# DSA 8070 R Session 10: Canonical Correlation Analysis

Whitney

October 24, 2021

## Contents

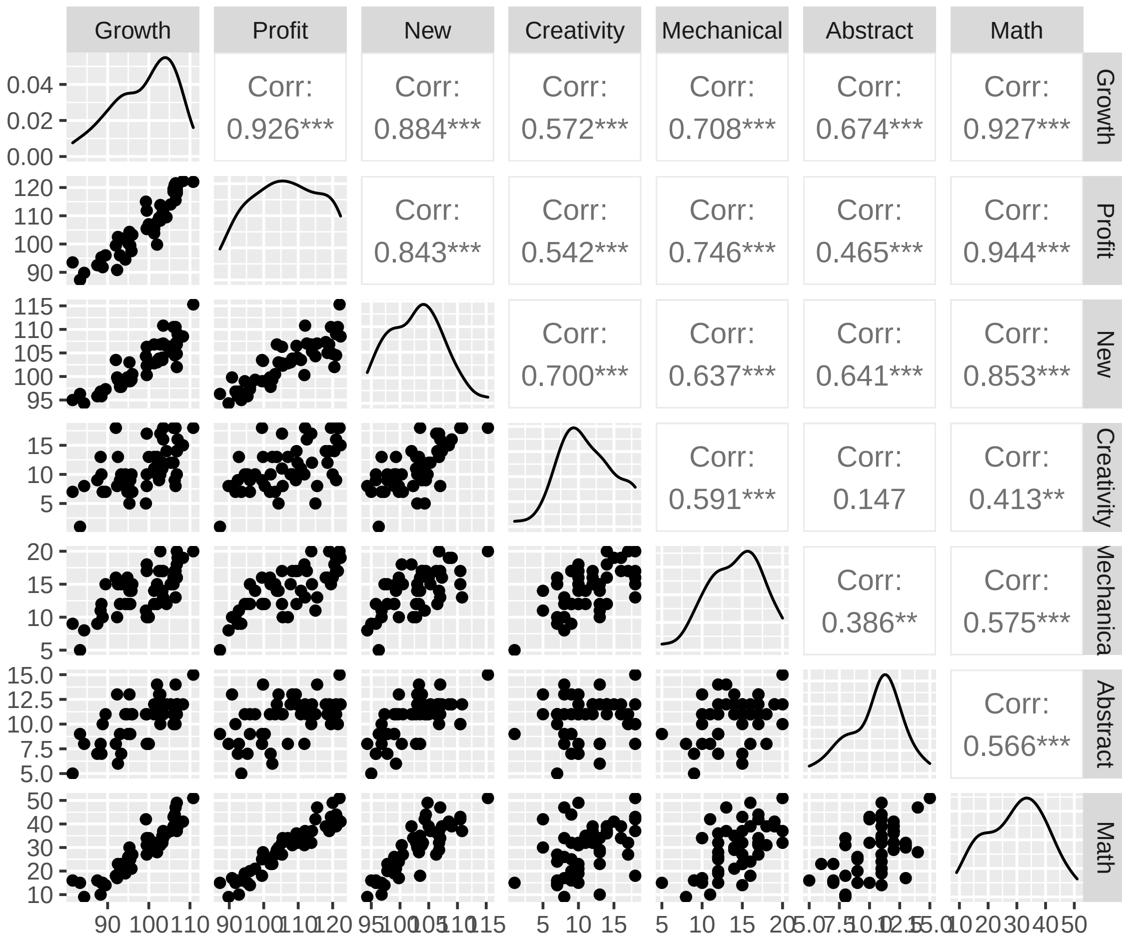


### Load the data and libraries

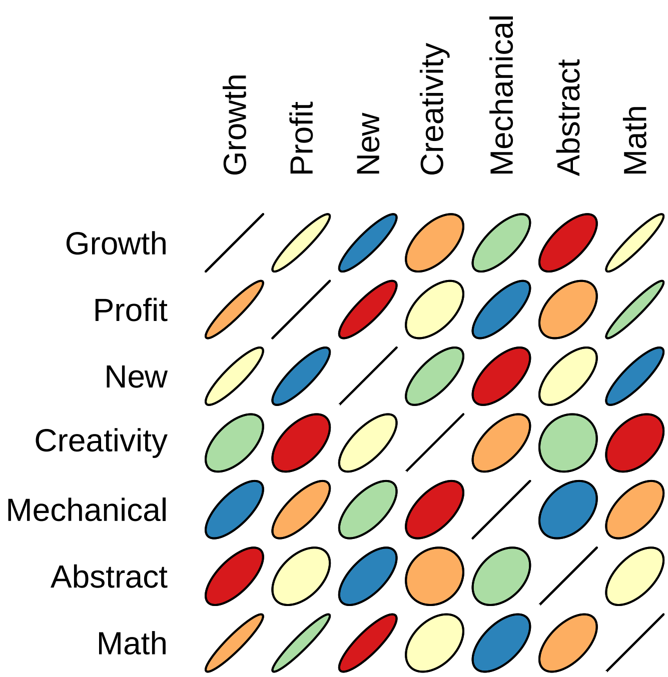
```r
library(ggplot2)
library(GGally)
library(ellipse)
library(RColorBrewer)
library(CCA)
library(CCP)
dat1 <- read.table("sales.txt")
colnames(dat1) <- c("Growth", "Profit", "New",
                    "Creativity", "Mechanical", "Abstract", "Math")
summary(dat1)
```

```
##      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
```

```r
ggpairs(dat1)
```

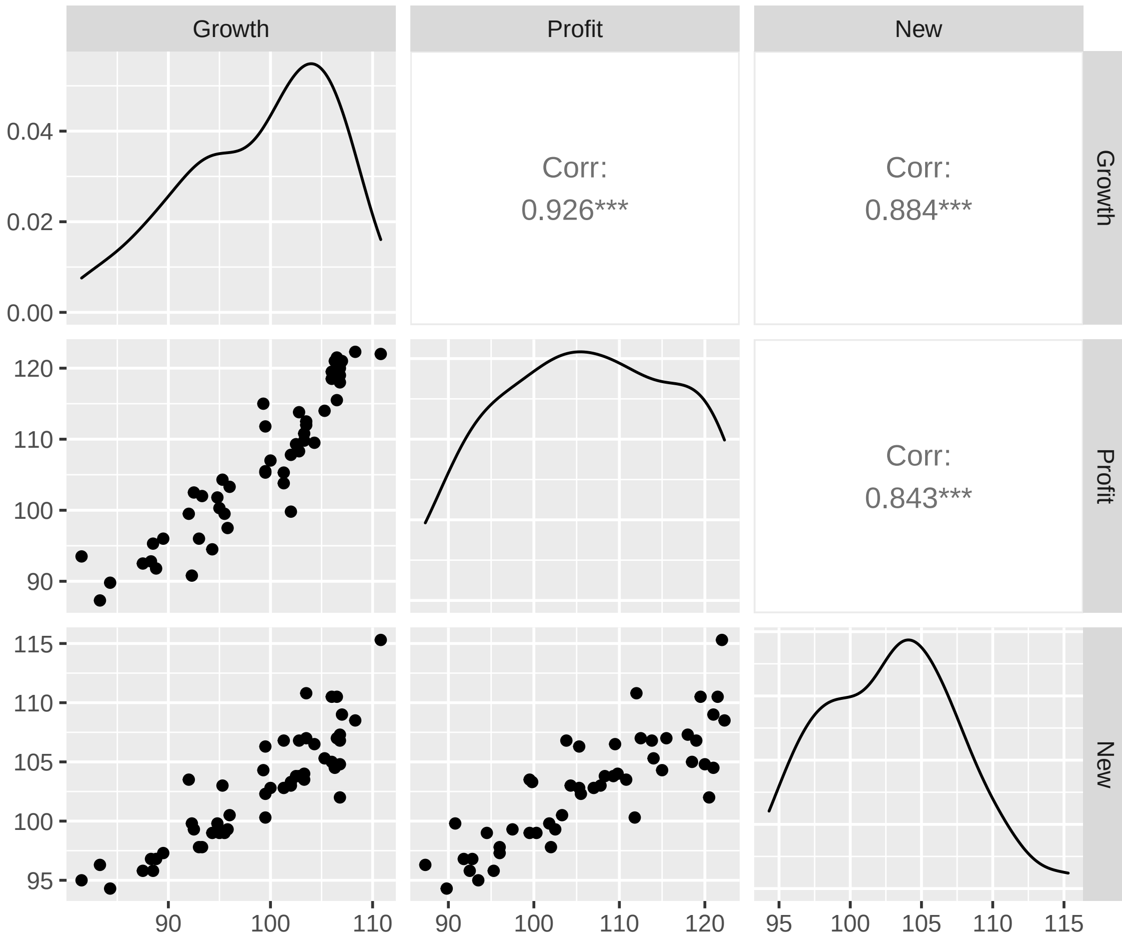



```r
my_colors <- brewer.pal(5, "Spectral")
plotcorr(cor(dat1), col = my_colors)
```

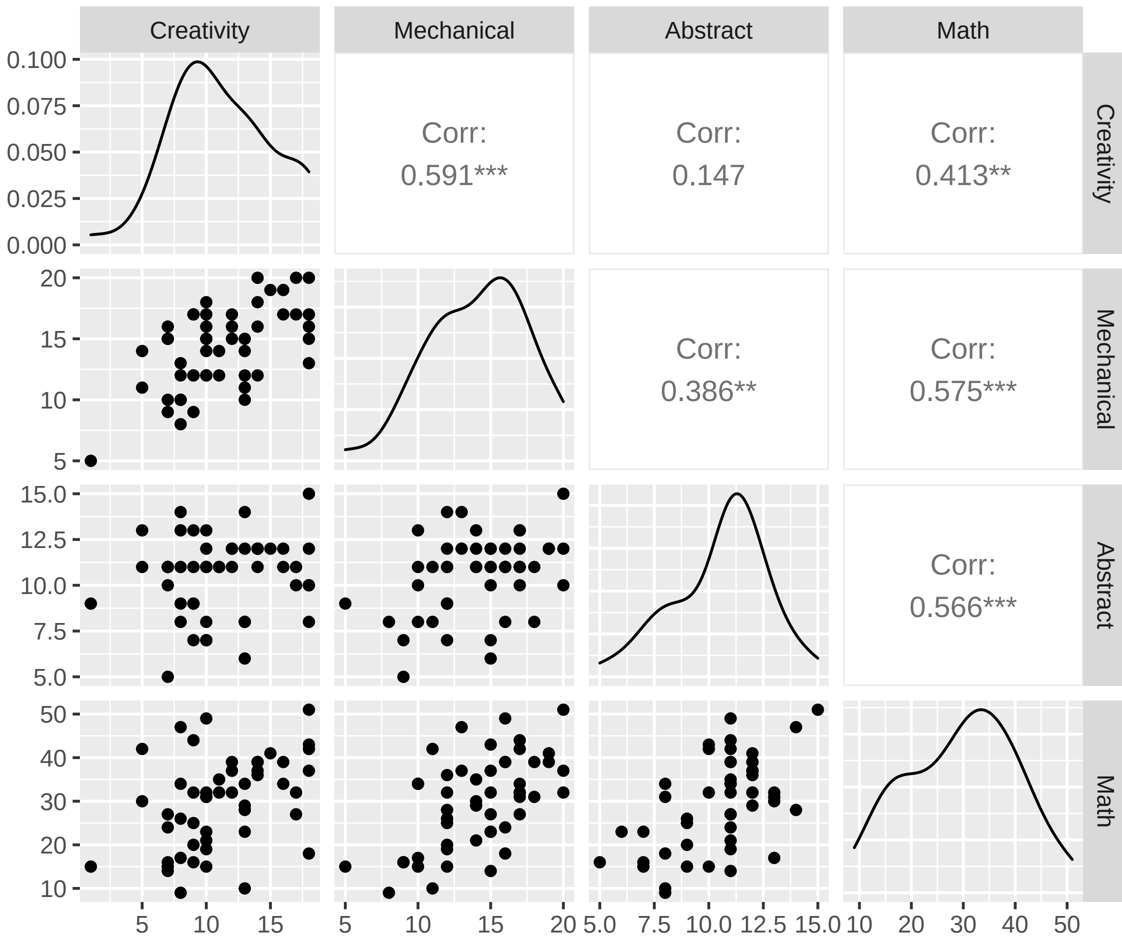


### Let's examine *sales* and *intelligence*

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




```
ggpairs(intelligence)
```




```r
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 $H_0 : \Sigma_{XY} = 0$

```
# tests of canonical dimensions
rho <- cc(sales, intelligence)$cor
## Define number of observations, number of variables in first set, and number of variables in the seco
n <- dim(sales)[1]
p <- length(sales)
q <- length(intelligence)
## Calculate p-values using the F-approximations of different test statistics:
library(CCP)
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
```

## 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)
```

```
##
## 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 $\{Y_j\}_{i=1}^4$

```
# 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
```

```
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
```

```
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
```

```
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
```

```
# 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
```

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

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

```
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
```