.241247978 0.26234179

SecondServeReturnsWon 0.13556848 -0.30358565 -0.404359859 -0.397499499 -0.08812446

FirstServesIn 0.46875952 -0.01812074 -0.326892184 -0.188532278 0.10380643

DoubleFaults 0.27786436 -0.21598416 -0.157093786 0.054974387 -0.70441043

FirstServePercentage -0.05423925 0.45923720 -0.261023158 -0.317897990 0.33598190

# Multiplying each column of the eigenvector’s matrix by the square-root of the corresponding eigenvalue in order to get the factor loadings

> unrot.fact.AO\_Finalist <- sweep(pcafactors.AO\_Finalists,MARGIN=2,AustralianOpen\_Finalists\_allstats\_pca$sdev[1:5],`\*`)

> unrot.fact.AO\_Finalist

PC1 PC2 PC3 PC4 PC5

Age -0.007232325 0.60097835 -0.312775054 0.333562015 -0.04397665

Rank 0.237440856 -0.53596388 0.327624953 0.165194025 0.33048054

avgOdds 0.205688019 0.47971549 -0.369202617 0.505528449 -0.13870116

SP\_Percent 0.842977134 0.27296435 0.415294670 -0.008424858 -0.03125294

RP\_Percent -0.842977134 -0.27296435 -0.415294670 0.008424858 0.03125294

BP\_Win\_Percentage 0.453988059 0.25574654 -0.009194735 -0.419074698 0.10652159

Aces 0.552777545 -0.27058094 -0.022120471 0.392732271 0.36035529

firstServeReturnsWon 0.192610853 -0.33681431 -0.681879621 0.255780956 0.24903598

SecondServeReturnsWon 0.236157833 -0.42597337 -0.543845384 -0.421445198 -0.08365484

FirstServesIn 0.816570577 -0.02542595 -0.439654929 -0.199889618 0.09854144

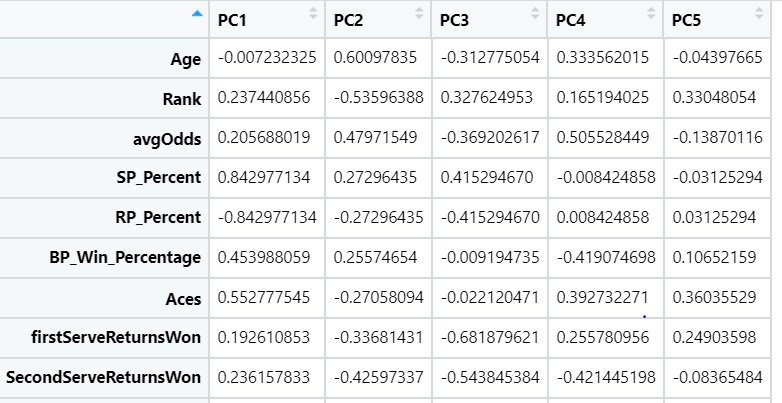
DoubleFaults 0.484034681 -0.30305615 -0.211283905 0.058286089 -0.66868316

FirstServePercentage -0.094483797 0.64437439 -0.351064123 -0.337048429 0.31894110

> # Performing the varimax rotation. The default in the varimax function is norm=TRUE thus, Kaiser normalization is carried out

> rot.fact.AO\_Finalist<-varimax(unrot.fact.AO\_Finalist)

> View(unrot.fact.AO\_Finalist)



> fact.load.AO\_Finalist <- rot.fact.AO\_Finalist$loadings[1:12,1:5]

> fact.load.AO\_Finalist

PC1 PC2 PC3 PC4 PC5

Age 0.01417243 0.72814415 0.11457330 0.06957223 0.154567083

Rank 0.07246725 -0.41564789 -0.62758870 0.06356468 -0.107504872

avgOdds 0.08176690 0.81725989 -0.05724614 -0.02261173 -0.071792323

SP\_Percent 0.94441580 0.06781892 -0.18709674 0.09197388 -0.136573701

RP\_Percent -0.94441580 -0.06781892 0.18709674 -0.09197388 0.136573701

BP\_Win\_Percentage 0.56347814 -0.04306566 0.13352199 -0.26351016 0.227900105

Aces 0.23943920 0.08421228 -0.75317506 -0.16083968 -0.076131758

firstServeReturnsWon -0.28675406 0.25351564 -0.46588950 -0.61594245 -0.003041345

SecondServeReturnsWon -0.05632020 -0.21434218 0.05986868 -0.80670947 -0.118856493

FirstServesIn 0.54306186 0.15813837 -0.21544506 -0.73753524 -0.009898431

DoubleFaults 0.20507606 0.09311755 0.05949905 -0.40314063 -0.777507787

FirstServePercentage 0.10563288 0.32105809 0.35105743 -0.18175837 0.701571521

> scale.AO\_Finalist <- scale(AustralianOpen\_Finalists\_allstats\_Numeric)

> scale.AO\_Finalist

Age Rank avgOdds SP\_Percent RP\_Percent BP\_Win\_Percentage

[1,] 0.86020070 -0.49690020 -0.6817896 1.846062706 -1.846062706 0.62668822

[2,] 0.86020070 -0.49690020 -0.6817896 -0.339407071 0.339407071 -0.24428066

[3,] 0.86020070 -0.49690020 -0.6817896 -0.239027116 0.239027116 -1.81202464

[4,] 0.86020070 -0.49690020 -0.6817896 1.543274808 -1.543274808 0.97507578

[5,] 0.86020070 -0.49690020 -0.6817896 -0.727893792 0.727893792 1.32346333

[6,] 0.86020070 -0.49690020 -0.6817896 -0.313099754 0.313099754 -1.81202464

[7,] 0.86020070 -0.49690020 -0.6817896 0.075496756 -0.075496756 -0.02031723

[8,] 1.12913478 -0.19715857 -0.6817896 -0.714242303 0.714242303 0.27830067

fit.pc

Principal Components Analysis

Call: principal(r = AustralianOpen\_Finalists\_allstats\_Numeric, nfactors = 5,

rotate = "varimax")

Standardized loadings (pattern matrix) based upon correlation matrix

RC1 RC4 RC2 RC3 RC5 h2 u2 com

Age 0.01 -0.07 0.73 -0.11 -0.15 0.57 0.428 1.2

Rank 0.07 -0.06 -0.42 0.63 0.11 0.59 0.413 1.9

avgOdds 0.08 0.02 0.82 0.06 0.07 0.68 0.316 1.0

SP\_Percent 0.94 -0.09 0.07 0.19 0.14 0.96 0.041 1.2

RP\_Percent -0.94 0.09 -0.07 -0.19 -0.14 0.96 0.041 1.2

BP\_Win\_Percentage 0.56 0.26 -0.04 -0.13 -0.23 0.46 0.541 1.9

Aces 0.24 0.16 0.08 0.75 0.08 0.66 0.337 1.4

firstServeReturnsWon -0.29 0.62 0.25 0.47 0.00 0.74 0.257 2.7

SecondServeReturnsWon -0.06 0.81 -0.21 -0.06 0.12 0.72 0.282 1.2

FirstServesIn 0.54 0.74 0.16 0.22 0.01 0.91 0.090 2.1

DoubleFaults 0.21 0.40 0.09 -0.06 0.78 0.82 0.179 1.7

FirstServePercentage 0.11 0.18 0.32 -0.35 -0.70 0.76 0.237 2.2

RC1 RC4 RC2 RC3 RC5

SS loadings 2.60 1.89 1.64 1.46 1.25

Proportion Var 0.22 0.16 0.14 0.12 0.10

Cumulative Var 0.22 0.37 0.51 0.63 0.74

Proportion Explained 0.29 0.21 0.19 0.17 0.14

Cumulative Proportion 0.29 0.51 0.69 0.86 1.00

Mean item complexity = 1.6

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 248.7 with prob < 9.6e-44

Fit based upon off diagonal values = 0.89

> round(fit.pc$values, 3)

[1] 3.035 1.969 1.809 1.124 0.901 0.762 0.686 0.661 0.567 0.438 0.049 0.000

> fit.pc$loadings

Loadings:

RC1 RC4 RC2 RC3 RC5

Age 0.728 -0.115 -0.155

Rank -0.416 0.628 0.108

avgOdds 0.817

SP\_Percent 0.944 0.187 0.137

RP\_Percent -0.944 -0.187 -0.137

BP\_Win\_Percentage 0.563 0.264 -0.134 -0.228

Aces 0.239 0.161 0.753

firstServeReturnsWon -0.287 0.616 0.254 0.466

SecondServeReturnsWon 0.807 -0.214 0.119

FirstServesIn 0.543 0.738 0.158 0.215

DoubleFaults 0.205 0.403 0.778

FirstServePercentage 0.106 0.182 0.321 -0.351 -0.702

RC1 RC4 RC2 RC3 RC5

SS loadings 2.604 1.891 1.636 1.459 1.247

Proportion Var 0.217 0.158 0.136 0.122 0.104

Cumulative Var 0.217 0.375 0.511 0.633 0.736

> for (i in c(1,3,2,4,5)) { print(fit.pc$loadings[[1,i]])}

[1] 0.01417243

[1] 0.7281442

[1] -0.06957223

[1] -0.1145733

[1] -0.1545671

> fit.pc$communality

Age Rank avgOdds SP\_Percent

0.5722531 0.5874800 0.6835421 0.9586374

RP\_Percent BP\_Win\_Percentage Aces firstServeReturnsWon

0.9586374 0.4585664 0.6633609 0.7429455

SecondServeReturnsWon FirstServesIn DoubleFaults FirstServePercentage

0.7176058 0.9103967 0.8213079 0.7627166

> fit.pc$scores

RC1 RC4 RC2 RC3 RC5

[1,] 5.177146171 0.912893568 0.55582435 -0.844971235 0.222439605

[2,] -1.209413101 -1.309093868 -0.27211345 -1.947875711 -1.186757152

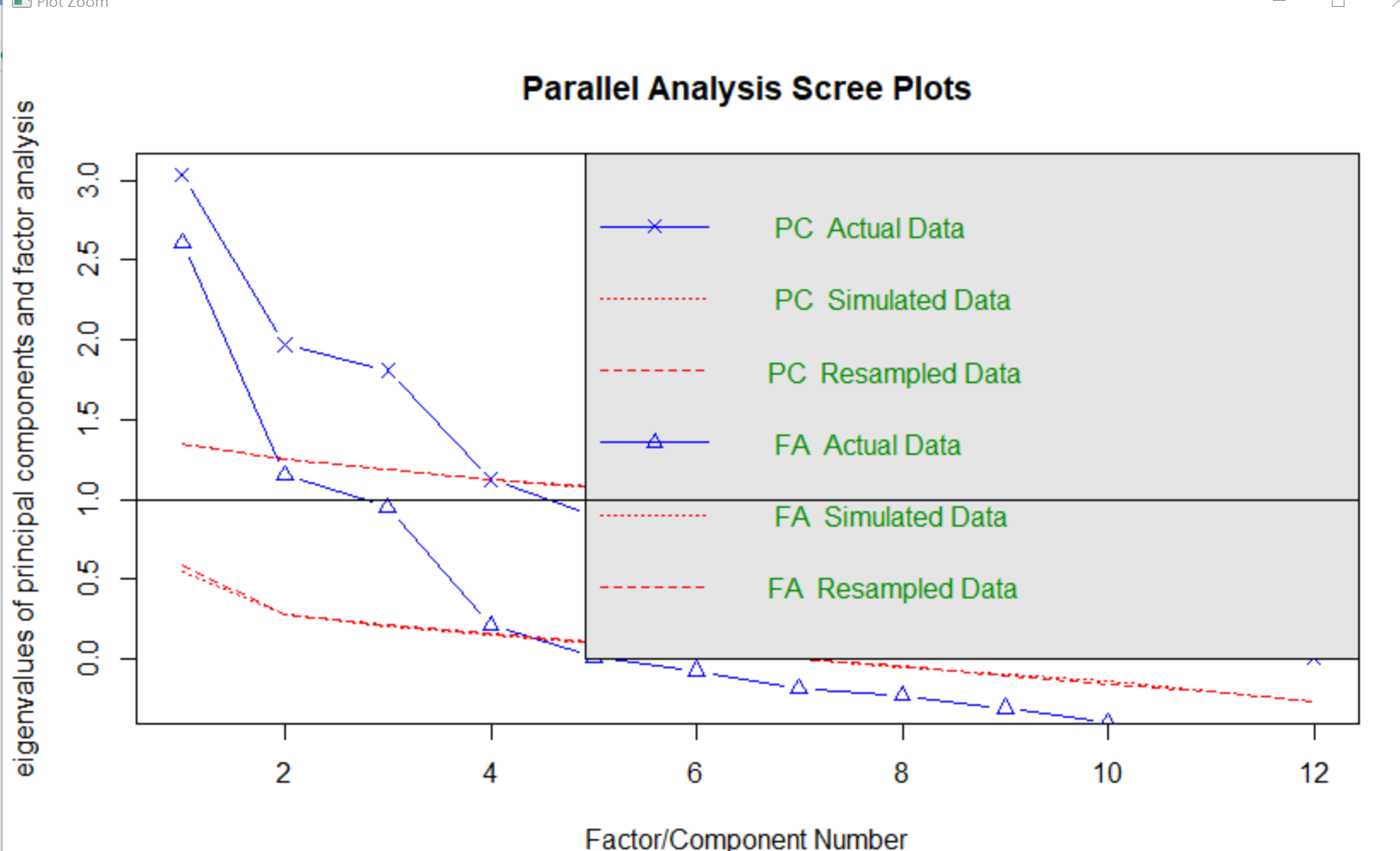
[3,] -1.661026519 -1.609687230 -0.17625175 -1.425147783 -0.735841174

[4,] 4.702099770 1.914816483 0.77426101 0.292659498 -0.178578817

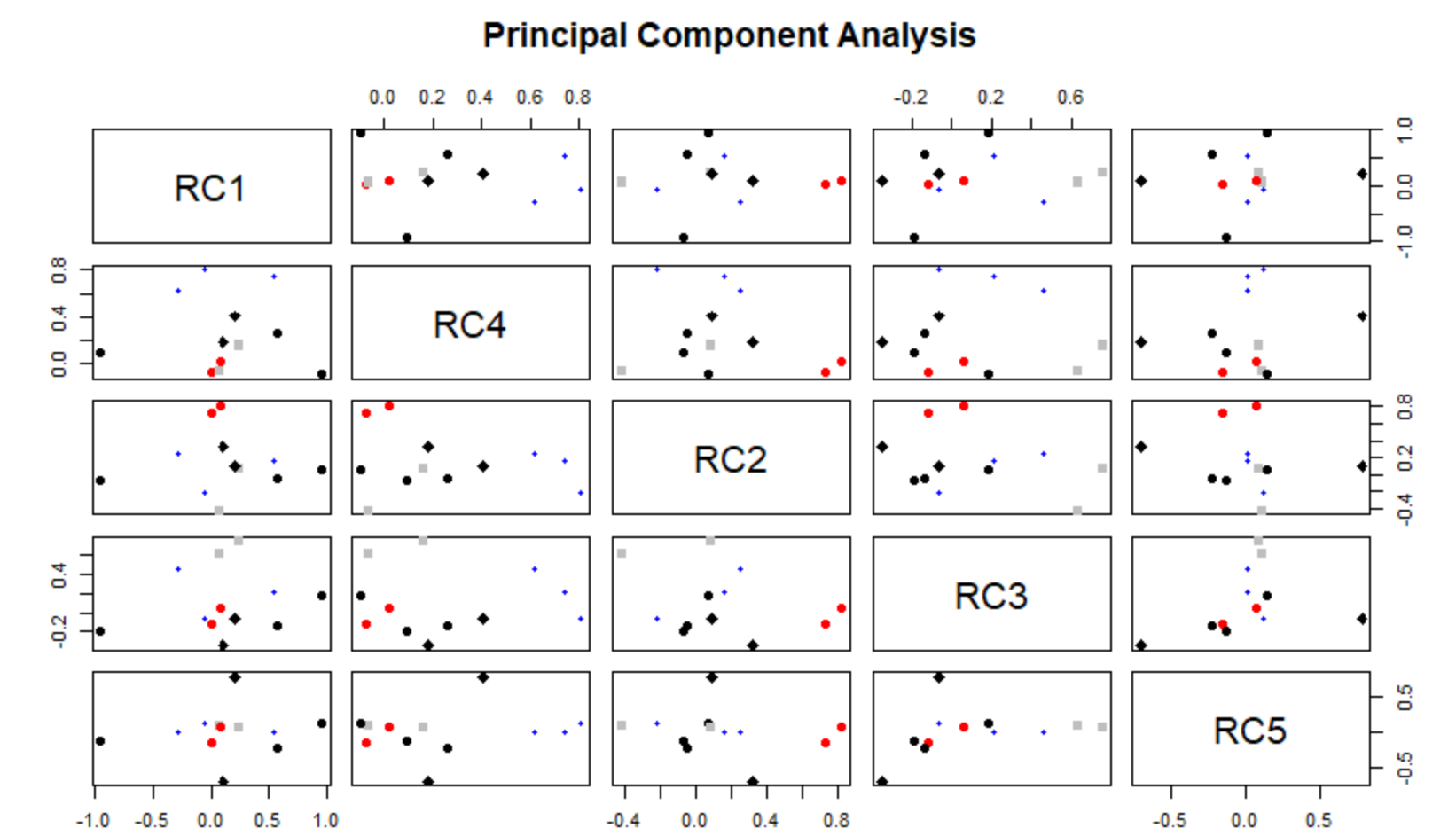
[5,] -1.348665943 -1.113806609 -0.14219933 -1.966954902 -1.670455839

[6,] -2.199329966 -2.030564070 -0.62733517 -0.910022167 1.237189439

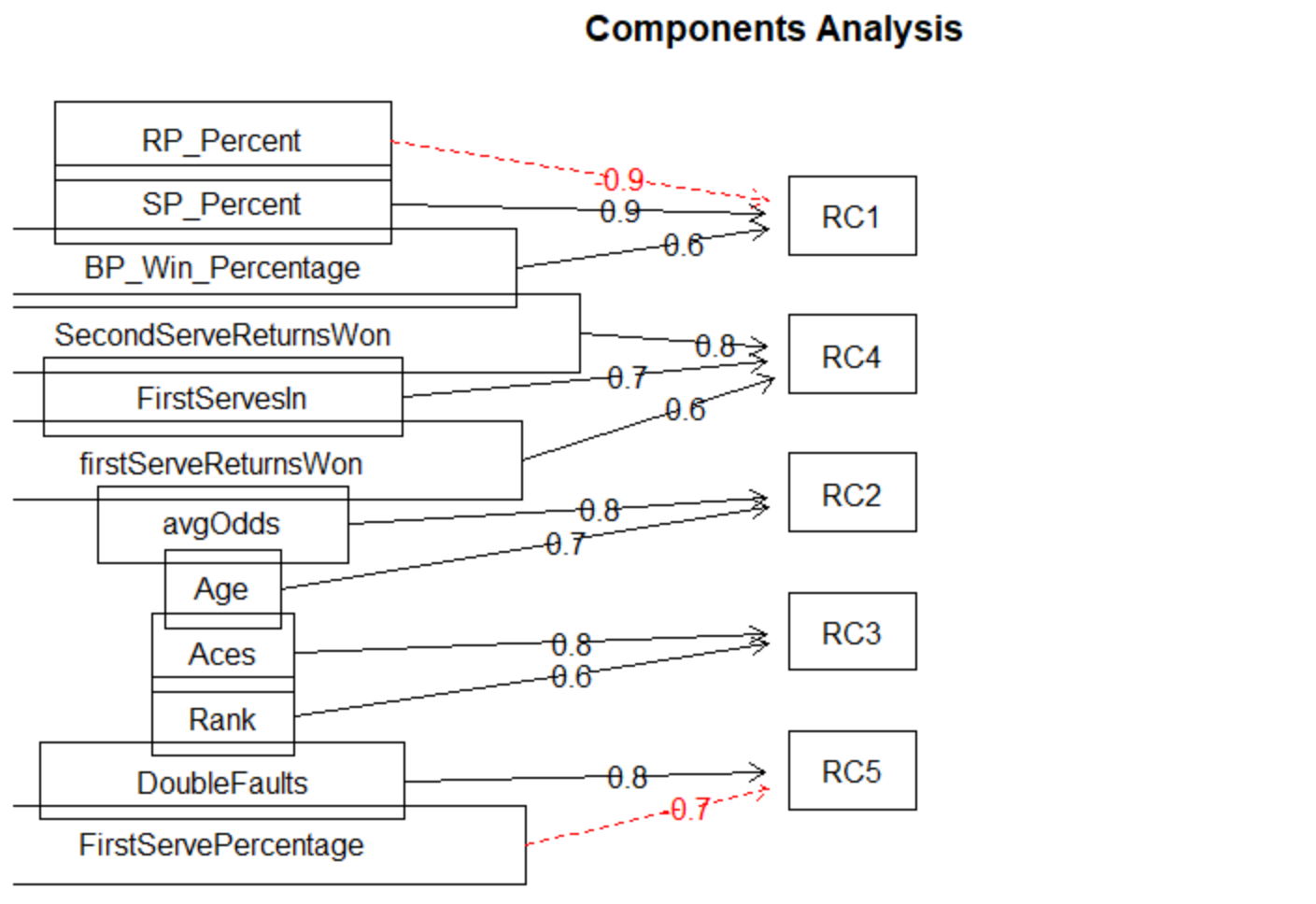
fa.parallel(AustralianOpen\_Finalists\_allstats\_Numeric) #factor recommendation



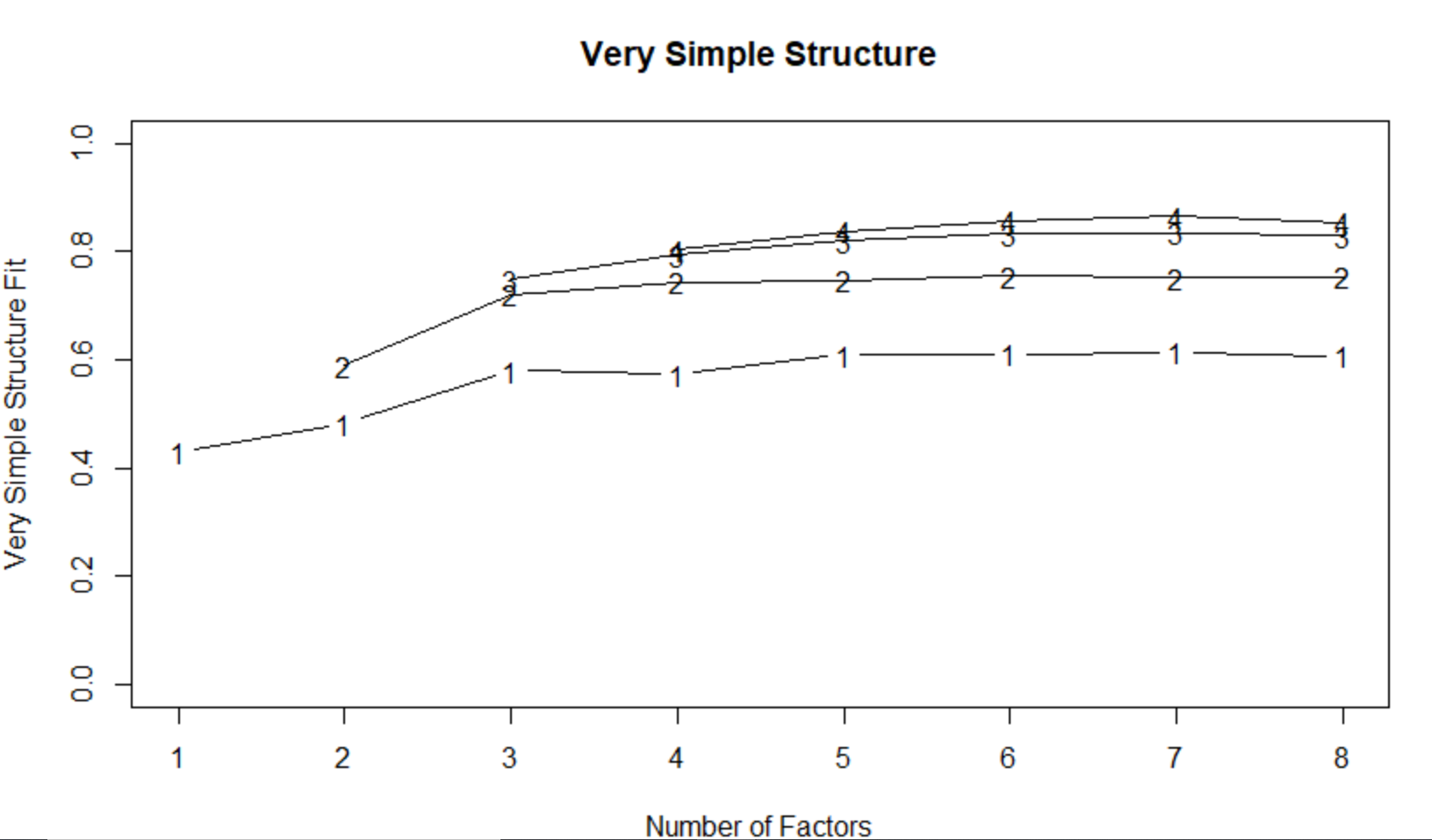
fa.plot(fit.pc) #see correlations within factors



> fa.diagram(fit.pc)#Visualize the realtionship



vss(AustralianOpen\_Finalists\_allstats\_Numeric)



Renaming the 5 factor groups:

RC1 as TotalPoints

RC2 as OddsVsAge

RC3 as RankVsAces,

RC4 as ReturnPointVsFirstServe

RC5 as ServeStats