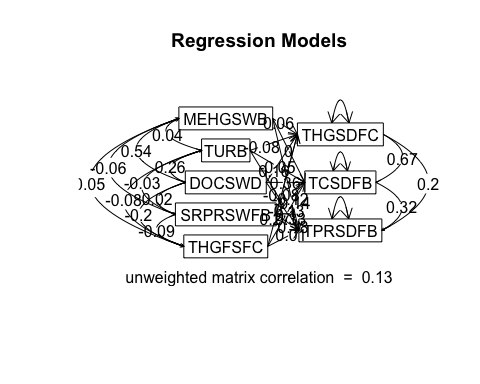
# get the data  
library(readr)  
# removed a row with label units and transformed to csv file format  
marsh <- read\_csv("data\_marsh\_cleaned\_hw2.csv")  
# create two groups for water & soil variables  
water <- marsh[, 2:6]  
soil <- marsh[, 7:9]  
# CCA summary  
library(psych)  
a = set.cor(x=2:6, y=7:9, data = marsh)



print(a) # print produces the t and p-values

## Call: setCor(y = y, x = x, data = data, z = z, n.obs = n.obs, use = use,   
## std = std, square = square, main = main)  
##   
## Multiple Regression from raw data   
##   
## Beta weights   
## THGSDFC TCSDFB TPRSDFB  
## MEHGSWB 0.06 0.00 -0.14  
## TURB -0.08 -0.05 -0.13  
## DOCSWD 0.16 0.36 0.33  
## SRPRSWFB -0.08 -0.12 0.18  
## THGFSFC 0.27 0.10 0.01  
##   
## Multiple R   
## THGSDFC TCSDFB TPRSDFB   
## 0.33 0.36 0.33   
## multiple R2   
## THGSDFC TCSDFB TPRSDFB   
## 0.11 0.13 0.11   
##   
## Unweighted multiple R   
## THGSDFC TCSDFB TPRSDFB   
## 0.25 0.27 0.21   
## Unweighted multiple R2   
## THGSDFC TCSDFB TPRSDFB   
## 0.06 0.07 0.04   
##   
## SE of Beta weights   
## THGSDFC TCSDFB TPRSDFB  
## MEHGSWB 13.01 2.45 37.61  
## TURB 0.47 0.09 1.35  
## DOCSWD 0.72 0.14 2.09  
## SRPRSWFB 730.75 137.52 2112.49  
## THGFSFC 0.08 0.01 0.22  
##   
## t of Beta Weights   
## THGSDFC TCSDFB TPRSDFB  
## MEHGSWB 0.00 0.00 0.00  
## TURB -0.16 -0.52 -0.10  
## DOCSWD 0.23 2.65 0.16  
## SRPRSWFB 0.00 0.00 0.00  
## THGFSFC 3.48 6.65 0.03  
##   
## Probability of t <   
## THGSDFC TCSDFB TPRSDFB  
## MEHGSWB 1.00000 1.0e+00 1.00  
## TURB 0.87000 6.0e-01 0.92  
## DOCSWD 0.82000 8.9e-03 0.87  
## SRPRSWFB 1.00000 1.0e+00 1.00  
## THGFSFC 0.00064 4.4e-10 0.98  
##   
## Shrunken R2   
## THGSDFC TCSDFB TPRSDFB   
## 0.081 0.104 0.084   
##   
## Standard Error of R2   
## THGSDFC TCSDFB TPRSDFB   
## 0.044 0.047 0.044   
##   
## F   
## THGSDFC TCSDFB TPRSDFB   
## 3.88 4.80 4.01   
##   
## Probability of F <   
## THGSDFC TCSDFB TPRSDFB   
## 0.002430 0.000412 0.001880   
##   
## degrees of freedom of regression   
## [1] 5 159  
##   
## Various estimates of between set correlations  
## Squared Canonical Correlations   
## [1] 0.149 0.119 0.072  
## Chisq of canonical correlations   
## [1] 26 20 12  
##   
## Average squared canonical correlation = 0.11  
## Cohen's Set Correlation R2 = 0.3  
## Shrunken Set Correlation R2 = 0.23  
## F and df of Cohen's Set Correlation 3.92 15 420.01  
## Unweighted correlation between the two sets = 0.13

# the test statistic   
(1 - a$cancor2[1]) \* (1 - a$cancor2[2]) \* (1 - a$cancor2[3])

## [1] 0.6963021

# df (n - 1) where n = 165  
a$df[1] + a$df[2]

## [1] 164

# p-value  
ct = corr.test(water, soil)  
ct$p

## THGSDFC TCSDFB TPRSDFB  
## MEHGSWB 0.40453258 0.1320323220 1.0000000  
## TURB 1.00000000 1.0000000000 1.0000000  
## DOCSWD 1.00000000 0.0003389441 0.0558705  
## SRPRSWFB 1.00000000 1.0000000000 0.1372601  
## THGFSFC 0.01453233 1.0000000000 1.0000000

# the test statistic   
(1 - a$cancor2[2]) \* (1 - a$cancor2[3])

## [1] 0.8179043

# df (n - 1) where n = 165  
a$df[1] + a$df[2]

## [1] 164

# p-value  
ct = corr.test(water, soil)  
ct$p

## THGSDFC TCSDFB TPRSDFB  
## MEHGSWB 0.40453258 0.1320323220 1.0000000  
## TURB 1.00000000 1.0000000000 1.0000000  
## DOCSWD 1.00000000 0.0003389441 0.0558705  
## SRPRSWFB 1.00000000 1.0000000000 0.1372601  
## THGFSFC 0.01453233 1.0000000000 1.0000000

# the test statistic   
(1 - a$cancor2[3])

## [1] 0.9284064

# df (n - 1) where n = 165  
a$df[1] + a$df[2]

## [1] 164

# p-value  
ct = corr.test(water, soil)  
ct$p

## THGSDFC TCSDFB TPRSDFB  
## MEHGSWB 0.40453258 0.1320323220 1.0000000  
## TURB 1.00000000 1.0000000000 1.0000000  
## DOCSWD 1.00000000 0.0003389441 0.0558705  
## SRPRSWFB 1.00000000 1.0000000000 0.1372601  
## THGFSFC 0.01453233 1.0000000000 1.0000000

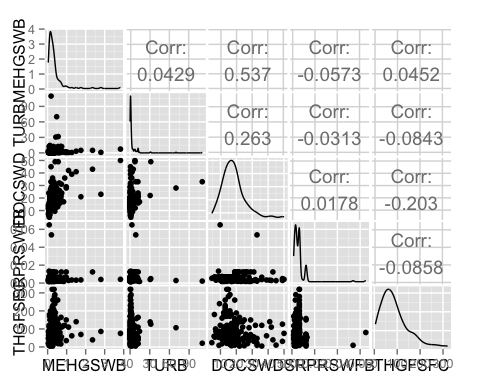
library(CCA)

## Loading required package: fda  
## Loading required package: splines  
## Loading required package: Matrix  
##   
## Attaching package: 'fda'  
##   
## The following object is masked from 'package:graphics':  
##   
## matplot  
##   
## Loading required package: fields  
## Loading required package: spam  
## Loading required package: grid  
## Spam version 1.3-0 (2015-10-24) is loaded.  
## Type 'help( Spam)' or 'demo( spam)' for a short introduction   
## and overview of this package.  
## Help for individual functions is also obtained by adding the  
## suffix '.spam' to the function name, e.g. 'help( chol.spam)'.  
##   
## Attaching package: 'spam'  
##   
## The following objects are masked from 'package:base':  
##   
## backsolve, forwardsolve  
##   
## Loading required package: maps  
##   
## # ATTENTION: maps v3.0 has an updated 'world' map. #  
## # Many country borders and names have changed since 1990. #  
## # Type '?world' or 'news(package="maps")'. See README\_v3. #  
##   
##   
##   
## Attaching package: 'fields'  
##   
## The following object is masked from 'package:psych':  
##   
## describe

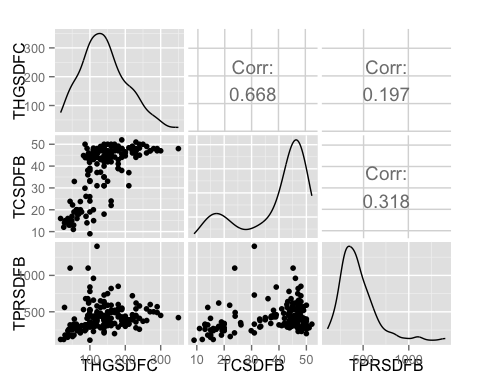
library(ggplot2)

##   
## Attaching package: 'ggplot2'  
##   
## The following object is masked from 'package:psych':  
##   
## %+%

library(GGally)  
ggpairs(water)



ggpairs(soil)



# correlations between the two groups of variables  
matcor(water, soil)

## $Xcor  
## MEHGSWB TURB DOCSWD SRPRSWFB THGFSFC  
## MEHGSWB 1.00000000 0.04286195 0.53653344 -0.05729504 0.04523356  
## TURB 0.04286195 1.00000000 0.26262016 -0.03127880 -0.08426556  
## DOCSWD 0.53653344 0.26262016 1.00000000 0.01784706 -0.20284406  
## SRPRSWFB -0.05729504 -0.03127880 0.01784706 1.00000000 -0.08581679  
## THGFSFC 0.04523356 -0.08426556 -0.20284406 -0.08581679 1.00000000  
##   
## $Ycor  
## THGSDFC TCSDFB TPRSDFB  
## THGSDFC 1.0000000 0.6677804 0.1966074  
## TCSDFB 0.6677804 1.0000000 0.3178176  
## TPRSDFB 0.1966074 0.3178176 1.0000000  
##   
## $XYcor  
## MEHGSWB TURB DOCSWD SRPRSWFB THGFSFC  
## MEHGSWB 1.00000000 0.04286195 0.53653344 -0.05729504 0.04523356  
## TURB 0.04286195 1.00000000 0.26262016 -0.03127880 -0.08426556  
## DOCSWD 0.53653344 0.26262016 1.00000000 0.01784706 -0.20284406  
## SRPRSWFB -0.05729504 -0.03127880 0.01784706 1.00000000 -0.08581679  
## THGFSFC 0.04523356 -0.08426556 -0.20284406 -0.08581679 1.00000000  
## THGSDFC 0.15971021 -0.05151880 0.11909492 -0.09647552 0.25310209  
## TCSDFB 0.19749008 0.04374098 0.32344092 -0.11800127 0.03809560  
## TPRSDFB 0.02092839 -0.05980083 0.22121653 0.19411633 -0.07060351  
## THGSDFC TCSDFB TPRSDFB  
## MEHGSWB 0.15971021 0.19749008 0.02092839  
## TURB -0.05151880 0.04374098 -0.05980083  
## DOCSWD 0.11909492 0.32344092 0.22121653  
## SRPRSWFB -0.09647552 -0.11800127 0.19411633  
## THGFSFC 0.25310209 0.03809560 -0.07060351  
## THGSDFC 1.00000000 0.66778043 0.19660738  
## TCSDFB 0.66778043 1.00000000 0.31781764  
## TPRSDFB 0.19660738 0.31781764 1.00000000

# display the canonical correlations  
cc1 <- cc(water, soil)  
cc1$cor

## [1] 0.3855843 0.3449978 0.2675698

# raw canonical coefficients  
cc1[3:4]

## $xcoef  
## [,1] [,2] [,3]  
## MEHGSWB 0.720571333 -0.613310304 0.442819677  
## TURB 0.014902006 0.003947628 0.046585662  
## DOCSWD -0.122898091 -0.045649299 -0.038307498  
## SRPRSWFB -15.972715690 77.864165952 -98.959103678  
## THGFSFC 0.004124619 -0.009849176 -0.009493841  
##   
## $ycoef  
## [,1] [,2] [,3]  
## THGSDFC 0.011415578 -0.010169482 -0.014106076  
## TCSDFB -0.077556675 -0.037720634 0.072787341  
## TPRSDFB -0.002969355 0.002268621 -0.004222605

# compute canonical loadings  
cc2 <- comput(water, soil, cc1)  
  
# display canonical loadings/latent variables  
cc2[3:6]

## $corr.X.xscores  
## [,1] [,2] [,3]  
## MEHGSWB -0.2138288 -0.54424426 0.05580913  
## TURB -0.1207027 -0.03435814 0.49853147  
## DOCSWD -0.8920181 -0.39006177 0.02464817  
## SRPRSWFB -0.1719363 0.58138401 -0.63983875  
## THGFSFC 0.4914315 -0.62009828 -0.52589688  
##   
## $corr.Y.xscores  
## [,1] [,2] [,3]  
## THGSDFC -0.003665011 -0.30485575 -0.12523874  
## TCSDFB -0.246423901 -0.26504660 0.00980968  
## TPRSDFB -0.275332457 0.05094524 -0.18310544  
##   
## $corr.X.yscores  
## [,1] [,2] [,3]  
## MEHGSWB -0.08244902 -0.18776307 0.014932836  
## TURB -0.04654108 -0.01185348 0.133391950  
## DOCSWD -0.34394820 -0.13457045 0.006595106  
## SRPRSWFB -0.06629592 0.20057620 -0.171201505  
## THGFSFC 0.18948827 -0.21393254 -0.140714106  
##   
## $corr.Y.yscores  
## [,1] [,2] [,3]  
## THGSDFC -0.009505083 -0.8836455 -0.46806012  
## TCSDFB -0.639092107 -0.7682559 0.03666214  
## TPRSDFB -0.714065477 0.1476683 -0.68432782

# tests of canonical dimensions  
ev <- (1 - cc1$cor^2)  
  
n <- dim(water)[1]  
p <- length(water)  
q <- length(soil)  
k <- min(p, q)  
m <- n - 3/2 - (p + q)/2  
  
w <- rev(cumprod(rev(ev)))  
  
# initialize  
d1 <- d2 <- f <- vector("numeric", k)  
  
for (i in 1:k) {  
 s <- sqrt((p^2 \* q^2 - 4)/(p^2 + q^2 - 5))  
 si <- 1/s  
 d1[i] <- p \* q  
 d2[i] <- m \* s - p \* q/2 + 1  
 r <- (1 - w[i]^si)/w[i]^si  
 f[i] <- r \* d2[i]/d1[i]  
 p <- p - 1  
 q <- q - 1  
}  
  
pv <- pf(f, d1, d2, lower.tail = FALSE)  
(dmat <- cbind(WilksL = w, F = f, df1 = d1, df2 = d2, p = pv))

## WilksL F df1 df2 p  
## [1,] 0.6963021 4.051995 15 433.8093 6.185853e-07  
## [2,] 0.8179043 4.176302 8 316.0000 9.094796e-05  
## [3,] 0.9284064 4.087068 3 159.0000 7.921523e-03