Q\_1.R

Nadim

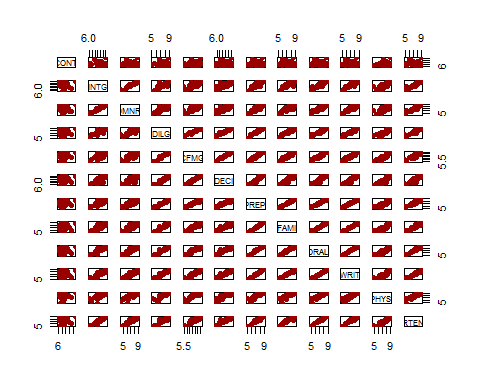
Thu Jan 31 02:10:15 2019

#####################################################################################  
# 1Q) Examine the USJudgeRatings data in the datasets library. This dataset contains   
# the ratings of 43 US Superior Court judges by attorneys. Each of the judges is evaluated   
# on each of 12 attributes such as demeanor, preparation for trial, sound rulings, and the   
# number of contacts each attorney had with that judge. See the R help file for more   
# information on this dataset.  
  
  
library('datasets')  
data\_<-USJudgeRatings  
#visulaising the datasetusing pairs plot  
pairs(data\_,pch=19,col="#990000")  
#visualising the dataset using correlation matrix  
library(GGally)

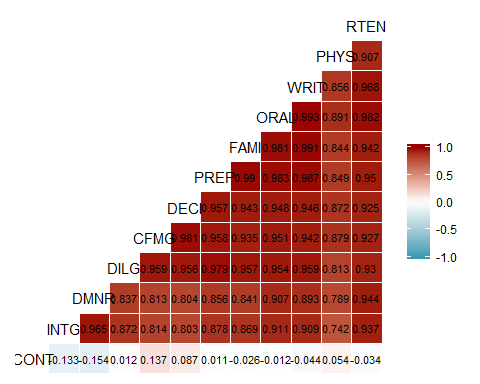
## Warning: package 'GGally' was built under R version 3.5.2

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 3.5.2



ggcorr(data\_, low = "#3B9AB2", mid = "#FFFFFF", high = "#990000",label = T, label\_color = "black",label\_size = 3, label\_round = 3)



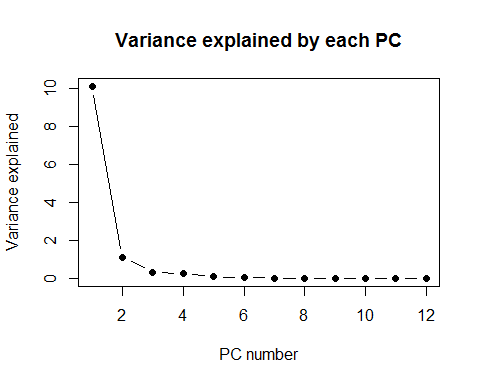
#to check if dataset has similar standard deviations, if not we need to   
#standardize them  
apply(data\_,2,sd)

## CONT INTG DMNR DILG CFMG DECI PREP   
## 0.9408768 0.7701447 1.1437054 0.9008978 0.8601102 0.8029362 0.9533702   
## FAMI ORAL WRIT PHYS RTEN   
## 0.9489868 1.0100437 0.9611328 0.9395753 1.1009711

##scaling  
std.data\_<-scale(data\_,center=T,scale=T)  
  
#Performing PCA  
  
data\_.pca<- prcomp(std.data\_)  
summary(data\_.pca)

## Importance of components:  
## PC1 PC2 PC3 PC4 PC5 PC6  
## Standard deviation 3.1833 1.05078 0.57698 0.50383 0.29061 0.19310  
## Proportion of Variance 0.8445 0.09201 0.02774 0.02115 0.00704 0.00311  
## Cumulative Proportion 0.8445 0.93647 0.96421 0.98537 0.99240 0.99551  
## PC7 PC8 PC9 PC10 PC11 PC12  
## Standard deviation 0.14030 0.12416 0.08851 0.07491 0.05708 0.04539  
## Proportion of Variance 0.00164 0.00128 0.00065 0.00047 0.00027 0.00017  
## Cumulative Proportion 0.99715 0.99844 0.99909 0.99956 0.99983 1.00000

lambda<- data\_.pca$sdev^2  
plot(lambda, type="b", pch = 19, main = "Variance explained by each PC",  
 xlab = "PC number", ylab = "Variance explained")



#first two components explain 93.6% of variability, so we can retain those two  
#since the loadings are the same as eigen vectors  
  
a<-round(data\_.pca$rotation,3)  
a[,1:2]

## PC1 PC2  
## CONT 0.003 -0.933  
## INTG -0.289 0.182  
## DMNR -0.287 0.198  
## DILG -0.304 -0.036  
## CFMG -0.303 -0.168  
## DECI -0.302 -0.128  
## PREP -0.309 -0.032  
## FAMI -0.307 0.001  
## ORAL -0.313 0.004  
## WRIT -0.311 0.031  
## PHYS -0.281 -0.089  
## RTEN -0.310 0.039

####Interpreting the loadings of the PCs  
####(i)First eigen shows component for every variable except CONT and second the PC   
####shows the opposite to the first  
####(ii)the same can be observed in the first plot where cont does not show any correlation  
  
###We can see from the matrix in part(a) that the covariance of cont with the rest\  
###of the variables is negligible. Hence the variance defined by CONT variable will have   
####have to be defined by a PC which has high weightage to the CONT variable. This   
###can be confirmed with the PC loadings