Loading data

```r
college.data <- read.csv("college_data_final.csv")
college.data$CC_BASIC <- as.factor(college.data$CC_BASIC) # converting the types of universities as fac

## str(college.data)
## nrow(college.data)
## ncol(college.data)
## head(college.data)
## tail(college.data)
```

Explanatory Data Analysis

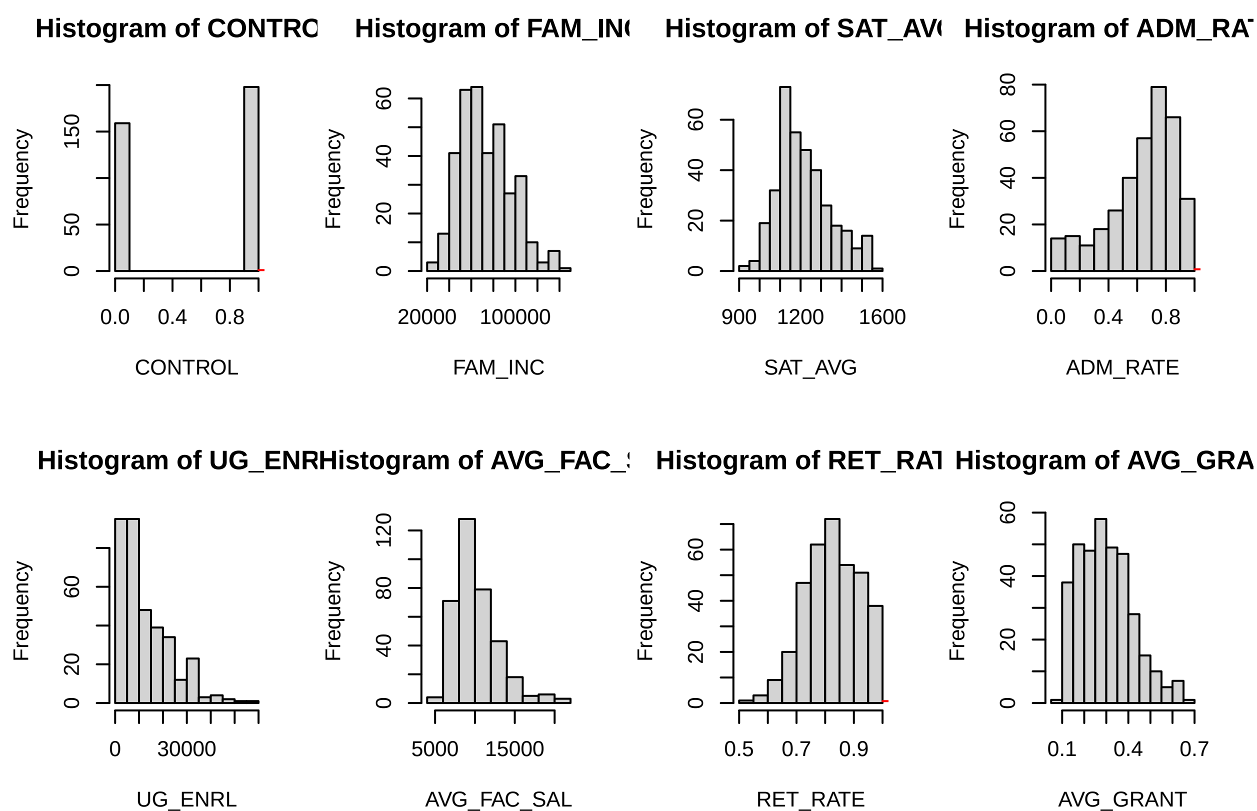
We perform an explanatory data analysis to get a better understanding of the variables and relationships among them.

```r
summary(college.data)
```

```
##   INST_NAME            STATE             STATE_FIPS    CC_BASIC
##  Length:357         Length:357         Min.   : 1.00   R1:129
##  Class :character   Class :character   1st Qu.:17.00   R2:115
##  Mode  :character   Mode  :character   Median :29.00   R3:113
##                                        Mean   :29.08
##                                        3rd Qu.:42.00
##                                        Max.   :56.00
##     CONTROL          FAM_INC          SAT_AVG        ADM_RATE
##  Min.   :0.0000   Min.   : 27098   Min.   : 920   Min.   :0.0436
##  1st Qu.:0.0000   1st Qu.: 54889   1st Qu.:1127   1st Qu.:0.5185
##  Median :1.0000   Median : 69402   Median :1192   Median :0.6948
##  Mean   :0.5546   Mean   : 73005   Mean   :1218   Mean   :0.6379
##  3rd Qu.:1.0000   3rd Qu.: 87753   3rd Qu.:1289   3rd Qu.:0.8104
##  Max.   :1.0000   Max.   :142287   Max.   :1566   Max.   :0.9999
##     UG_ENRL       AVG_FAC_SAL       RET_RATE        AVG_GRANT
##  Min.   :  699   Min.   : 4945   Min.   :0.5404   Min.   :0.0892
##  1st Qu.: 4657   1st Qu.: 8249   1st Qu.:0.7572   1st Qu.:0.2009
##  Median : 8903   Median : 9586   Median :0.8237   Median :0.2896
##  Mean   :12645   Mean   :10160   Mean   :0.8251   Mean   :0.2974
##  3rd Qu.:18923   3rd Qu.:11684   3rd Qu.:0.8999   3rd Qu.:0.3706
##  Max.   :58392   Max.   :20484   Max.   :0.9919   Max.   :0.6914
##    MD_EARN_10       MN_EARN_10       UNEMP_RATE     GRAD_DEBT_MDN
##  Min.   : 30000   Min.   : 33700   Min.   :2.190   Min.   : 8700
##  1st Qu.: 42700   1st Qu.: 48300   1st Qu.:2.980   1st Qu.:19791
##  Median : 47700   Median : 54400   Median :3.260   Median :22250
##  Mean   : 50959   Mean   : 59385   Mean   :3.382   Mean   :21927
##  3rd Qu.: 57000   3rd Qu.: 65400   3rd Qu.:3.640   3rd Qu.:25000
##  Max.   :104700   Max.   :153600   Max.   :6.600   Max.   :30500
##   GRAD_RATE_6      GRAD_RATE_4
##  Min.   :0.2135   Min.   :0.0000
##  1st Qu.:0.5208   1st Qu.:0.2802
##  Median :0.6448   Median :0.4541
##  Mean   :0.6538   Mean   :0.4687
```

```
##   3rd Qu.:0.7982   3rd Qu.:0.6346
##   Max.   :0.9771   Max.   :0.9148
```

```r
# Histograms of all the variables
par(mfrow= c(2,4))
for (i in c(5:18)) {
  hist(college.data[,i],
       main=paste0("Histogram of ",colnames(college.data[i])),
       xlab=colnames(college.data[i]))
  lines(college.data[,i], col="red")
}
```




```r
dev.off() # reset par function
```

```
## null device
##           1
```

```r
# Scatter Plot Matrices
library(car)
for (i in list(5:9, 10:14, 15:18)) {
  spm(college.data[,i])
}
# Co-variance and correlation matrix:
cov.m <- cov(college.data[, 5:ncol(college.data)])
cor.m <- cor(college.data[,5:ncol(college.data)])
round(cov.m, 3) # co-variance matrix
```

```
##                 CONTROL       FAM_INC     SAT_AVG  ADM_RATE       UG_ENRL
## CONTROL           0.248     -4211.343     -14.240     0.032      3153.596
## FAM_INC       -4211.343 538572561.839 1846931.055 -1618.624 -20620312.999
## SAT_AVG         -14.240   1846931.055   16647.761   -21.794    223398.299
## ADM_RATE          0.032     -1618.624     -21.794     0.053       -87.682
## UG_ENRL        3153.596 -20620312.999  223398.299   -87.682 109705385.907
## AVG_FAC_SAL    -126.737  24276111.677  284512.334  -451.921   6418248.910
## RET_RATE         -0.007      1366.281      10.579    -0.014       300.068
## AVG_GRANT         0.016     -2333.847     -11.491     0.009       -31.177
## MD_EARN_10    -2280.875 169336309.868 1288463.965 -1926.206  -4308430.974
## MN_EARN_10    -3221.429 218162345.629 1918855.834 -2991.349  -6541526.041
## UNEMP_RATE        0.047     -8189.806     -27.883     0.002      -308.653
## GRAD_DEBT_MDN  -285.960   3516792.560 -234099.482   403.156 -11827455.819
## GRAD_RATE_6      -0.024      2782.792      18.810    -0.025       313.899
## GRAD_RATE_4      -0.046      3472.416      20.779    -0.030       -56.481
##                AVG_FAC_SAL RET_RATE AVG_GRANT    MD_EARN_10    MN_EARN_10
## CONTROL           -126.737   -0.007     0.016     -2280.875     -3221.429
## FAM_INC       24276111.677 1366.281 -2333.847 169336309.868 218162345.629
## SAT_AVG         284512.334   10.579   -11.491   1288463.965   1918855.834
## ADM_RATE          -451.921   -0.014     0.009     -1926.206     -2991.349
## UG_ENRL        6418248.910  300.068   -31.177  -4308430.974  -6541526.041
## AVG_FAC_SAL    7664651.931  192.541  -152.601  26638516.435  40985316.544
## RET_RATE           192.541    0.009    -0.008       862.133      1227.584
## AVG_GRANT         -152.601   -0.008     0.015      -884.906     -1280.778
## MD_EARN_10    26638516.435  862.133  -884.906 154325523.715 215282050.386
## MN_EARN_10    40985316.544 1227.584 -1280.778 215282050.386 326943134.037
## UNEMP_RATE        -169.591   -0.018     0.044     -2013.768     -2524.647
## GRAD_DEBT_MDN -5419939.266 -140.402    74.075 -15994171.014 -30302010.198
## GRAD_RATE_6        339.127    0.014    -0.014      1617.808      2297.460
## GRAD_RATE_4        364.265    0.015    -0.017      1790.490      2649.273
##                UNEMP_RATE GRAD_DEBT_MDN GRAD_RATE_6 GRAD_RATE_4
## CONTROL             0.047      -285.960      -0.024      -0.046
## FAM_INC         -8189.806   3516792.560    2782.792    3472.416
## SAT_AVG           -27.883   -234099.482      18.810      20.779
## ADM_RATE            0.002       403.156      -0.025      -0.030
## UG_ENRL          -308.653 -11827455.819     313.899     -56.481
## AVG_FAC_SAL      -169.591  -5419939.266     339.127     364.265
## RET_RATE           -0.018      -140.402       0.014       0.015
## AVG_GRANT           0.044        74.075      -0.014      -0.017
## MD_EARN_10      -2013.768 -15994171.014    1617.808    1790.490
## MN_EARN_10      -2524.647 -30302010.198    2297.460    2649.273
## UNEMP_RATE          0.405      -189.177      -0.041      -0.052
## GRAD_DEBT_MDN    -189.177  14252660.643    -226.474    -225.611
## GRAD_RATE_6        -0.041      -226.474       0.028       0.031
## GRAD_RATE_4        -0.052      -225.611       0.031       0.044
```

```r
round(cor.m, 3) # co-relation matrix
```

```
##                CONTROL FAM_INC SAT_AVG ADM_RATE UG_ENRL AVG_FAC_SAL RET_RATE
## CONTROL          1.000  -0.365  -0.222    0.277   0.605      -0.092   -0.160
## FAM_INC         -0.365   1.000   0.617   -0.302  -0.085       0.378    0.629
## SAT_AVG         -0.222   0.617   1.000   -0.731   0.165       0.796    0.876
## ADM_RATE         0.277  -0.302  -0.731    1.000  -0.036      -0.706   -0.640
## UG_ENRL          0.605  -0.085   0.165   -0.036   1.000       0.221    0.306
```

```
## AVG_FAC_SAL    -0.092   0.378   0.796   -0.706   0.221       1.000    0.743
## RET_RATE       -0.160   0.629   0.876   -0.640   0.306       0.743    1.000
## AVG_GRANT       0.267  -0.831  -0.736    0.337  -0.025      -0.455   -0.663
## MD_EARN_10     -0.369   0.587   0.804   -0.671  -0.033       0.775    0.741
## MN_EARN_10     -0.358   0.520   0.822   -0.716  -0.035       0.819    0.725
## UNEMP_RATE      0.148  -0.554  -0.339    0.014  -0.046      -0.096   -0.301
## GRAD_DEBT_MDN  -0.152   0.040  -0.481    0.462  -0.299      -0.519   -0.397
## GRAD_RATE_6    -0.289   0.721   0.877   -0.657   0.180       0.737    0.931
## GRAD_RATE_4    -0.445   0.716   0.770   -0.620  -0.026       0.629    0.776
##               AVG_GRANT MD_EARN_10 MN_EARN_10 UNEMP_RATE GRAD_DEBT_MDN
## CONTROL           0.267     -0.369     -0.358      0.148        -0.152
## FAM_INC          -0.831      0.587      0.520     -0.554         0.040
## SAT_AVG          -0.736      0.804      0.822     -0.339        -0.481
## ADM_RATE          0.337     -0.671     -0.716      0.014         0.462
## UG_ENRL          -0.025     -0.033     -0.035     -0.046        -0.299
## AVG_FAC_SAL      -0.455      0.775      0.819     -0.096        -0.519
## RET_RATE         -0.663      0.741      0.725     -0.301        -0.397
## AVG_GRANT         1.000     -0.588     -0.585      0.570         0.162
## MD_EARN_10       -0.588      1.000      0.958     -0.255        -0.341
## MN_EARN_10       -0.585      0.958      1.000     -0.219        -0.444
## UNEMP_RATE        0.570     -0.255     -0.219      1.000        -0.079
## GRAD_DEBT_MDN     0.162     -0.341     -0.444     -0.079         1.000
## GRAD_RATE_6      -0.716      0.783      0.764     -0.391        -0.361
## GRAD_RATE_4      -0.681      0.689      0.701     -0.394        -0.286
##               GRAD_RATE_6 GRAD_RATE_4
## CONTROL            -0.289      -0.445
## FAM_INC             0.721       0.716
## SAT_AVG             0.877       0.770
## ADM_RATE           -0.657      -0.620
## UG_ENRL             0.180      -0.026
## AVG_FAC_SAL         0.737       0.629
## RET_RATE            0.931       0.776
## AVG_GRANT          -0.716      -0.681
## MD_EARN_10          0.783       0.689
## MN_EARN_10          0.764       0.701
## UNEMP_RATE         -0.391      -0.394
## GRAD_DEBT_MDN      -0.361      -0.286
## GRAD_RATE_6         1.000       0.896
## GRAD_RATE_4         0.896       1.000
```

Applying Multivariate Analysis Methods:

1. Principal Component Analysis (PCA)

```r
# PCA on correlation matrix
college.pc <- prcomp(college.data[,6:ncol(college.data)],
                     center = TRUE, scale. = TRUE)

# eigenvalues of each principal component
college.pc$sdev ^ 2
```

```
##  [1] 7.67607883 1.86677708 1.17980245 0.56744864 0.45444578 0.39295582
```

```
##  [7] 0.26284585 0.18720821 0.16247641 0.12426940 0.06610943 0.03592464
## [13] 0.02365747
```

```r
# compute proportion of total variance explained by each component
## Compute cumulative proportion of total variance
## explained by the components
summary(college.pc)
```

```
## Importance of components:
##                           PC1    PC2     PC3     PC4     PC5     PC6     PC7
## Standard deviation     2.7706 1.3663 1.08619 0.75329 0.67413 0.62686 0.51268
## Proportion of Variance 0.5905 0.1436 0.09075 0.04365 0.03496 0.03023 0.02022
## Cumulative Proportion  0.5905 0.7341 0.82482 0.86847 0.90343 0.93365 0.95387
##                           PC8    PC9    PC10    PC11    PC12    PC13
## Standard deviation     0.4327 0.4031 0.35252 0.25712 0.18954 0.15381
## Proportion of Variance 0.0144 0.0125 0.00956 0.00509 0.00276 0.00182
## Cumulative Proportion  0.9683 0.9808 0.99033 0.99542 0.99818 1.00000
```

```r
# library(factoextra)
# fviz_eig(college.pc)

par(mfrow=c(1,2))
# How many PCs to choose?
screeplot(college.pc, type = "l",ylim = c(0,9), npcs = 13, main = "Screeplot (for D&A)")
abline(h = 1, col="red", lty=5)
legend("topright", legend=c("Eigenvalue = 1"),
       col=c("red"), lty=5, cex=0.6)

# according to screeplot let's just choose 3 PCs
# correlation between the principal components and original variables
corr <- cor(college.data[6:ncol(college.data)], college.pc$x[,1:3])
round(corr, 3)
```

```
##                 PC1    PC2    PC3
## FAM_INC       0.721 -0.585 -0.010
## SAT_AVG       0.947  0.075 -0.046
## ADM_RATE     -0.741 -0.381 -0.250
## UG_ENRL       0.126  0.358 -0.876
## AVG_FAC_SAL   0.829  0.366  0.034
## RET_RATE      0.910  0.059 -0.196
## AVG_GRANT    -0.770  0.452  0.131
## MD_EARN_10    0.883  0.064  0.255
## MN_EARN_10    0.889  0.153  0.267
## UNEMP_RATE   -0.394  0.660  0.331
## GRAD_DEBT_MDN -0.446 -0.640  0.176
## GRAD_RATE_6   0.949 -0.060 -0.094
## GRAD_RATE_4   0.872 -0.168  0.072
```

```r
x <-round(corr, 3)
names(dimnames(x)) <- list("", "Table 1S")
x
```

```
##                Table 1S
##                    PC1    PC2    PC3
##   FAM_INC        0.721 -0.585 -0.010
##   SAT_AVG        0.947  0.075 -0.046
##   ADM_RATE      -0.741 -0.381 -0.250
##   UG_ENRL        0.126  0.358 -0.876
##   AVG_FAC_SAL    0.829  0.366  0.034
##   RET_RATE       0.910  0.059 -0.196
##   AVG_GRANT     -0.770  0.452  0.131
##   MD_EARN_10     0.883  0.064  0.255
##   MN_EARN_10     0.889  0.153  0.267
##   UNEMP_RATE    -0.394  0.660  0.331
##   GRAD_DEBT_MDN -0.446 -0.640  0.176
##   GRAD_RATE_6    0.949 -0.060 -0.094
##   GRAD_RATE_4    0.872 -0.168  0.072
```

```r
# New data-set for Classification and Discrimination
college.data.a <- college.pc$x[,1:3]
colnames(college.data.a) <- c("GRAD_RATE_6", "UNEMP_RATE", "UG_ENRLL")
head(college.data.a)
```

```
##      GRAD_RATE_6  UNEMP_RATE   UG_ENRLL
## [1,]  -3.0862439  2.59130224  1.6665597
## [2,]  -1.4504681  0.27631190 -0.5328531
## [3,]  -1.0617638  0.08558002  0.1011173
## [4,]   0.7477203 -0.44542503 -1.9393123
## [5,]   1.2580840 -0.65131322 -1.6261204
## [6,]   1.3830937 -2.00739174 -0.2682109
```

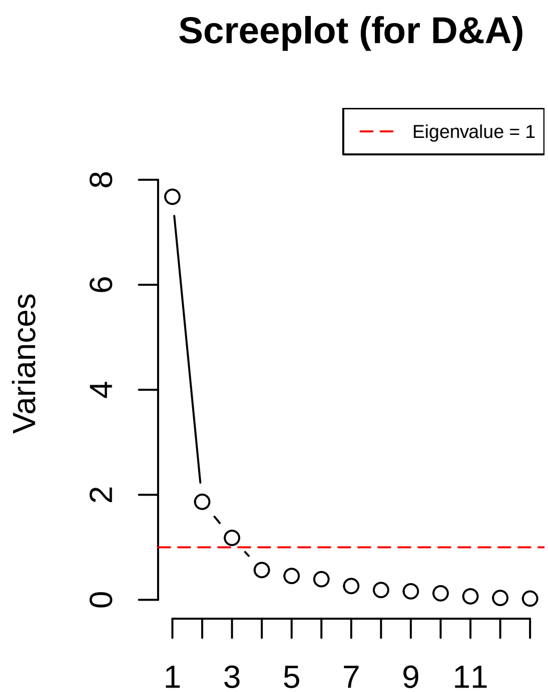
```r
# For Multivariate Regression Analysis
# PCA on correlation matrix
college.pc <- prcomp(college.data[,c(-1:-5, -13, -16, -17)],
                     center = TRUE, scale. = TRUE)

# eigenvalues of each principal component
college.pc$sdev ^ 2
```

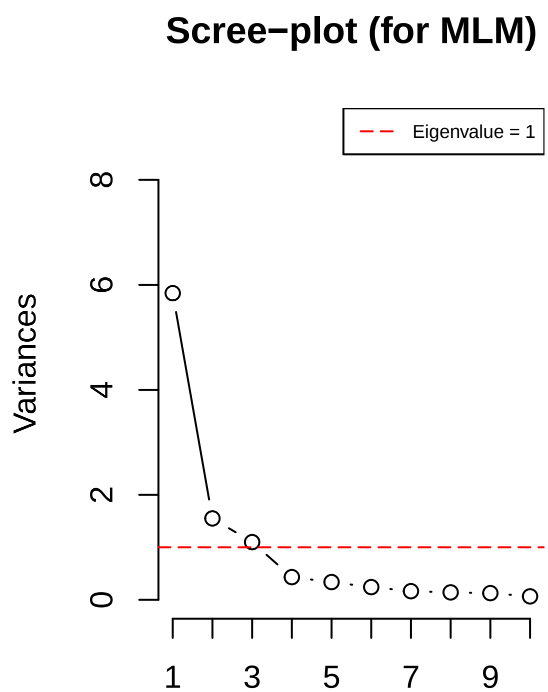
```
##  [1] 5.83980235 1.55042603 1.09880219 0.43207769 0.33805177 0.24205582
##  [7] 0.16425678 0.14159059 0.12699231 0.06594447
```

```r
##compute proportion of total variance explained by each component
## Compute cumulative proportion of total variance
## explained by the components
summary(college.pc)
```

```
## Importance of components:
##                          PC1   PC2    PC3     PC4     PC5     PC6     PC7
## Standard deviation     2.417 1.245 1.0482 0.65733 0.58142 0.49199 0.40529
## Proportion of Variance 0.584 0.155 0.1099 0.04321 0.03381 0.02421 0.01643
## Cumulative Proportion  0.584 0.739 0.8489 0.89211 0.92592 0.95012 0.96655
##                            PC8    PC9    PC10
## Standard deviation     0.37629 0.3564 0.25680
## Proportion of Variance 0.01416 0.0127 0.00659
## Cumulative Proportion  0.98071 0.9934 1.00000
```

```r
# How many PCs to choose?
screeplot(college.pc, type = "l",ylim =c(0,9), npcs = 10, main = "Scree-plot (for MLM)")
abline(h = 1, col="red", lty=5)
legend("topright", legend=c("Eigenvalue = 1"),
       col=c("red"), lty=5, cex=0.6)
```






```r
# according to scree-plot let's just choose 3 PCs
# correlation between the principal components and original variables
corr <- cor(college.data[,c(-1:-5, -13, -16, -17)], college.pc$x[,1:3])
x <-round(corr, 3)
names(dimnames(x)) <- list("", "Table 2")
x
```

```
##              Table 2
##                  PC1    PC2    PC3
##   FAM_INC      0.755 -0.518  0.049
##   SAT_AVG      0.947  0.119 -0.039
##   ADM_RATE    -0.726 -0.473 -0.223
##   UG_ENRL      0.116  0.301 -0.934
##   AVG_FAC_SAL  0.810  0.432 -0.021
##   RET_RATE     0.905  0.124 -0.200
##   AVG_GRANT   -0.804  0.440  0.061
##   MN_EARN_10   0.865  0.215  0.220
##   UNEMP_RATE  -0.434  0.708  0.251
##   GRAD_RATE_4  0.879 -0.098  0.128
```

```r
# new data-set based on PCA
college.data.b <- college.pc$x[, 1:3]
colnames(college.data.b) <- c("SAT_AVG", "UNEMP_RATE", "UG_ENRL")
college.data.c <- cbind(college.data[, c(13,16,17)], data.frame(college.data.b))
# head(college.data.c)
```

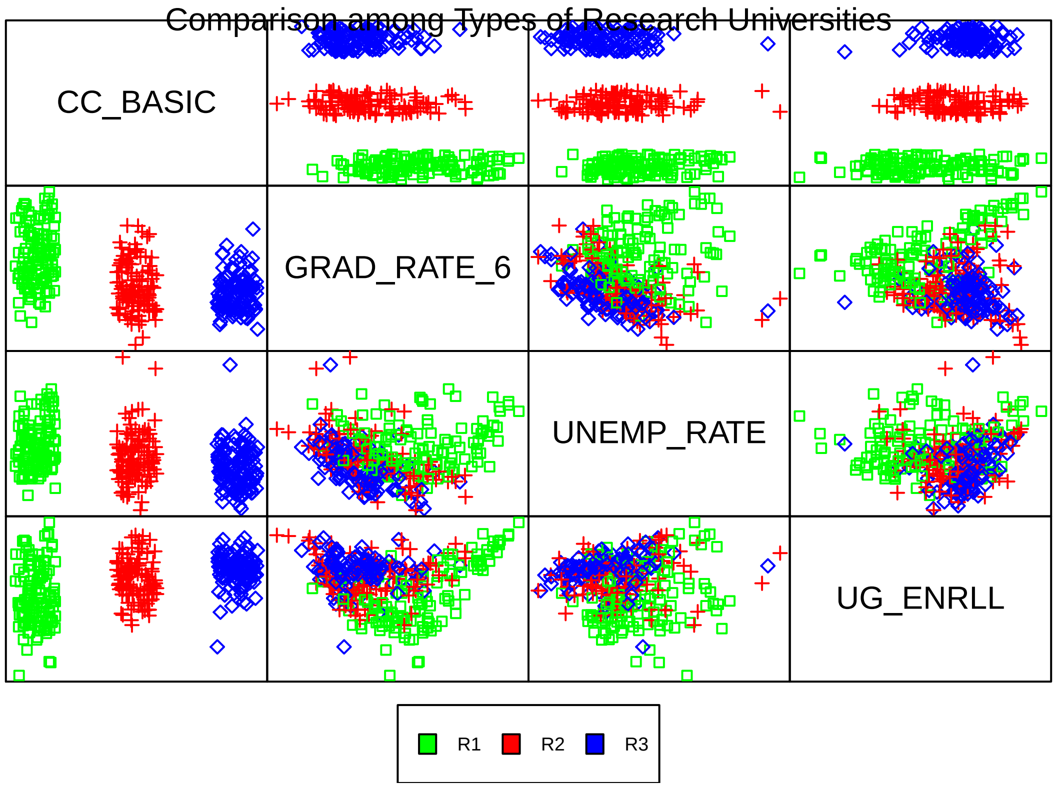
Results:

After running two PCAs, each for the subsequent analysis(Discrimination & Classification, Multivariate Regression), followed by a scree plot on our original data, we found out that the first, second, and third PCs simultaneously explained around 60%, 15%, and 10% of the variation in the data which altogether explained over 82% of the total variability. Only those PCs were chosen which had eigenvalue more than 1. Although the original data had 18 variables and 13 continuous variables, its dimensionality is significantly reduced to 3 variables with keeping maximum variability in the data.

Table 1 and Table 2 from above show the correlation between the original variable and PCs that were respectively used for Discrimination and Classification and Multivariate regression analysis. The new PCs were named accordingly to those original variables with whom they had maximum absolute correlation. For instance, in table 1, PC1 was named GRAD_RATE_6.

2. Discriminant and Classification Analysis

```r
college.data.a1 <- data.frame(college.data[,4], college.data.a)
colnames(college.data.a1)[1] <- colnames(college.data)[4]
# green = R1, red = R2, blue = R3
colors <- c("green", "red" ,"blue")[college.data[ , 4]]
# new data-frame with jittered Class variable
new_college <- cbind(jitter(as.integer(college.data[ , 4])),
                     college.data.a)
shapes = c(0,3,5)
shapes <- shapes[as.numeric(college.data.a1$CC_BASIC)]

colnames(new_college)[1] <- colnames(college.data)[4] # old name to new variable
pairs(new_college, pch = shapes, cex = 1, gap = 0, col = colors,
      xaxt = "n", yaxt = "n")
legend("bottom", fill = c("green", "red", "blue"),
       legend = c(levels(college.data$CC_BASIC)), horiz = T, cex = 0.6,
       xpd = T)
mtext("Comparison among Types of Research Universities", line = 1)
```



Results:

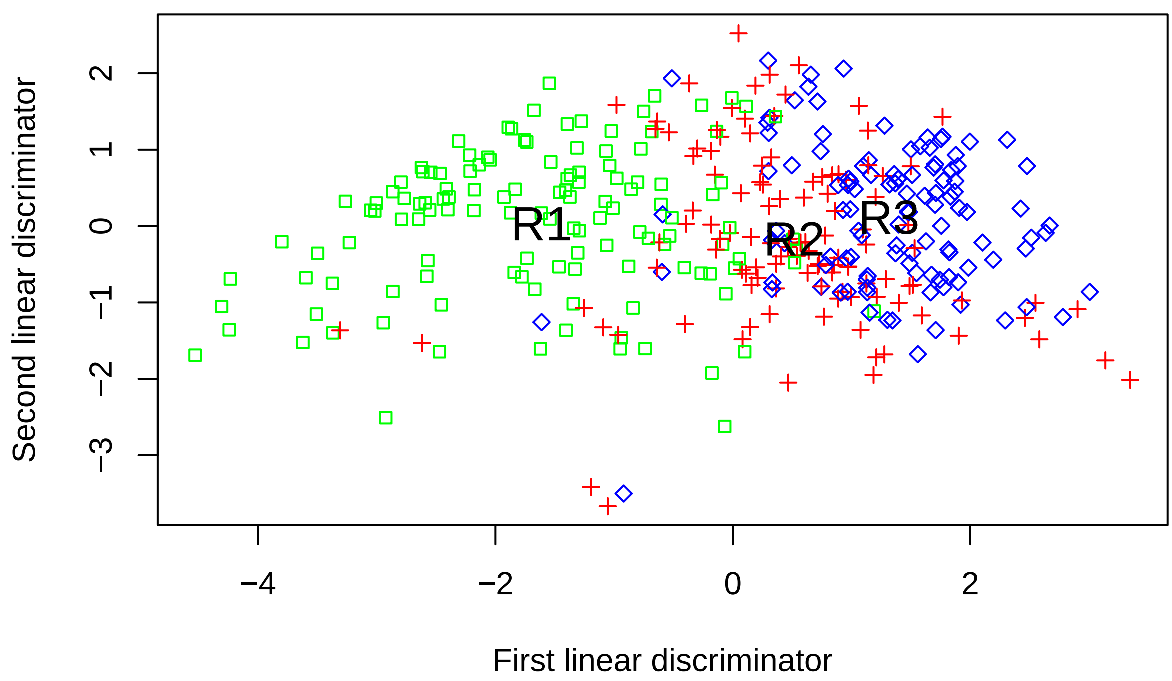
Before performing a classification and discrimination, the scatter plot of the new dataset consisting of all PCs was obtained. Just by eyeballing, it appeared as if the R1 and R3 universities are distinct from one another and R3 universities seemed overlapped with R1 and R3. After running the Linear Discriminant Analysis (LDA), we became more certain that some variables within our data differentiate among the types of universities.

```
library(MASS)
ld <- lda(CC_BASIC ~ ., data = college.data.a1, CV = F)
ld
```

```
## Call:
## lda(CC_BASIC ~ ., data = college.data.a1, CV = F)
##
## Prior probabilities of groups:
##        R1        R2        R3
## 0.3613445 0.3221289 0.3165266
##
## Group means:
##    GRAD_RATE_6  UNEMP_RATE   UG_ENRLL
## R1   2.0240845  0.59252343 -0.5855738
## R2  -0.9003394 -0.03426302  0.1654495
## R3  -1.3944058 -0.64155111  0.5001091
##
## Coefficients of linear discriminants:
##                     LD1        LD2
```

```
## GRAD_RATE_6 -0.4067301  0.2232970
## UNEMP_RATE  -0.5411710 -0.5693897
## UG_ENRLL     0.7994012  0.1292818
##
## Proportion of trace:
##    LD1    LD2
## 0.9916 0.0084
```

```r
loading <- as.matrix(college.data.a1[ , -1]) %*% ld$scaling
plot(loading, col = c("green", "red", "blue")[ college.data.a1[ , 1] ],
  pch = shapes, cex = 0.8,
  xlab = "First linear discriminator",
  ylab = "Second linear discriminator")
for (i in c("R1", "R2", "R3")) { # add class number to each centroid
  centx <- mean(loading[college.data.a1[,1] == i, ] [ , 1] )
  centy <- mean(loading[college.data.a1[,1] == i, ] [ , 2] )
  text(centx, centy, i, cex = 1.5)
}
```




```r
# using linear discrimination analysis
ld1 <- lda(CC_BASIC ~ ., data = college.data.a1, CV = T)
mat <- table(college.data.a1$CC_BASIC, ld1$class)

# Estimated AER using holdout procedure
n <- sum(mat)
```

```
eaer <- (n - sum(diag(mat))) / n
eaer
```

```
## [1] 0.280112
```

Results:

In the above plot, we applied the linear discrimination analysis to the university data and plot the resulting groups in colors and identifying class number. The first and second discriminators are linear combinations of variables that best discriminate between the three research categories of the colleges. This figure illustrates a clear distinction between only the two (R1 and R3) of the three types of universities. And R2 universities seem to have an intersection with both R1 and R3. Based on the confusion matrix in Fig VII, the LDA model was overall successful in classifying the types of universities with around 28% AER.

```
levels(college.data.a1$CC_BASIC) <- c(1:3)
college.data.c1 <- data.frame((college.data.a1$CC_BASIC), college.data.c)
colnames(college.data.c1)[1] <- c("CC_BASIC")
college.mlm.1 <- lm(cbind(GRAD_RATE_6, GRAD_DEBT_MDN, MD_EARN_10) ~ .,
                    data = college.data.c1)
library(car)
Manova(college.mlm.1, type="II",test = c("Wilks"))
```

3. Multivariate Regression Analysis

```
##
## Type II MANOVA Tests: Wilks test statistic
##            Df test stat approx F num Df den Df    Pr(>F)
## CC_BASIC    2   0.96370     2.17      6    698   0.04401 *
## SAT_AVG     1   0.10615   979.64      3    349 < 2.2e-16 ***
## UNEMP_RATE  1   0.72389    44.37      3    349 < 2.2e-16 ***
## UG_ENRL     1   0.85234    20.15      3    349 4.545e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# The following R function (pred.mlm) used for computing Confidence Intervals and Prediction Intervals

pred.mlm <- function(object, newdata, level=0.95,
                     interval = c("confidence", "prediction")){
    form <- as.formula(paste("~",as.character(formula(object))[3]))
    xnew <- model.matrix(form, newdata)
    fit <- predict(object, newdata)
    Y <- model.frame(object)[,1]
    X <- model.matrix(object)
    n <- nrow(Y)
    m <- ncol(Y)
    p <- ncol(X) - 1
    sigmas <- colSums((Y - object$fitted.values)^2) / (n - p - 1)
    fit.var <- diag(xnew %*% tcrossprod(solve(crossprod(X)), xnew))
    if(interval[1]=="prediction") fit.var <- fit.var + 1
    const <- qf(level, df1=m, df2=n-p-m) * m * (n - p - 1) / (n - p - m)
```

```r
    vmat <- (n/(n-p-1)) * outer(fit.var, sigmas)
    lwr <- fit - sqrt(const) * sqrt(vmat)
    upr <- fit + sqrt(const) * sqrt(vmat)
    if(nrow(xnew)==1L){
    ci <- rbind(fit, lwr, upr)
    rownames(ci) <- c("fit", "lwr", "upr")
    } else {
    ci <- array(0, dim=c(nrow(xnew), m, 3))
    dimnames(ci) <- list(1:nrow(xnew), colnames(Y), c("fit", "lwr", "upr") )
    ci[,,1] <- fit
    ci[,,2] <- lwr
    ci[,,3] <- upr
    }
    ci
}
library(usefun) # to print an empty line
for (i in 1:3) {
# For a sample student with following features:
newdata <- data.frame(CC_BASIC = factor(i, levels = c(1,2,3)),
                      SAT_AVG = median(college.data.c1$SAT_AVG),
                      UNEMP_RATE = median(college.data.c1$UNEMP_RATE),
                      UG_ENRL = median(college.data.c1$UG_ENRL))

# 95% Confidence Interval
print(paste0("CONFIDENCE INTERVAL for R", i))
print(pred.mlm(college.mlm.1, newdata))
print_empty_line(html.output = FALSE)
}
```

```
## [1] "CONFIDENCE INTERVAL for R1"
##     GRAD_RATE_6 GRAD_DEBT_MDN MD_EARN_10
## fit   0.6051530      22192.47   48892.03
## lwr   0.5826845      20940.51   46499.97
## upr   0.6276215      23444.44   51284.10
##
## [1] "CONFIDENCE INTERVAL for R2"
##     GRAD_RATE_6 GRAD_DEBT_MDN MD_EARN_10
## fit   0.6163603      22825.39   49603.91
## lwr   0.6020914      22030.31   48084.80
## upr   0.6306292      23620.46   51123.02
##
## [1] "CONFIDENCE INTERVAL for R3"
##     GRAD_RATE_6 GRAD_DEBT_MDN MD_EARN_10
## fit   0.6382129      22531.68   49398.87
## lwr   0.6221380      21635.97   47687.49
## upr   0.6542878      23427.39   51110.26
```

```r
## 95% Prediction Interval
## print(paste0("PREDICTION INTERVAL for R", i))
## print(pred.mlm(college.mlm.1, newdata, interval="prediction"))
## print_empty_line(html.output = FALSE)
## }
```

Results:

Interpreting the confidence interval for the mean graduation rate for a student in our example:

With a confidence coefficient of 0.95, the mean graduation rate for the hypothetical student in an R1 school is somewhere between 61.6% to 63.06%.