DSA 8070 R Session 10: Canonical Correlation Analysis

Whitney

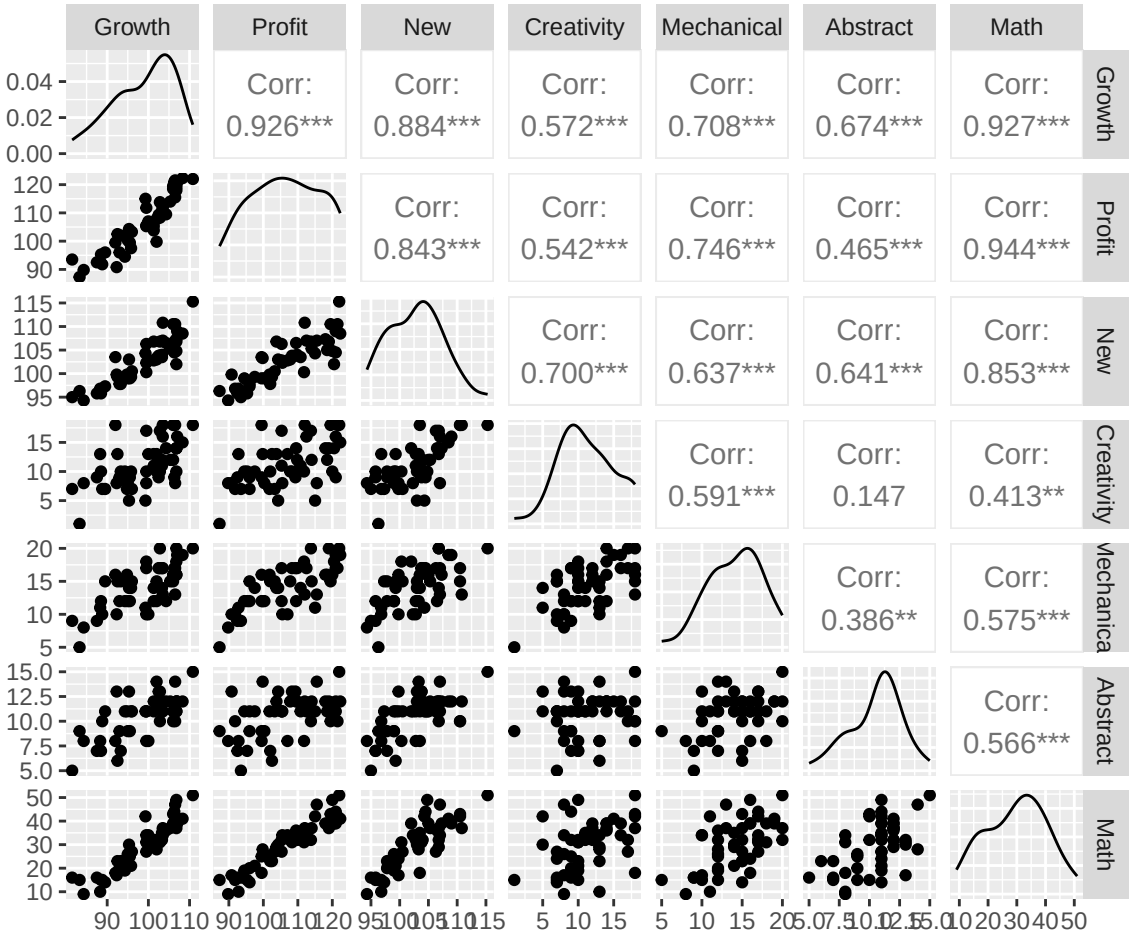
Contents

Load the data and libraries

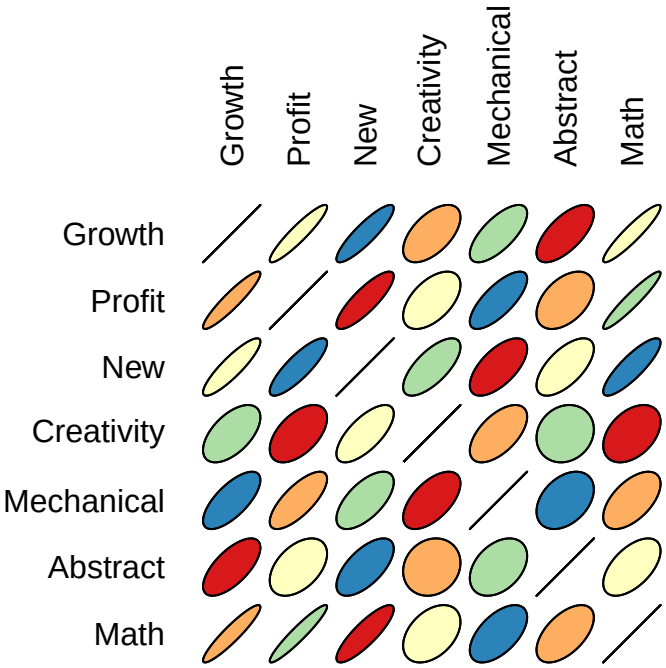
```
library(ggplot2)
library(GGally)
library(ellipse)
library(RColorBrewer)
library(CCA)
library(CCP)
dat1 <- read.table("sales.txt")
colnames(dat1) <- c("Growth", "Profit", "New",
                    "Creativity", "Mechanical", "Abstract", "Math")
summary(dat1)
```

```
##      Growth          Profit           New          Creativity
##  Min.   : 81.50   Min.   : 87.3   Min.   : 94.30   Min.   : 1.00
##  1st Qu.: 93.55   1st Qu.: 99.5   1st Qu.: 99.08   1st Qu.: 8.25
##  Median :100.65   Median :106.2   Median :103.15   Median :10.00
##  Mean   : 98.84   Mean   :106.6   Mean   :102.81   Mean   :11.22
##  3rd Qu.:105.05   3rd Qu.:114.8   3rd Qu.:106.45   3rd Qu.:14.00
##  Max.   :110.80   Max.   :122.3   Max.   :115.30   Max.   :18.00
##    Mechanical       Abstract          Math
##  Min.   : 5.00   Min.   : 5.00   Min.   : 9.00
##  1st Qu.:12.00   1st Qu.: 9.00   1st Qu.:21.50
##  Median :15.00   Median :11.00   Median :31.50
##  Mean   :14.18   Mean   :10.56   Mean   :29.76
##  3rd Qu.:17.00   3rd Qu.:12.00   3rd Qu.:37.00
##  Max.   :20.00   Max.   :15.00   Max.   :51.00
```

```
ggpairs(dat1)
```

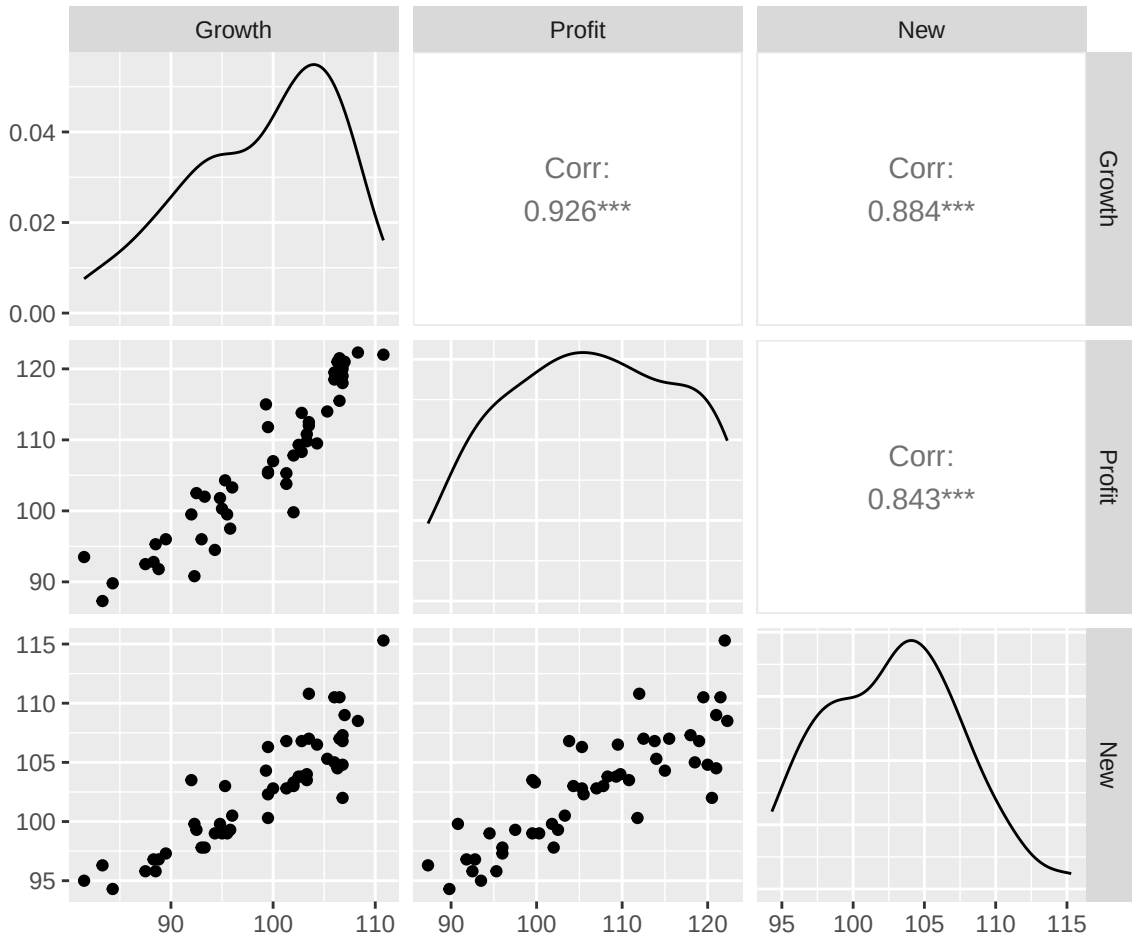



```
my_colors <- brewer.pal(5, "Spectral")
plotcorr(cor(dat1), col = my_colors)
```

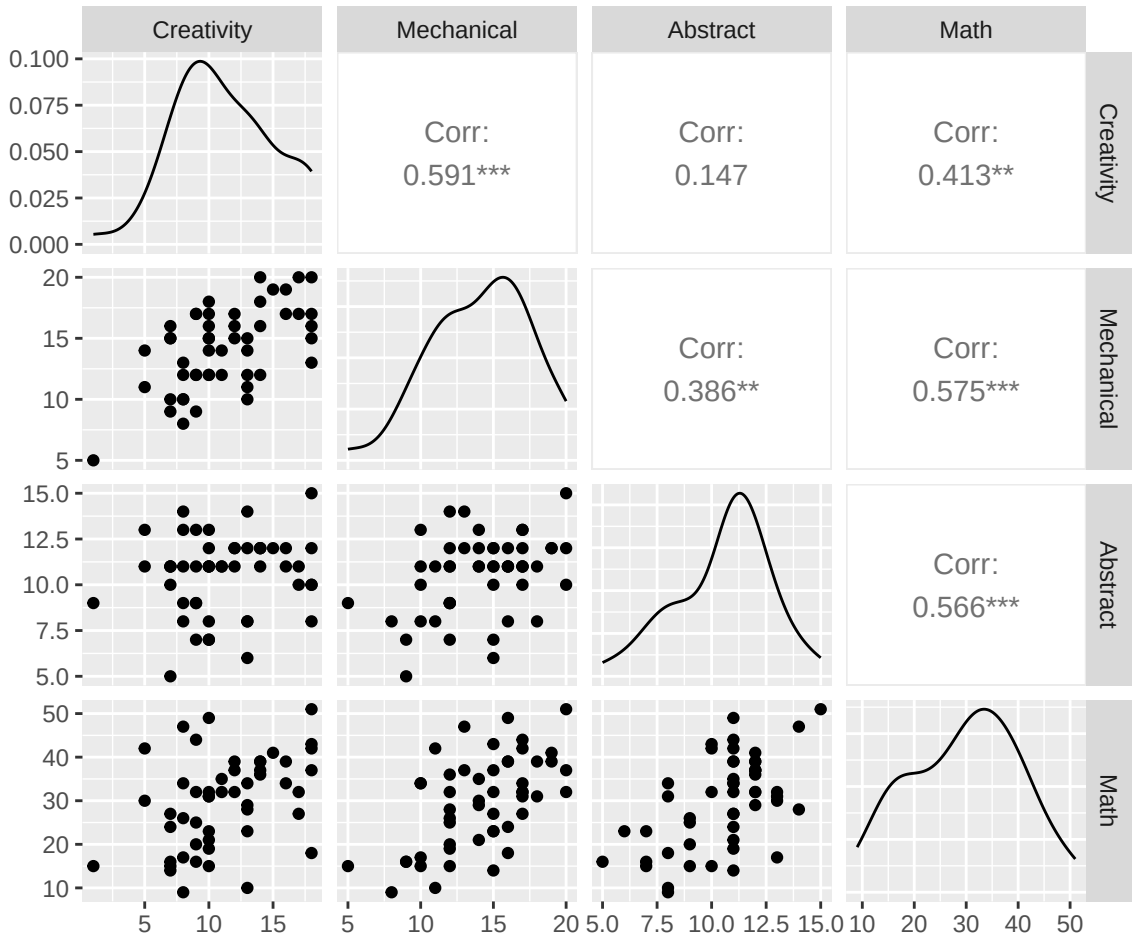


Let's examine *sales* and *intelligence*

```
sales <- dat1[, 1:3]
intelligence <- dat1[, 4:7]
ggpairs(sales)
```




```
ggpairs(intelligence)
```




```
matcor(sales, intelligence)
```

```
## $Xcor
##           Growth    Profit       New
## Growth 1.0000000 0.9260758 0.8840023
## Profit 0.9260758 1.0000000 0.8425232
## New    0.8840023 0.8425232 1.0000000
##
## $Ycor
##            Creativity Mechanical  Abstract      Math
## Creativity  1.0000000  0.5907360 0.1469074 0.4126395
## Mechanical  0.5907360  1.0000000 0.3859502 0.5745533
## Abstract    0.1469074  0.3859502 1.0000000 0.5663721
## Math        0.4126395  0.5745533 0.5663721 1.0000000
##
## $XYcor
##               Growth    Profit       New Creativity Mechanical  Abstract
## Growth     1.0000000 0.9260758 0.8840023  0.5720363  0.7080738 0.6744073
## Profit     0.9260758 1.0000000 0.8425232  0.5415080  0.7459097 0.4653880
## New        0.8840023 0.8425232 1.0000000  0.7003630  0.6374712 0.6410886
## Creativity 0.5720363 0.5415080 0.7003630  1.0000000  0.5907360 0.1469074
## Mechanical 0.7080738 0.7459097 0.6374712  0.5907360  1.0000000 0.3859502
## Abstract   0.6744073 0.4653880 0.6410886  0.1469074  0.3859502 1.0000000
```

```
## Math       0.9273116 0.9442960 0.8525682  0.4126395  0.5745533 0.5663721
##                 Math
## Growth     0.9273116
## Profit     0.9442960
## New        0.8525682
## Creativity 0.4126395
## Mechanical 0.5745533
## Abstract   0.5663721
## Math       1.0000000
```

Test $H_0 : \Sigma_{XY} = 0$

```
rho <- cc(sales, intelligence)$cor
n <- dim(sales)[1]
p <- length(sales); q <- length(intelligence)
## Calculate p-values using the F-approximations
p.asym(rho, n, p, q, tstat = "Wilks")
```

```
## Wilks' Lambda, using F-approximation (Rao's F):
##                 stat    approx df1      df2      p.value
## 1 to 3:  0.002148472 87.391525  12 114.0588 0.000000e+00
## 2 to 3:  0.195241267 18.526265   6  88.0000 8.248957e-14
## 3 to 3:  0.852846693  3.882233   2  45.0000 2.783536e-02
```

Canonical Correlation Analysis using *cc* function from *CCA* package

```
cc1 <- cc(sales, intelligence)
names(cc1)
```

```
## [1] "cor"    "names"  "xcoef"  "ycoef"  "scores"
```

```
cc1$cor
```

```
## [1] 0.9944827 0.8781065 0.3836057
```

```
cc1$xcoef
```

```
##               [,1]       [,2]       [,3]
## Growth -0.06237788 -0.1740703  0.3771529
## Profit -0.02092564  0.2421641 -0.1035150
## New    -0.07825817 -0.2382940 -0.3834151
```

```
cc1$ycoef
```

```
##                   [,1]        [,2]        [,3]
## Creativity -0.06974814 -0.19239132 -0.24655659
## Mechanical -0.03073830  0.20157438  0.14189528
## Abstract   -0.08956418 -0.49576326  0.28022405
## Math       -0.06282997  0.06831607 -0.01133259
```

```
cc1$scores
```

```
## $xscores
##              [,1]         [,2]        [,3]
##  [1,]  0.97838292 -0.362539552  0.81938141
##  [2,]  1.40651588 -0.410239408  0.05351720
##  [3,]  0.66973709  0.044672581  0.66847466
##  [4,] -0.40689705 -2.063089470 -0.30840196
##  [5,] -0.23688307 -0.310765017  0.99852234
##  [6,]  0.65494914 -0.844131320  1.14501451
##  [7,]  0.65528867 -0.236093843  0.93986313
##  [8,] -2.04552806 -1.334870222 -1.86845037
##  [9,] -0.35985473 -0.519574441  0.94175512
## [10,] -0.72379436  2.167475467  1.87763089
## [11,] -0.43808377 -0.291022268  0.89837606
## [12,]  0.04665613  1.736460981  0.67680067
## [13,] -0.74183310 -0.386875380 -0.45592909
## [14,]  0.02197133 -0.265760835  0.56211507
## [15,] -0.07973516 -0.108696870  0.40371149
## [16,]  1.96716552  1.701082028 -2.18552760
## [17,] -0.12525281 -0.746667218  1.06998582
## [18,] -0.41988032  0.070288835  0.98656859
## [19,]  0.25428846  0.007931722 -1.16609979
## [20,] -0.28687624 -1.267369772 -0.95084222
## [21,]  1.43024767  0.727850022 -0.03851613
## [22,] -0.32086304  1.593024005 -1.26353824
## [23,]  1.55121735  0.223860881 -0.12582704
## [24,] -0.75246285  0.068144053  0.71947930
## [25,] -1.29453998  0.585685240 -0.78260154
## [26,]  0.83411571  1.038223887  0.31143725
## [27,] -1.08624858  0.299106887  0.10431852
## [28,] -0.93245443  1.379170139  0.85582619
## [29,]  0.97434443 -1.976531657  0.32682226
## [30,] -0.89871368  1.779857589  0.67875923
## [31,] -1.31816090  0.039068369 -1.57960457
## [32,]  1.41677918 -0.081040167 -0.23857427
## [33,]  0.42719503  0.239653495  0.15995996
## [34,]  0.83477033 -1.238029918  1.00485465
## [35,] -1.39120112  0.436361392 -1.59805812
## [36,] -0.99174366 -0.182594033  0.36498468
## [37,]  0.52144946 -0.699170907 -2.10553997
## [38,] -0.09295538 -2.319565933  1.71161788
## [39,] -1.36370547  0.793354173 -0.23516470
## [40,] -1.06804514  0.660417991  0.19251105
## [41,] -0.36206701 -0.225189261  0.72509423
## [42,]  0.75616793  0.941121130 -0.61716522
## [43,] -0.70972028  0.097445954 -0.77782264
## [44,]  1.88288869 -0.423444109 -1.36329886
## [45,]  0.58821767  0.252097550  0.13103950
## [46,] -1.02875134 -1.413474738 -1.86114887
## [47,]  1.23583458  0.365853533 -0.30894632
## [48,]  1.92471305  0.484483866 -0.47810330
## [49,] -0.68982939 -1.133476875  0.34804581
```

```
## [50,] -0.86681530  1.107521445  0.63269334
##
## $yscores
##              [,1]        [,2]        [,3]
##  [1,]  0.97479103  0.09430244 -0.08851950
##  [2,]  1.40034960 -0.76140727  0.45769014
##  [3,]  0.66755933  0.69659017  0.09004153
##  [4,] -0.19984043 -1.14455925 -0.05227647
##  [5,] -0.20982423 -0.16086269  0.79529079
##  [6,]  0.60160796 -0.61815056  0.49782999
##  [7,]  0.66064116  0.43588278 -0.14518246
##  [8,] -2.38396289 -0.88140585  0.15766738
##  [9,] -0.29803503 -0.32179325  1.37063799
## [10,] -0.93127733  0.64827089 -0.12480193
## [11,] -0.41079711 -0.14249657  0.58596818
## [12,]  0.11904755  2.35859742  0.11141301
## [13,] -0.72588545 -0.27966647 -0.70314741
## [14,]  0.04726778 -0.15155656  0.27603833
## [15,] -0.03278039  0.37377660 -1.79741676
## [16,]  1.81607881  1.58415070 -1.09665800
## [17,] -0.09779329 -0.46221390 -0.15717568
## [18,] -0.34775982  0.14588306  0.09261709
## [19,]  0.20575055 -0.03287585  2.18906768
## [20,] -0.35582377 -0.95027027 -0.87037584
## [21,]  1.39832111  0.55185730 -0.83886122
## [22,] -0.27686589  1.17372033  1.06694262
## [23,]  1.49745418  0.20784154 -1.02932306
## [24,] -0.66347038 -0.20406500  0.24551465
## [25,] -1.19107609 -0.03070063 -0.19579577
## [26,]  0.80346642  1.70310899 -0.50383613
## [27,] -0.95936491 -0.25064113 -0.12836843
## [28,] -1.21910791  1.69784810  0.46430792
## [29,]  0.93624898 -2.30445623  1.02914053
## [30,] -0.86594820  1.75023346  0.90942276
## [31,] -1.27981072  0.04300999 -1.86226853
## [32,]  1.44440068 -1.06423465 -1.38353924
## [33,]  0.40313424  0.57049419  1.31139946
## [34,]  0.78874451 -1.15793146  0.23670462
## [35,] -1.27845734  0.37784268 -1.56714538
## [36,] -1.13042931 -0.14601431  1.39507261
## [37,]  0.43932868 -0.47179081 -1.99750652
## [38,] -0.25466221 -2.60755059  0.23571368
## [39,] -1.24698790  0.29832283  0.02809563
## [40,] -0.95665816  0.41902427  0.46187787
## [41,] -0.29111687 -0.06108586  1.60586199
## [42,]  0.68378617  1.62169828 -1.52372994
## [43,] -0.67262436  0.49179647 -0.78157702
## [44,]  2.06209412 -0.11916799  0.94732921
## [45,]  0.56088586  0.56712037  1.48729252
## [46,] -1.02048263 -1.76156170 -1.51761542
## [47,]  1.21992390 -0.31761469  1.45872318
## [48,]  1.94818626 -0.77531725 -0.56510955
## [49,] -0.64792179 -1.26188686 -0.66195176
## [50,] -0.73030445  0.62990478  0.08452069
```

```
##
## $corr.X.xscores
##              [,1]          [,2]         [,3]
## Growth -0.9798776  0.0006477883  0.199598477
## Profit -0.9464085  0.3228847489 -0.007504408
## New    -0.9518620 -0.1863009724 -0.243414776
##
## $corr.Y.xscores
##                  [,1]       [,2]        [,3]
## Creativity -0.6348095 -0.1894059 -0.24988439
## Mechanical -0.7171837  0.2086069  0.02598458
## Abstract   -0.6436782 -0.4402237  0.22027544
## Math       -0.9388771  0.1734549  0.03614570
##
## $corr.X.yscores
##              [,1]          [,2]         [,3]
## Growth -0.9744713  0.0005688272  0.076567107
## Profit -0.9411869  0.2835272081 -0.002878734
## New    -0.9466102 -0.1635921013 -0.093375287
##
## $corr.Y.yscores
##                  [,1]       [,2]        [,3]
## Creativity -0.6383313 -0.2156981 -0.65140953
## Mechanical -0.7211626  0.2375644  0.06773775
## Abstract   -0.6472493 -0.5013329  0.57422365
## Math       -0.9440859  0.1975329  0.09422619
```

Check

Compute the eigenvalues and eigenvectors of

$$\Sigma_X^{-1/2}\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-1/2}$$

and

$$\Sigma_Y^{-1/2}\Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\Sigma_Y^{-1/2}$$

```
library(expm)
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'Matrix'
```

```
## The following object is masked from 'package:spam':
##
##     det
```

```
##
## Attaching package: 'expm'
```

```
## The following object is masked from 'package:Matrix':
##
##     expm
```

```
a <- solve(sqrtm(var(dat1[, 1:3]))) %*% var(dat1)[1:3, 4:7] %*% solve(var(dat1[, 4:7])) %*% var(dat1)[4
```

```
eigen(a)$values
```

```
## [1] 0.9889958 0.7710711 0.1471533
```

```
cc1$cor^2
```

```
## [1] 0.9889958 0.7710711 0.1471533
```

```
u_vec <- eigen(a)$vectors

u_vec[, 1] %*% solve(sqrtm(var(dat1[, 1:3])))
```

```
##              [,1]        [,2]        [,3]
## [1,] -0.06237788 -0.02092564 -0.07825817
```

```
cc1$xcoef[, 1]
```

```
##      Growth      Profit         New
## -0.06237788 -0.02092564 -0.07825817
```

```
b <- solve(sqrtm(var(dat1[, 4:7]))) %*% var(dat1)[4:7, 1:3] %*% solve(var(dat1[, 1:3])) %*% var(dat1)[1
```

```
eigen(b)$values
```

```
## [1] 9.889958e-01 7.710711e-01 1.471533e-01 7.771561e-16
```

```
cc1$cor^2
```

```
## [1] 0.9889958 0.7710711 0.1471533
```

```
v_vec <- eigen(b)$vectors

v_vec[, 1] %*% solve(sqrtm(var(dat1[, 4:7])))
```

```
##             [,1]      [,2]       [,3]       [,4]
## [1,] 0.06974814 0.0307383 0.08956418 0.06282997
```

```
cc1$ycoef[, 1]
```

```
##  Creativity  Mechanical    Abstract        Math
## -0.06974814 -0.03073830 -0.08956418 -0.06282997
```

Compute the correlations between $\{(U_i, V_i)\}_{i=1}^{3}$ and $\{X_i\}_{i=1}^{3}$ and $\{Y_j\}_{i=1}^{4}$

```r
# compute canonical loadings
cc2 <- comput(sales, intelligence, cc1)
# display canonical loadings
cc2$corr.X.xscores
```

```
##              [,1]          [,2]         [,3]
## Growth -0.9798776  0.0006477883  0.199598477
## Profit -0.9464085  0.3228847489 -0.007504408
## New    -0.9518620 -0.1863009724 -0.243414776
```

```r
cc2$corr.Y.xscores
```

```
##                  [,1]       [,2]        [,3]
## Creativity -0.6348095 -0.1894059 -0.24988439
## Mechanical -0.7171837  0.2086069  0.02598458
## Abstract   -0.6436782 -0.4402237  0.22027544
## Math       -0.9388771  0.1734549  0.03614570
```

```r
cc2$corr.X.yscores
```

```
##              [,1]          [,2]         [,3]
## Growth -0.9744713  0.0005688272  0.076567107
## Profit -0.9411869  0.2835272081 -0.002878734
## New    -0.9466102 -0.1635921013 -0.093375287
```

```r
cc2$corr.Y.yscores
```

```
##                  [,1]       [,2]        [,3]
## Creativity -0.6383313 -0.2156981 -0.65140953
## Mechanical -0.7211626  0.2375644  0.06773775
## Abstract   -0.6472493 -0.5013329  0.57422365
## Math       -0.9440859  0.1975329  0.09422619
```

```r
# check
cc1$xcoef[, 1] %*% var(dat1[, 1:3]) %*% diag(diag(var(dat1[, 1:3])))^(-0.5), 3)
```

```
##            [,1]       [,2]      [,3]
## [1,] -0.9798776 -0.9464085 -0.951862
```

```r
cc2$corr.X.xscores[, 1]
```

```
##     Growth     Profit        New
## -0.9798776 -0.9464085 -0.9518620
```

```r
cc1$ycoef[, 1] %*% var(dat1[, 4:7]) %*% diag(diag(var(dat1[, 4:7])))^(-0.5), 4)
```

```
##            [,1]       [,2]       [,3]       [,4]
## [1,] -0.6383313 -0.7211626 -0.6472493 -0.9440859
```

```
cc2$corr.Y.yscores[, 1]
```

```
## Creativity Mechanical   Abstract       Math
## -0.6383313 -0.7211626 -0.6472493 -0.9440859
```