Project 5: Canonical Correlation Analysis

DA 410

Marjorie Blanco

```
library(readr)
library(knitr)
library(kableExtra)
library(CCA)
require(GGally)
library(CCA)
library(yacca)
options(scipen=999)
```

Use R to solve Chapter 11 Page 402: #11.9

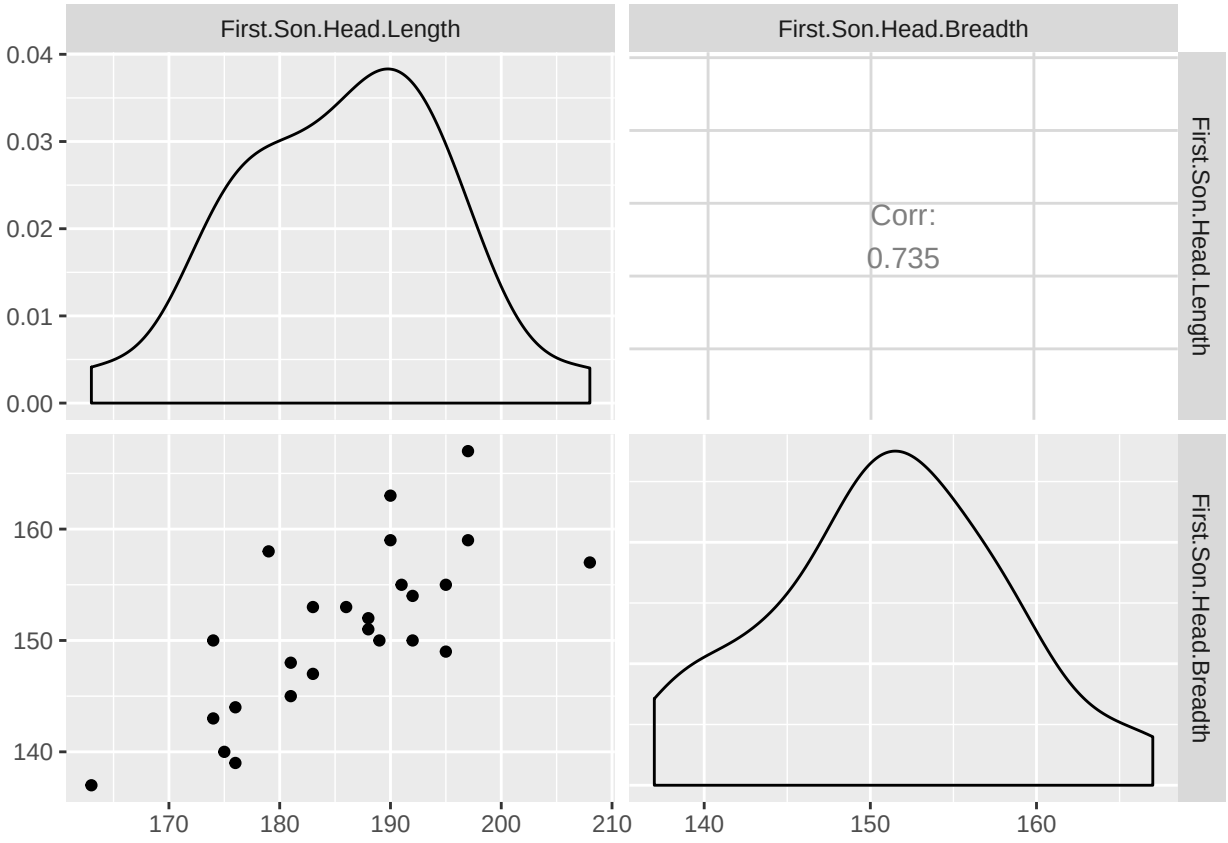
Make sure you include the commands and outputs, as well as the interpretations of the outputs.

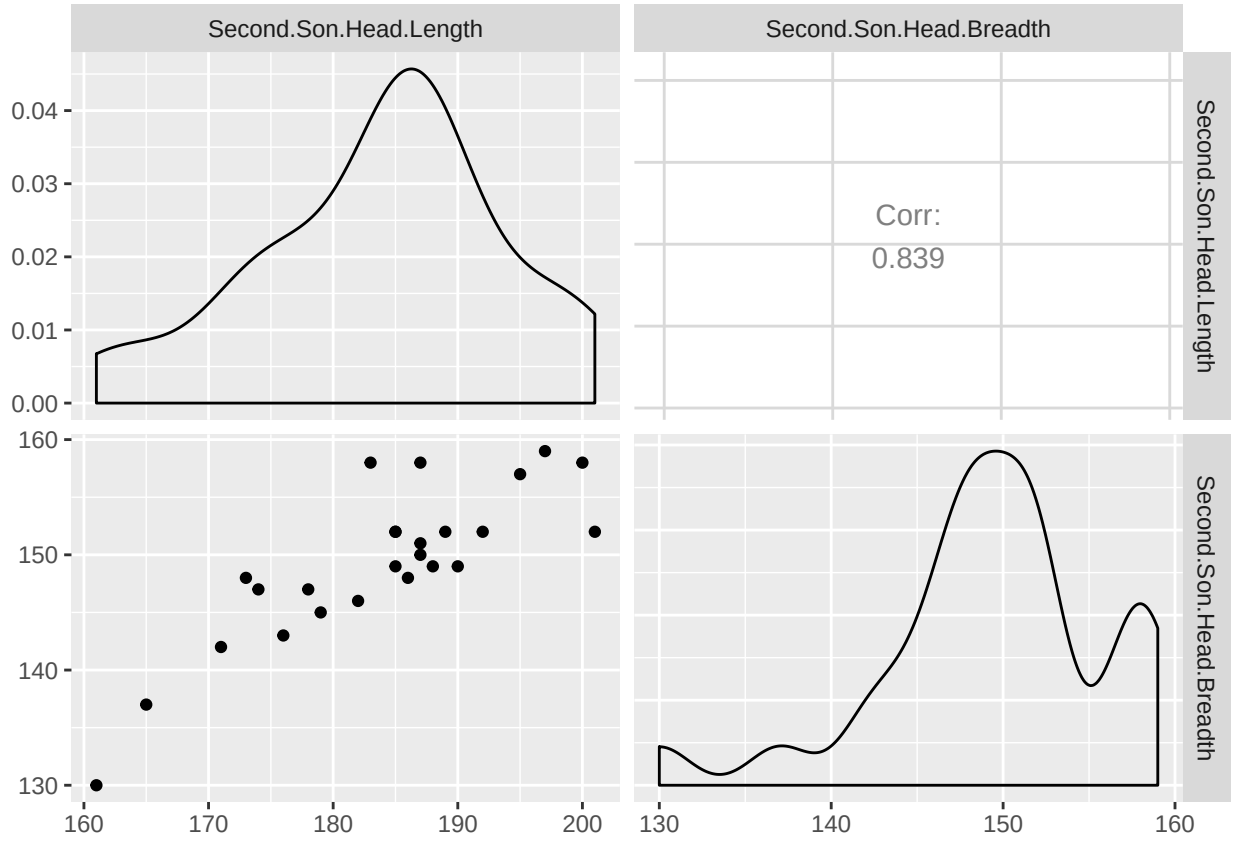
(a) Find the canonical correlations between (y_1, y_2) and (x_1, x_2).

```
sons <- as.data.frame(read_table("Software-Files/T3_8_SONS.DAT",
                                 col_names = c("First.Son.Head.Length",
                                               "First.Son.Head.Breadth",
                                               "Second.Son.Head.Length",
                                               "Second.Son.Head.Breadth"),
                                 cols(
                                   First.Son.Head.Length = col_integer(),
                                   First.Son.Head.Breadth = col_integer(),
                                   Second.Son.Head.Length = col_integer(),
                                   Second.Son.Head.Breadth = col_integer())))
```

For this analysis, we are going to investigate the associations between first son measures and second son measures.

We have a data file, sons.dat, with 25 observations on four variables. The first son variables are head length and head breadth. The second son variables are head length and head breadth.


First.Son.Head.Length
First.Son.Head.Breadth
First.Son.Head.Length
First.Son.Head.Breadth
Corr:
0.735
0.04
0.03
0.02
0.01
0.00
160
150
140
170
180
190
200
210
140
150
160



Correlations

```
matcor(first.son, second.son)
```

```
## $Xcor
##                         First.Son.Head.Length First.Son.Head.Breadth
## First.Son.Head.Length               1.0000000               0.7345555
## First.Son.Head.Breadth              0.7345555               1.0000000
##
## $Ycor
##                          Second.Son.Head.Length Second.Son.Head.Breadth
## Second.Son.Head.Length                1.0000000                0.8392519
## Second.Son.Head.Breadth               0.8392519                1.0000000
##
## $XYcor
##                          First.Son.Head.Length First.Son.Head.Breadth
## First.Son.Head.Length                1.0000000               0.7345555
## First.Son.Head.Breadth               0.7345555               1.0000000
## Second.Son.Head.Length               0.7107518               0.6931573
## Second.Son.Head.Breadth              0.7039807               0.7085504
##                         Second.Son.Head.Length Second.Son.Head.Breadth
## First.Son.Head.Length                0.7107518                0.7039807
## First.Son.Head.Breadth               0.6931573                0.7085504
## Second.Son.Head.Length               1.0000000                0.8392519
## Second.Son.Head.Breadth              0.8392519                1.0000000
```

Display the canonical correlations

```
cc1 <- cc(first.son, second.son)
cc1$cor
```

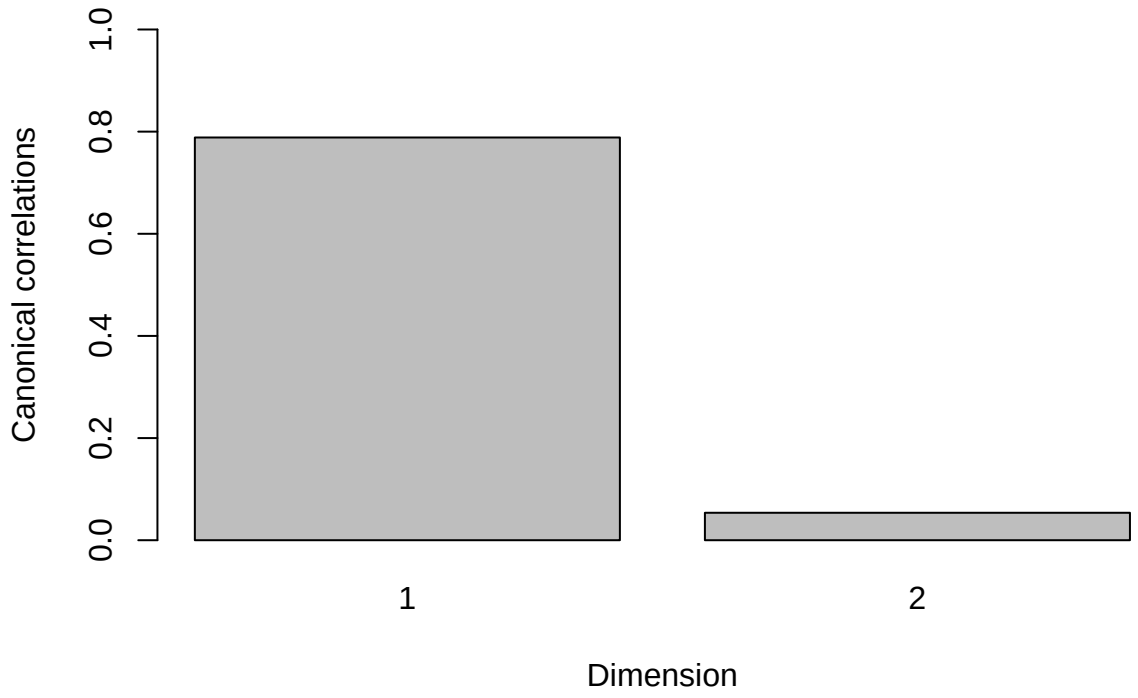
```
## [1] 0.7885079 0.0537397
```

$r_1 = 0.7885079$

$r_2 = 0.0537397$

The first canonical variate captures the most explained variance, canonical r = 0.7885079.

The scree graph of canonical correlations:



(b) Find the standardized coefficients for the canonical variates.

Raw canonical coefficients

```
cc1[3:4]
```

```
## $xcoef
##                               [,1]       [,2]
## First.Son.Head.Length  -0.05656620 -0.1399711
## First.Son.Head.Breadth -0.07073683  0.1869496
##
```

```
## $ycoef
##                                [,1]       [,2]
## Second.Son.Head.Length  -0.0502426 -0.1761479
## Second.Son.Head.Breadth -0.0802224  0.2620836
```

The variable First.Son.Head.Length, a one unit increase leads to a 0.0565662 decrease in the first canonical variate of set 1 when all of the other variables are held constant.

The variable First.Son.Head.Breadth, a one unit increase leads to a 0.0707368 decrease in the first canonical variate of set 1 when all of the other variables are held constant.

The variable First.Son.Head.Length, a one unit increase leads to a 0.1399711 decrease in the second canonical variate of set 1 when all of the other variables are held constant.

The variable First.Son.Head.Breadth, a one unit increase leads to a 0.1869496 increase in the second canonical variate of set 1 when all of the other variables are held constant.

The variable Second.Son.Head.Length, a one unit increase leads to a 0.0502426 decrease in the first canonical variate of set 2 when all of the other variables are held constant.

The variable Second.Son.Head.Breadth, a one unit increase leads to a 0.0802224 decrease in the first canonical variate of set 2 when all of the other variables are held constant.

The variable Second.Son.Head.Length, a one unit increase leads to a 0.1761479 decrease in the second canonical variate of set 2 when all of the other variables are held constant.

The variable Second.Son.Head.Breadth, a one unit increase leads to a 0.2620836 increase in the second canonical variate of set 2 when all of the other variables are held constant.

Compute canonical loadings

```
cc2 <- comput(first.son, second.son, cc1)
cc2[3:6]
```

```
## $corr.X.xscores
##                               [,1]       [,2]
## First.Son.Head.Length  -0.9352877 -0.3538884
## First.Son.Head.Breadth -0.9271512  0.3746875
##
## $corr.Y.xscores
##                                [,1]        [,2]
## Second.Son.Head.Length  -0.7539771 -0.01572908
## Second.Son.Head.Breadth -0.7582663  0.01474027
##
## $corr.X.yscores
##                               [,1]        [,2]
## First.Son.Head.Length  -0.7374817 -0.01901786
## First.Son.Head.Breadth -0.7310660  0.02013559
##
## $corr.Y.yscores
##                                [,1]       [,2]
## Second.Son.Head.Length  -0.9562074 -0.2926900
## Second.Son.Head.Breadth -0.9616470  0.2742901
```

Standardized first.son canonical coefficients diagonal matrix of first.son sd's

```
s1 <- diag(sqrt(diag(cov(first.son))))
s1 %*% cc1$xcoef
```

```
##            [,1]      [,2]
## [1,] -0.5521896 -1.366374
## [2,] -0.5215372  1.378365
```

The variable First.Son.Head.Length, a one standard deviation increase in reading leads to a 0.5521896 standard deviation decrease in the score on the first canonical variate for set 1 when the other variables in the model are held constant.

The variable First.Son.Head.Length, a one standard deviation increase in reading leads to a 1.3663741 standard deviation decrease in the score on the second canonical variate for set 1 when the other variables in the model are held constant.

The variable First.Son.Head.Breadth, a one standard deviation increase in reading leads to a 0.5215372 standard deviation decrease in the score on the first canonical variate for set 1 when the other variables in the model are held constant.

The variable First.Son.Head.Breadth, a one standard deviation increase in reading leads to a 1.3783651 standard deviation increase in the score on the second canonical variate for set 1 when the other variables in the model are held constant.

Standardized second.son canonical coefficients diagonal matrix of second.son sd's

```
s2 <- diag(sqrt(diag(cov(second.son))))
s2 %*% cc1$ycoef
```

```
##            [,1]      [,2]
## [1,] -0.5044484 -1.768570
## [2,] -0.5382877  1.758566
```

The variable Second.Son.Head.Length, a one standard deviation increase in reading leads to a 0.5044484 standard deviation decrease in the score on the first canonical variate for set 2 when the other variables in the model are held constant.

The variable Second.Son.Head.Length, a one standard deviation increase in reading leads to a 1.7685698 standard deviation decrease in the score on the second canonical variate for set 2 when the other variables in the model are held constant.

The variable Second.Son.Head.Breadth, a one standard deviation increase in reading leads to a 0.5382877 standard deviation decrease in the score on the first canonical variate for set 2 when the other variables in the model are held constant.

The variable Second.Son.Head.Breadth, a one standard deviation increase in reading leads to a 1.7585657 standard deviation increase in the score on the second canonical variate for set 2 when the other variables in the model are held constant.

(c) Test the significance of each canonical correlation.

Table 1: Tests of Canonical Dimensions

```
##      Dimension Canonical Corr    WilksL          F df1 df2           p
## [1,]         1      0.7885079 0.3771629 6.59719349   4  42 0.0003256458
## [2,]         2      0.0537397 0.9971120 0.06371905   1  22 0.8030550074
```

```
cca.fit = cca(first.son, second.son)
F.test.cca(cca.fit)
```

```
##
##  F Test for Canonical Correlations (Rao's F Approximation)
##
##          Corr        F   Num df Den df    Pr(>F)
## CV 1 0.788508 6.597193 4.000000     42 0.0003256 ***
## CV 2 0.053740 0.063719 1.000000     22 0.8030550
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

H_0: there is no (linear) relationship between the y's and the x's all canonical correlations r_1, r_2 are non-significant.

H_1: there is (linear) relationship between the y's and the x's at least one canonical correlations r_1, r_2 is significant.

We reject the null hypothesis in favor of the alternative. This implies that at least r_1^2 is significantly different from zero.

We conclude that $r_1 = 0.7885079$ is significant since the p-value < 0.05

We conclude that $r_2 = 0.0537397$ is not significant since the p-value > 0.05

Tests of dimensionality for the canonical correlation analysis, as shown in Table 1, indicate that one of the two canonical dimensions are statistically significant at the 0.05 level. Dimension 1 has a canonical correlation of 0.7885079 between the sets of variables, while for dimension 2 the canonical correlation was much lower at 0.0537397

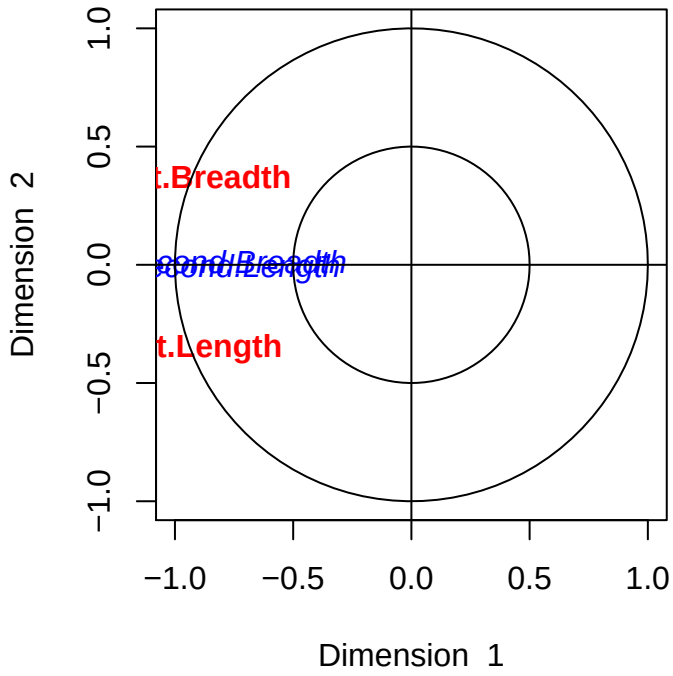
The first test of the canonical dimensions tests whether first dimension is significant (F = 6.5971935, p-value = 0.0003256). First dimension is significant.

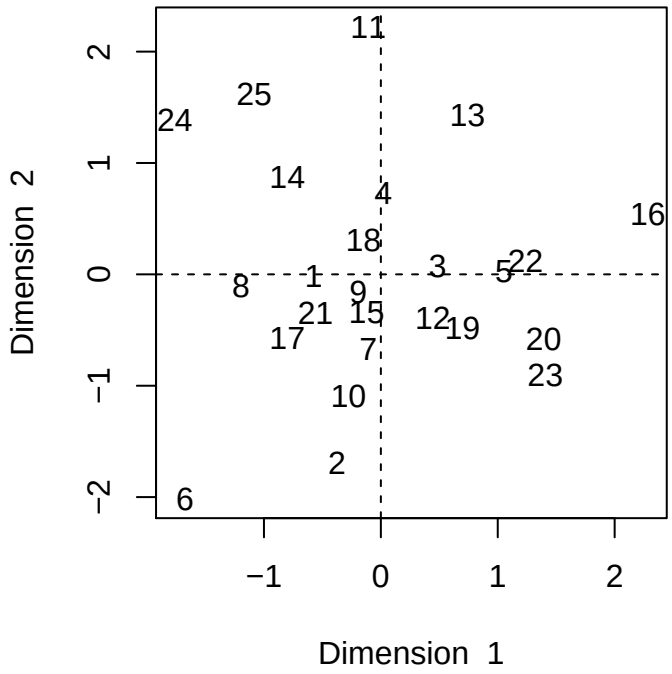
The second test of the canonical dimensions tests whether second dimension is significant (F = 0.063719 p-value = 0.803055). Second dimension is not significant.

```
## $cor
## [1] 0.7885079 0.0537397
##
## $names
## $names$Xnames
## [1] "First.Son.Head.Length"  "First.Son.Head.Breadth"
##
## $names$Ynames
## [1] "Second.Son.Head.Length"  "Second.Son.Head.Breadth"
##
## $names$ind.names
##  [1] "1"  "2"  "3"  "4"  "5"  "6"  "7"  "8"  "9"  "10" "11" "12" "13" "14"
## [15] "15" "16" "17" "18" "19" "20" "21" "22" "23" "24" "25"
##
##
## $xcoef
##                              [,1]      [,2]
## First.Son.Head.Length  -0.5521896 -1.366374
## First.Son.Head.Breadth -0.5215372  1.378365
##
```

```
## $ycoef
##                                [,1]      [,2]
## Second.Son.Head.Length  -0.5044484 -1.768570
## Second.Son.Head.Breadth -0.5382877  1.758566
##
## $scores
## $scores$xscores
##              [,1]        [,2]
##  [1,] -0.57312842 -0.01368291
##  [2,] -0.37497221 -1.69526490
##  [3,]  0.48769136  0.07738080
##  [4,]  0.02087481  0.73218662
##  [5,]  1.05346966  0.02943785
##  [6,] -1.67622740 -2.01929228
##  [7,] -0.10631187 -0.66848874
##  [8,] -1.19547291 -0.10571105
##  [9,] -0.19121934 -0.15461844
## [10,] -0.27601046 -1.08840202
## [11,] -0.10654457  2.22681901
## [12,]  0.44529580 -0.38951099
## [13,]  0.74218106  1.43107765
## [14,] -0.79950955  0.87408659
## [15,] -0.12048251 -0.34156804
## [16,]  2.28398802  0.54041483
## [17,] -0.79939320 -0.57356728
## [18,] -0.14882378  0.31227335
## [19,]  0.69990185 -0.48346801
## [20,]  1.39298318 -0.57838947
## [21,] -0.55895778 -0.34060361
## [22,]  1.23733888  0.12243043
## [23,]  1.40715381 -0.90531017
## [24,] -1.76136757  1.38988577
## [25,] -1.08245687  1.62188500
##
## $scores$yscores
##               [,1]        [,2]
##  [1,]  0.583317147 -0.25867829
##  [2,] -1.083576808 -2.29934799
##  [3,] -0.039028038 -0.26723166
##  [4,] -0.189755833 -0.79567548
##  [5,]  1.225925130  0.36425453
##  [6,] -0.631393423 -0.71401655
##  [7,] -0.290241030 -1.14797136
##  [8,] -0.480665628 -0.18557273
##  [9,] -1.444163206  0.23982870
## [10,] -0.299958033 -0.09536041
## [11,] -0.009048238 -0.70546317
## [12,]  0.674085341  1.14622854
## [13,] -0.279695235  0.51901903
## [14,] -1.183233212  0.06795746
## [15,] -0.861514824  1.73922453
## [16,]  2.691019899 -1.01926884
## [17,] -0.660544431  2.44381628
## [18,]  0.644105541  1.58446004
```

```
## [19,]  0.352366953 -0.52503854
## [20,]  1.928492714  0.11072435
## [21,] -0.279695235  0.51901903
## [22,]  0.473114948  0.44163678
## [23,]  0.894489740 -0.25440160
## [24,] -1.514668603 -0.55069868
## [25,] -0.219735634 -0.35744398
##
## $scores$corr.X.xscores
##                              [,1]       [,2]
## First.Son.Head.Length  -0.9352877 -0.3538884
## First.Son.Head.Breadth -0.9271512  0.3746875
##
## $scores$corr.Y.xscores
##                               [,1]        [,2]
## Second.Son.Head.Length  -0.7539771 -0.01572908
## Second.Son.Head.Breadth -0.7582663  0.01474027
##
## $scores$corr.X.yscores
##                              [,1]        [,2]
## First.Son.Head.Length  -0.7374817 -0.01901786
## First.Son.Head.Breadth -0.7310660  0.02013559
##
## $scores$corr.Y.yscores
##                               [,1]       [,2]
## Second.Son.Head.Length  -0.9562074 -0.2926900
## Second.Son.Head.Breadth -0.9616470  0.2742901
```


Dimension 2
1.0
0.5
0.0
−0.5
−1.0
t.Breadth
t.Length
−1.0
−0.5
0.0
0.5
1.0
Dimension 1


11
25
24
13
14
4
16
18
3
22
5
8
1
9
21
15
12
19
20
17
7
23
10
2
6
Dimension 2
Dimension 1
−2
−1
0
1
2