DSC324/424
Assignment #3 **(Due Sunday, August 14, 2022 at midnight)**

1) (20 points, for data projects) Choose a technique that we have covered so far in this course, and try applying that technique to your data. You may choose any of

a) Model building and Multiple Regression
b) PCA
c) CFA
d) CCA
e) CA (correspondence analysis)

If you are working as a group, each member of your group should try a different technique, or the same technique with different aspects of the data.

I performed the CCA technique on our group's data (Online New popularity data set). Our data is not really categorical data except for some variables on weekdays. Our dependent variable only had one variable so when creating the categorical variables seen in the code below we only had 1 CV. This limits our data and the information we are going to receive. For reference see Helio plots below the code.

```
library(yacca)
```

```
#Read in Data
```

```
setwd("C:/Users/doret/Documents/DSC 424/Project")


ONP = read.csv("OnlineNewsPopularity.csv", header = TRUE, sep = ",")

head(ONP)


#See the first six lines of the data

head(ONP)


names(ONP)


shares = ONP[, 61]

numbers = ONP[, 3:13]

data = ONP[, 14:19]

keyword = ONP[, 20:28]

selfRef = ONP[, 29:31]

day = ONP[, 32:39]
```

```
LDA = ONP[, 40:44]
global = ONP[, 45:50]
polarity = ONP[, 51:56]
tital = ONP[, 57:60]

#Numbers
# This gives us the cannonical correlates, but no significance tests
# c = cca(shares,numbers)
# summary(c)

#CV1
# helio.plot(c, cv=1, x.name="shares Values",
           y.name="numbers Values")

#Function Names
# ls(c)
# Perform a chi-square test on C
# c
# ls(c)
# c$chisq
# c$df
# summary(c)
# round(pchisq(c$chisq, c$df, lower.tail=F), 3)

#Data
# This gives us the cannonical correlates, but no significance tests
c2 = cca(shares,data)
summary(c2)

#CV1
helio.plot(c2, cv=1, x.name="shares Values",
          y.name="Data Values")

#Function Names
ls(c2)

# Perform a chi-square test on C2
c2
ls(c2)
c2$chisq
c2$df
summary(c2)
round(pchisq(c2$chisq, c2$df, lower.tail=F), 3)

#Keywords
# This gives us the cannonical correlates, but no significance tests
c3 = cca(shares,keyword)
```

```
summary(c3)


#CV1

helio.plot(c3, cv=1, x.name="shares Values",

        y.name="keyword Values")


#Function Names

ls(c3)


# Perform a chi-square test on C2

c3

ls(c3)

c3$chisq

c3$df

summary(c3)

round(pchisq(c3$chisq, c3$df, lower.tail=F), 3)
```

```
#selfRef

# This gives us the cannonical correlates, but no significance tests

c4 = cca(shares,selfRef)

summary(c4)


#CV1

helio.plot(c4, cv=1, x.name="shares Values",

          y.name="selfRef Values")


#Function Names

ls(c4)


# Perform a chi-square test on C2

c4

ls(c4)
```

```
c4$chisq
c4$df
summary(c4)
round(pchisq(c4$chisq, c4$df, lower.tail=F), 3)


# #day
# # This gives us the cannonical correlates, but no significance tests
# c5 = cca(shares,day)
# summary(c5)
#
# #CV1
# helio.plot(c5, cv=1, x.name="shares Values",
#            y.name="day Values")
#
# #Function Names
# ls(c5)
```

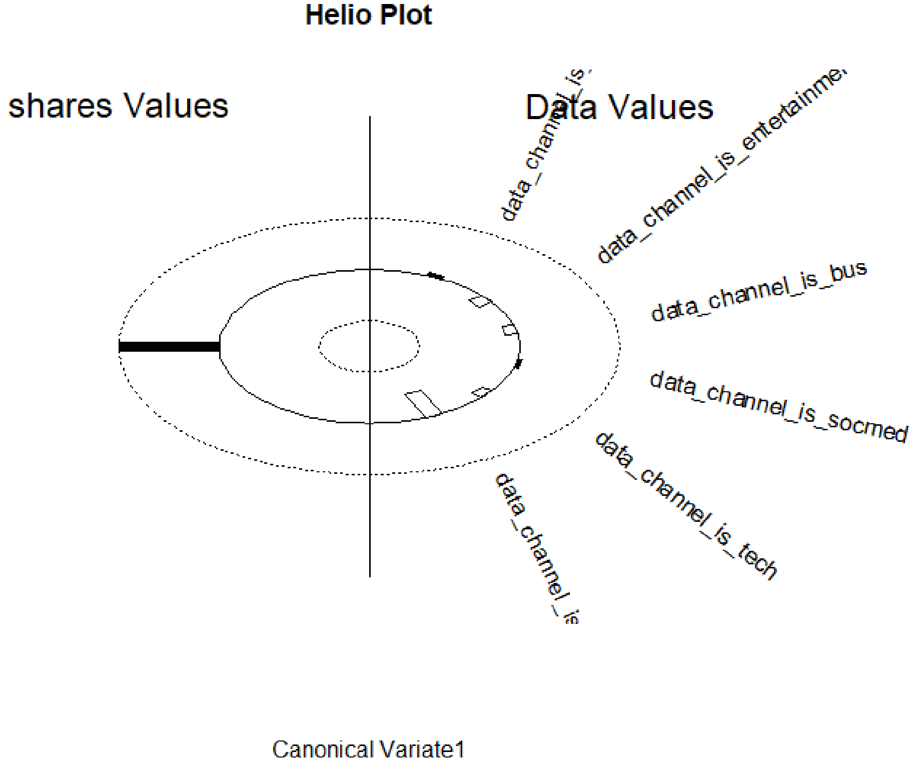
```
#
# # Perform a chi-square test on C2
# c5
# ls(c5)
# c5$chisq
# c5$df
# summary(c5)
# round(pchisq(c5$chisq, c5$df, lower.tail=F), 3)

#LDA
# This gives us the cannonical correlates, but no significance tests
c6 = cca(shares,LDA)
summary(c6)

#CV1
helio.plot(c6, cv=1, x.name="shares Values",
          y.name="LDA Values")

#Function Names
ls(c6)

# Perform a chi-square test on C2
c6
ls(c6)
c6$chisq
c6$df
summary(c6)
round(pchisq(c6$chisq, c6$df, lower.tail=F), 3)

#global
# This gives us the cannonical correlates, but no significance tests
c7 = cca(shares,global)
```

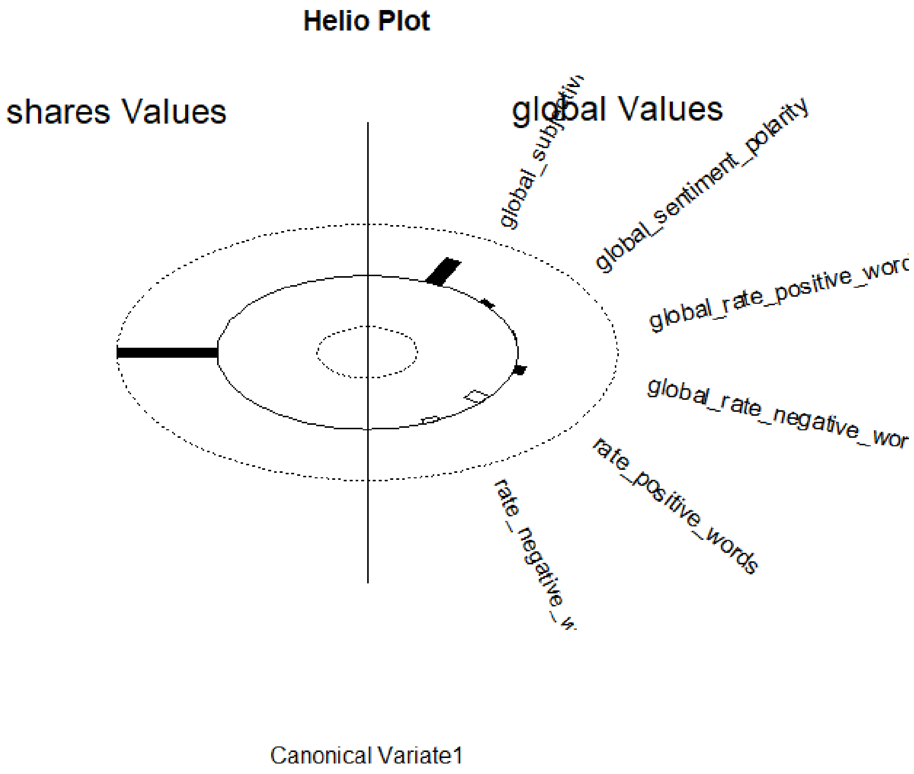
```
summary(c7)

#CV1
helio.plot(c7, cv=1, x.name="shares Values",
           y.name="global Values")

#Function Names
ls(c7)

# Perform a chi-square test on C2
c7
ls(c7)
c7$chisq
c7$df
summary(c7)
round(pchisq(c7$chisq, c7$df, lower.tail=F), 3)
```

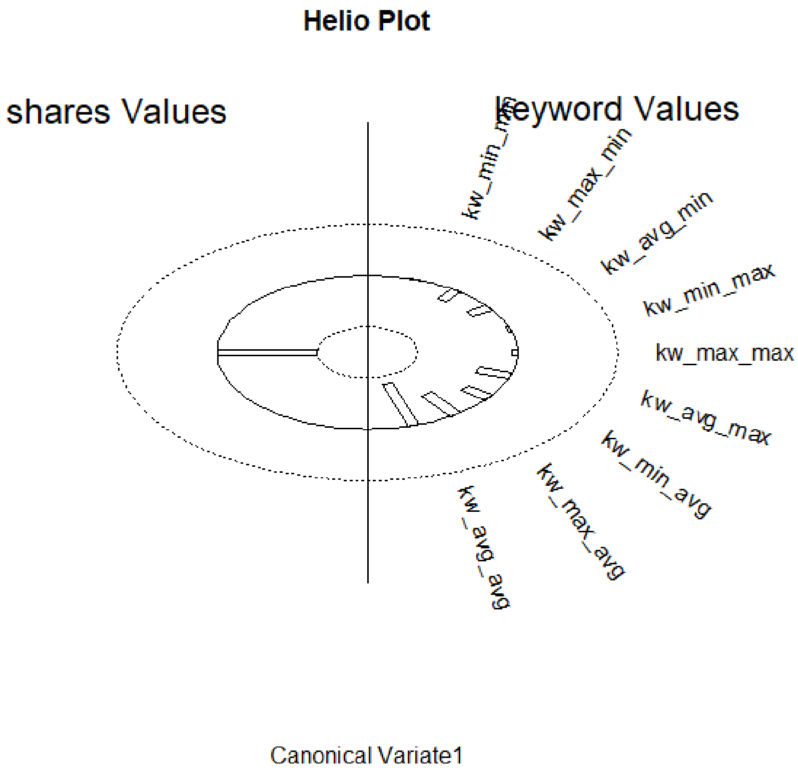
```
#polarity
# This gives us the cannonical correlates, but no significance tests
c8 = cca(shares,polarity)
summary(c8)

#CV1
helio.plot(c8, cv=1, x.name="shares Values",
          y.name="polarity Values")

#Function Names
ls(c8)

# Perform a chi-square test on C2
c8
ls(c8)
```

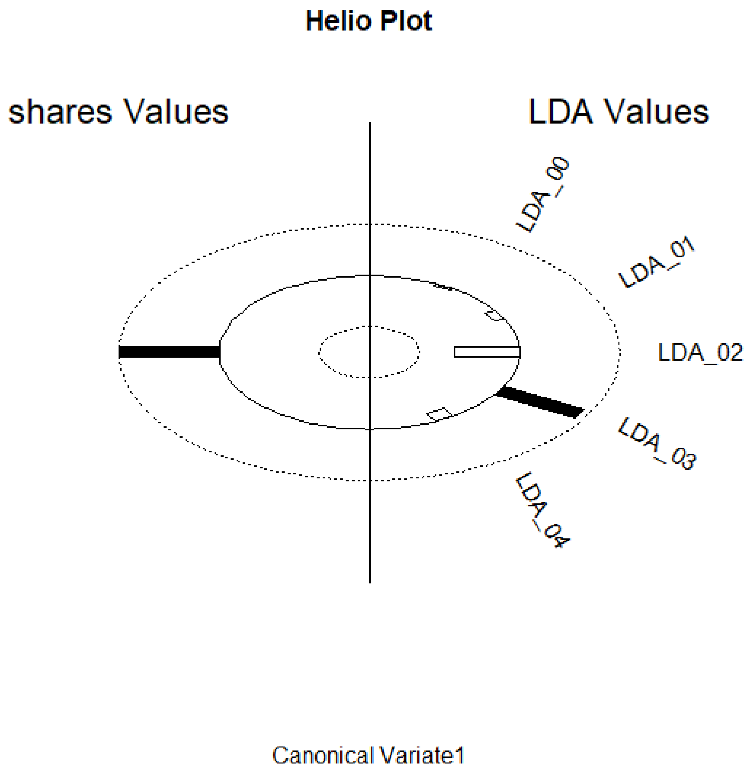
```
c8$chisq
c8$df
summary(c8)
round(pchisq(c8$chisq, c8$df, lower.tail=F), 3)

#tital
# This gives us the cannonical correlates, but no significance tests
c9 = cca(shares, tital)
summary(c9)

#CV1
helio.plot(c9, cv=1, x.name="shares Values",
        y.name="tital Values")

#Function Names
ls(c9)
```

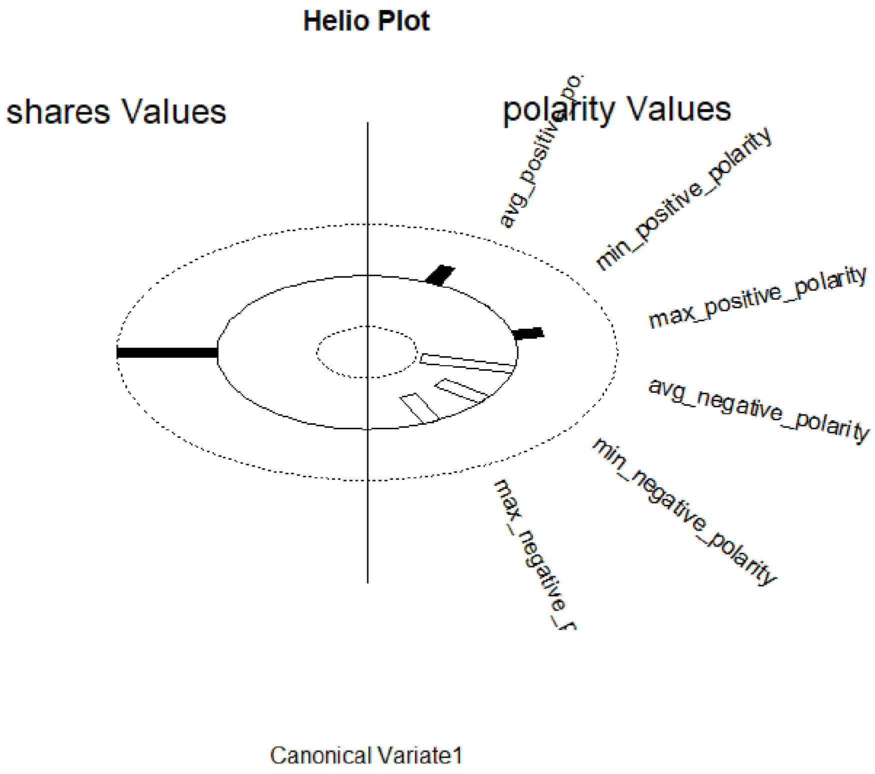
```
# Perform a chi-square test on C2

c9

ls(c9)

c9$chisq

c9$df

summary(c9)

round(pchisq(c9$chisq, c9$df, lower.tail=F), 3)
```

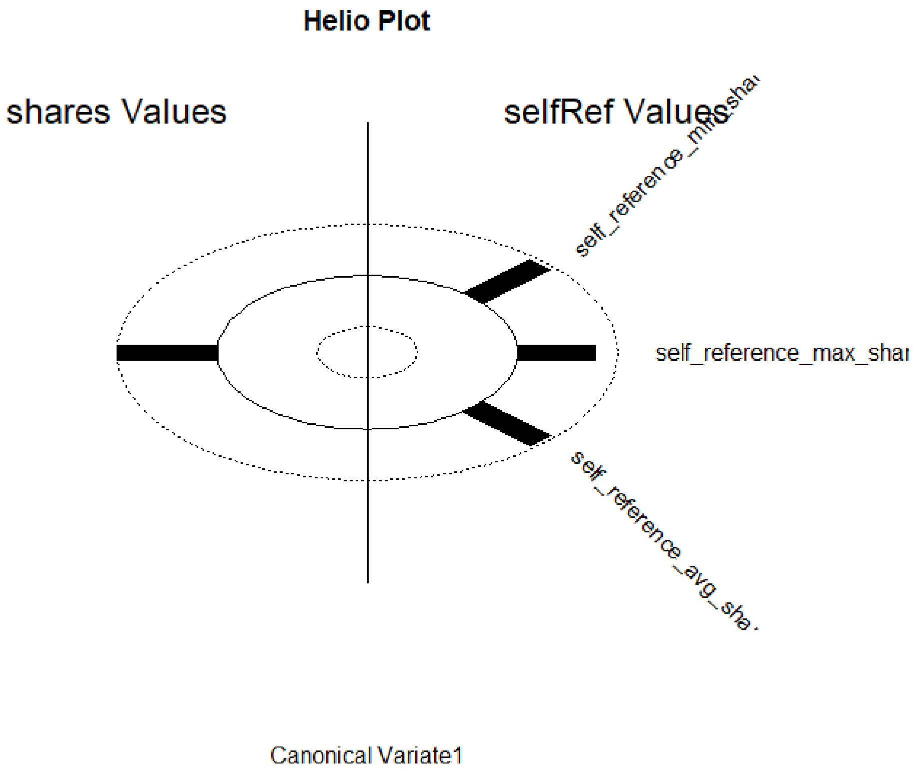

Helio Plot
shares Values
Data Values
data_channel_is_
data_channel_is_entertainme
data_channel_is_bus
data_channel_is_socmed
data_channel_is_tech
data_channel_is
Canonical Variate1

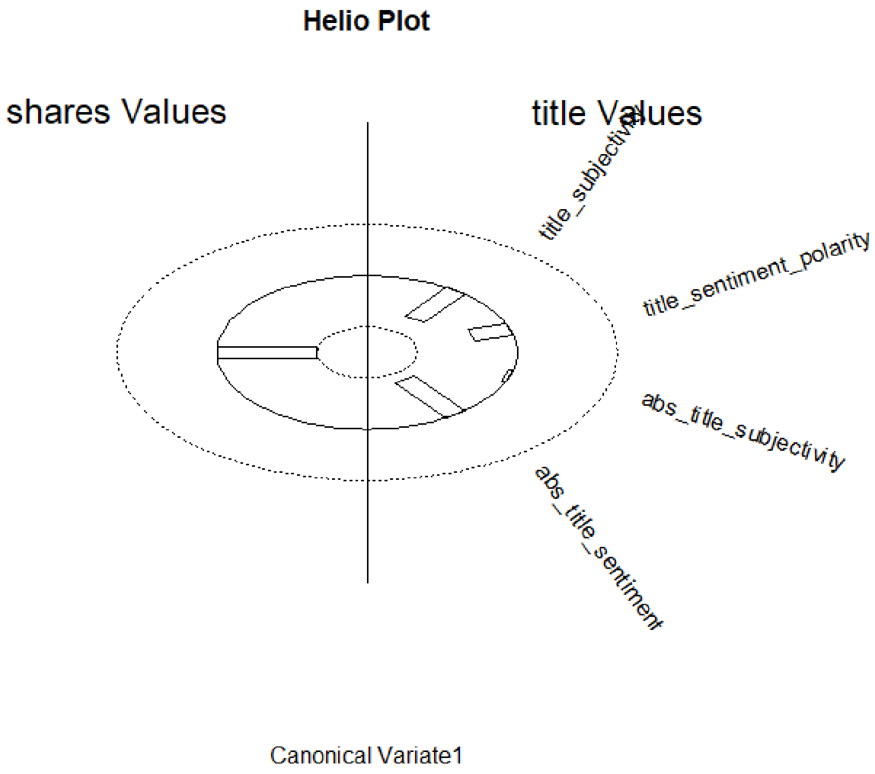

Helio Plot
shares Values
global Values
global_subjectivi
global_sentiment_polarity
global_rate_positive_word
global_rate_negative_wor
rate_positive_words
rate_negative_w
Canonical Variate1


Helio Plot
shares Values
keyword Values
kw_min_min
kw_max_min
kw_avg_min
kw_min_max
kw_max_max
kw_avg_max
kw_min_avg
kw_max_avg
kw_avg_avg
Canonical Variate1


Helio Plot
shares Values
LDA Values
LDA_00
LDA_01
LDA_02
LDA_03
LDA_04
Canonical Variate1


Helio Plot
shares Values
polarity Values
avg_positive_po
min_positive_polarity
max_positive_polarity
avg_negative_polarity
min_negative_polarity
max_negative_p
Canonical Variate1


Helio Plot
shares Values
selfRef Values
self_reference_min_shar
self_reference_max_shar
self_reference_avg_sha
Canonical Variate1


Helio Plot
shares Values
title Values
title_subjectivity
title_sentiment_polarity
abs_title_subjectivity
abs_title_sentiment
Canonical Variate1

2) Paper Review (10 points): An academic paper from a conference or Journal will be posted to the Homework 3 content section of D2L. It contains a usage of Canonical Correlation. Review the paper and evaluate their usage of Canonical Correlation. In particular, address **(Student burnout and work engagement a canonical correlation analysis)**

a) How suitable is their data for CC?

This study takes a look at student burnout in relation to work engagement. This is suitable for CCA because Canonical correlation analysis provides a way for explaining the relationship between 2 sets of variables using linear combinations of these variables.

b) How are they applying CC? What two groups of variables are being correlated? Are they metric, ordinal, nominal?

They are applying CCA to measure Burnout and Work engagement. These are metric.

c) What methods do they use to judge the quality of the correlation? Do they evaluate, and how do they evaluate the stability of the components?

The paper does not use KMO sampling adequacy or Bartlett's test; however, the paper does use Cronbach's alpha. Instead they use communalities across the two functions.

d) How many correlates do they concentrate on in their analysis, and do they attempt to interpret the correlates in terms of the original variables?

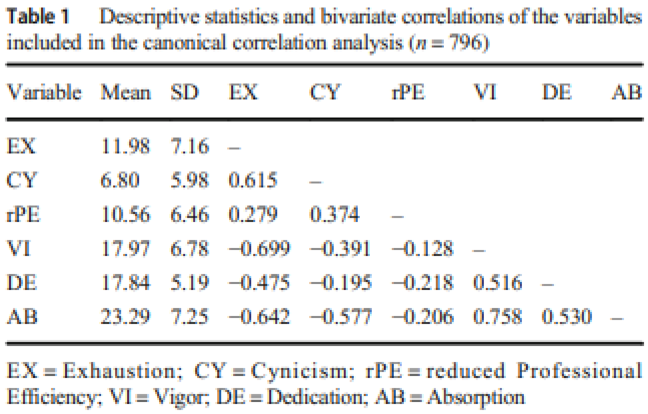


Table 1 Descriptive statistics and bivariate correlations of the variables included in the canonical correlation analysis ($n = 796$)

Variable	Mean	SD	EX	CY	rPE	VI	DE	AB
EX	11.98	7.16	–					
CY	6.80	5.98	0.615	–				
rPE	10.56	6.46	0.279	0.374	–			
VI	17.97	6.78	−0.699	−0.391	−0.128	–		
DE	17.84	5.19	−0.475	−0.195	−0.218	0.516	–	
AB	23.29	7.25	−0.642	−0.577	−0.206	0.758	0.530	–

EX = Exhaustion; CY = Cynicism; rPE = reduced Professional Efficiency; VI = Vigor; DE = Dedication; AB = Absorption

There are six correlated: (1) EX = Exhaustion, (2) CY = Cynicism, (3) rPE = reduced Professional Efficiency, (4) VI = Vigor, (5) DE = Dedication, (6) AB = Absorption.

EX, CY, rPE, when scored high would have high burnout. VI, DE, AB when scored high would have low burnout.

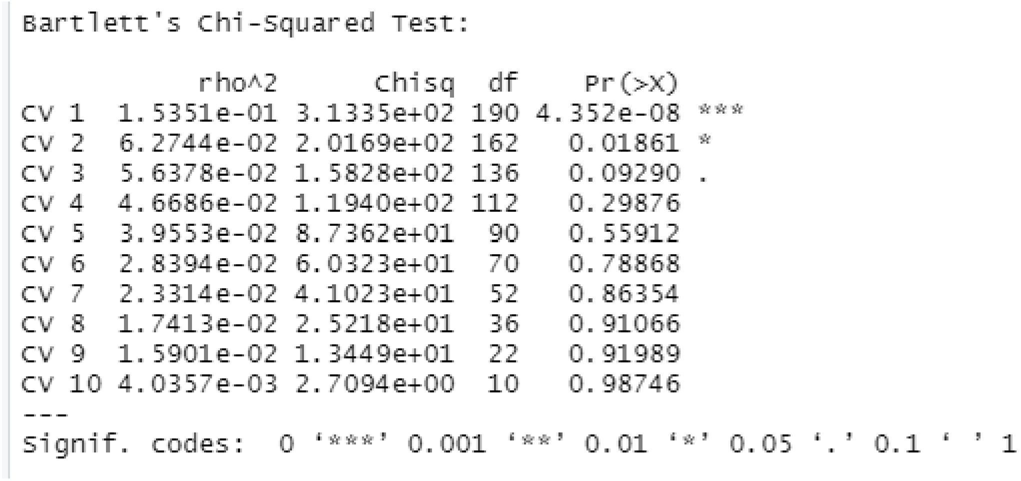
e) What conclusions does CC allow them to draw?

In this study they find out “there is a complex, yet, a strong relationship between burnout and work engagement among collegiate cycle students' (6, Conclusions).

3) **(20 points):** Perform the following Canonical Correlation Analysis on the Young People Survey from Lab 2: PCA/FA. Perform a canonical correlation analysis describing the relationships between the music and phobias variables using the data under the Lab 2: PCA/FA in the content folder).

1. Answer the following questions regarding the canonical correlations.

a. Test the null hypothesis that the canonical correlations are all equal to zero. Give your test statistic, d.f., and p-value.

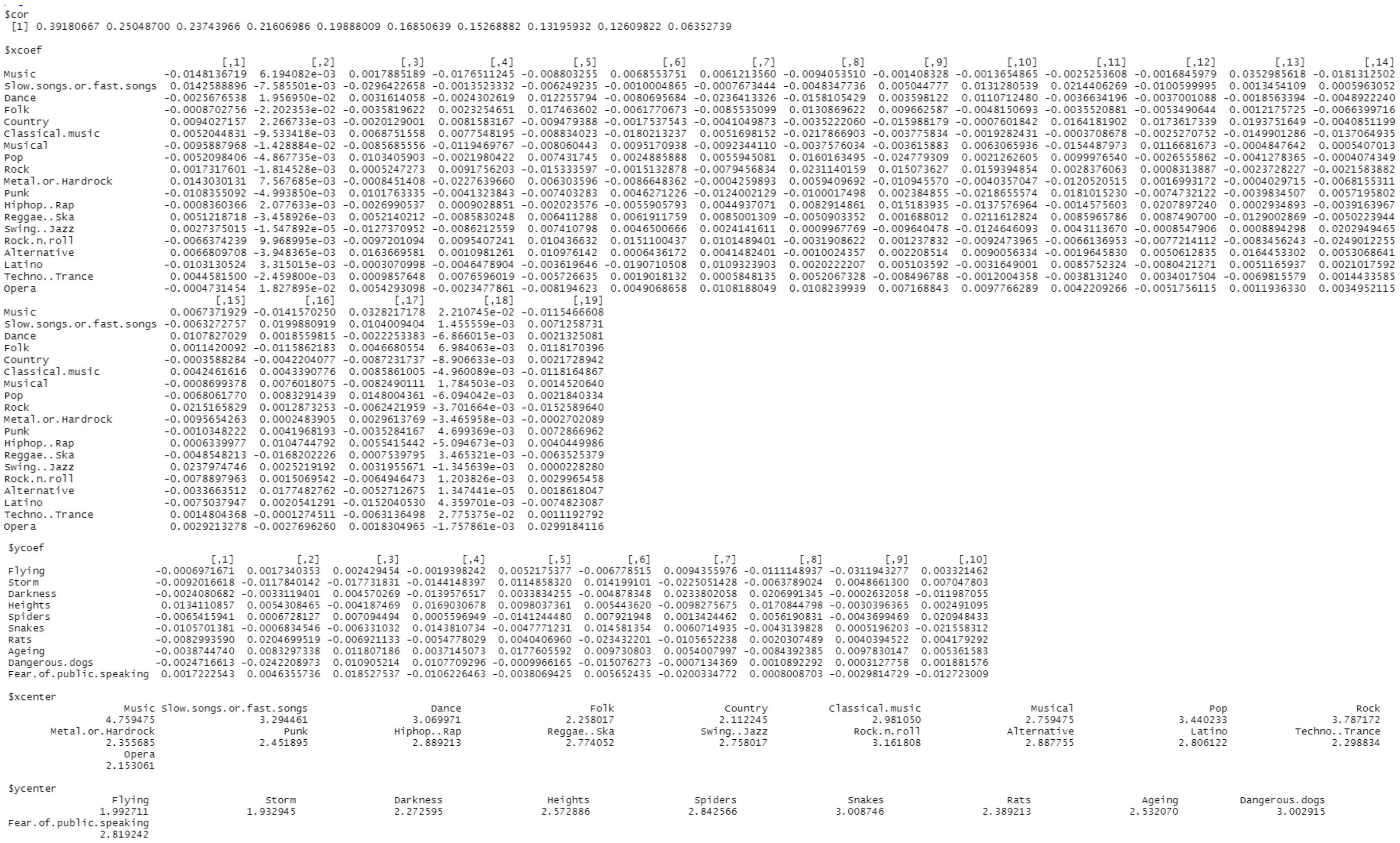


```
Bartlett's Chi-Squared Test:

           rho^2      Chisq  df    Pr(>X)    
CV 1  1.5351e-01 3.1335e+02 190 4.352e-08 ***
CV 2  6.2744e-02 2.0169e+02 162   0.01861 *  
CV 3  5.6378e-02 1.5828e+02 136   0.09290 .  
CV 4  4.6686e-02 1.1940e+02 112   0.29876    
CV 5  3.9553e-02 8.7362e+01  90   0.55912    
CV 6  2.8394e-02 6.0323e+01  70   0.78868    
CV 7  2.3314e-02 4.1023e+01  52   0.86354    
CV 8  1.7413e-02 2.5218e+01  36   0.91066    
CV 9  1.5901e-02 1.3449e+01  22   0.91989    
CV 10 4.0357e-03 2.7094e+00  10   0.98746    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

b. How many significant canonical variates are there?

Based on a, there are 10.

c. Present the first two canonical correlations (Cancor)?



```
$cor
 [1] 0.39180667 0.25048700 0.23743966 0.21606986 0.19888009 0.16850639 0.15268882 0.13195932 0.12609822 0.06352739

$xcoef
                                  [,1]          [,2]          [,3]          [,4]         [,5]          [,6]          [,7]          [,8]         [,9]         [,10]         [,11]         [,12]         [,13]         [,14]
Music                    -0.0148136719  6.194082e-03  0.0017885189 -0.0176511245 -0.008803255  0.0068553751  0.0061213560 -0.0094053510 -0.001408328 -0.0013654865 -0.0025253608 -0.0016845979  0.0352985618 -0.0181312502
Slow.songs.or.fast.songs  0.0142588896 -7.585501e-03 -0.0296422658 -0.0013523332 -0.006249235 -0.0010004865 -0.0007673444 -0.0048347736  0.005044777  0.0131280539  0.0214406269 -0.0100599995  0.0013454109  0.0005963052
Dance                    -0.0025676538  1.956950e-02  0.0031614058 -0.0024302619  0.012255794 -0.0080695684 -0.0236413326 -0.0158105429  0.003598122  0.0110712480 -0.0036634196 -0.0037001088 -0.0018563394 -0.0048922240
Folk                     -0.0008702756 -2.202353e-02 -0.0035819622  0.0023254651  0.017463602 -0.0061770673 -0.0085535099  0.0130869622  0.009662587 -0.0048150693 -0.0035520881 -0.0053490644  0.0012175725 -0.0066399716
Country                   0.0094027157  2.266733e-03 -0.0020129001  0.0081583167 -0.009479388 -0.0017537543 -0.0041049873 -0.0035222060 -0.015988179 -0.0007601842  0.0164181902  0.0173617339  0.0193751649 -0.0040851199
Classical.music           0.0052044831 -9.533418e-03  0.0068751558  0.0077548195 -0.008834023 -0.0180213237  0.0051698152 -0.0217866903 -0.003775834 -0.0019282431 -0.0003708678 -0.0025270752 -0.0149901286 -0.0137064935
Musical                  -0.0095887968 -1.428884e-02 -0.0085685556 -0.0119469767 -0.008060443  0.0095170938 -0.0092344110 -0.0037576034 -0.003615883  0.0063065936 -0.0154487973  0.0116681673 -0.0004847642  0.0005407013
Pop                      -0.0052098406 -4.867735e-03  0.0103405903 -0.0021980422  0.007431745  0.0024885888  0.0055945081  0.0160163495 -0.024779309  0.0021262605  0.0099976540 -0.0026555862 -0.0041278365 -0.0004074349
Rock                      0.0017317601 -1.814528e-03  0.0005247273  0.0091756203 -0.015333597 -0.0015132878 -0.0079456834  0.0231140159  0.015073627  0.0159394854  0.0028376063  0.0008313887 -0.0023728227 -0.0021583882
Metal.or.Hardrock         0.0143030131  7.567685e-03 -0.0008451408 -0.0227639660  0.006303596 -0.0086648362 -0.0004259893  0.0059409692 -0.010945570 -0.0040357047 -0.0120520515  0.0016993172 -0.0004029715 -0.0068155311
Punk                     -0.0108355092 -4.993850e-03  0.0101763335 -0.0041323843 -0.007403283  0.0046271226 -0.0124002129 -0.0100017498  0.002384855 -0.0218655574  0.0181015230 -0.0074732122 -0.0039834507  0.0057195802
Hiphop..Rap              -0.0008360366  2.077633e-03 -0.0026990537  0.0009028851 -0.002023576 -0.0055905793  0.0044937071  0.0082914861  0.015183935 -0.0137576964 -0.0014575603  0.0207897240  0.0002934893 -0.0039163967
Reggae..Ska               0.0051218718 -3.458926e-03  0.0052140212 -0.0085830248  0.006411288  0.0061911759  0.0085001309 -0.0050903352  0.001688012  0.0211612824  0.0085965786  0.0087490700 -0.0129002869 -0.0050223944
Swing..Jazz               0.0027375015 -1.547892e-05 -0.0127370952 -0.0086212559  0.007410798  0.0046500666  0.0024141611  0.0009967769 -0.009640478 -0.0124646093  0.0043113670 -0.0008547906  0.0008894298  0.0202949465
Rock.n.roll              -0.0066374239  9.968995e-03 -0.0097201094  0.0095407241  0.010436632  0.0151100437  0.0101489401 -0.0031908622  0.001237832 -0.0092473965 -0.0066136953 -0.0077214112 -0.0083456243 -0.0249012255
Alternative               0.0066809708 -3.948365e-03  0.0163669581  0.0010981261  0.010976142  0.0006436172  0.0041482401 -0.0010024357  0.002208514  0.0090056334 -0.0019645830  0.0050612835  0.0164453302  0.0053068641
Latino                   -0.0103130524  3.315015e-03 -0.0003070998 -0.0046478904 -0.003619646 -0.0190710508  0.0109323903  0.0020222207  0.005103592 -0.0031649001  0.0085752324 -0.0080421271  0.0051165937  0.0021017592
Techno..Trance            0.0044581500 -2.459800e-03  0.0009857648  0.0076596019 -0.005726635  0.0019018132  0.0005848135  0.0052067328 -0.008496788 -0.0012004358 -0.0038131240  0.0034017504 -0.0069815579  0.0014433585
Opera                    -0.0004731454  1.827895e-02  0.0054293098 -0.0023477861 -0.008194623  0.0049068658  0.0108188049  0.0108239939  0.007168843  0.0097766289  0.0042209266 -0.0051756115  0.0011936330  0.0034952115
                                 [,15]         [,16]         [,17]         [,18]         [,19]
Music                     0.0067371929 -0.0141570250  0.0328217178  2.210745e-02 -0.0115466608
Slow.songs.or.fast.songs -0.0063272757  0.0199880919  0.0104009404  1.455559e-03  0.0071258731
Dance                     0.0107827029  0.0018559815 -0.0022253383 -6.866015e-03  0.0021325081
Folk                      0.0011420092 -0.0115862183  0.0046680554  6.984063e-03  0.0118170396
Country                  -0.0003588284 -0.0042204077 -0.0087231737 -8.906633e-03  0.0021728942
Classical.music           0.0042461616  0.0043390776  0.0085861005 -4.960089e-03 -0.0118164867
Musical                  -0.0008699378  0.0076018075 -0.0082490111  1.784503e-03  0.0014520640
Pop                      -0.0068061770  0.0083291439  0.0148004361 -6.094042e-03  0.0021840334
Rock                      0.0215165829  0.0012873253 -0.0062421959 -3.701664e-03 -0.0152589640
Metal.or.Hardrock        -0.0095654263  0.0002483905  0.0029613769 -3.465958e-03 -0.0002702089
Punk                     -0.0010348222  0.0041968193 -0.0035284167  4.699369e-03  0.0072866962
Hiphop..Rap               0.0006339977  0.0104744792  0.0055415442 -5.094673e-03  0.0040449986
Reggae..Ska              -0.0048548213 -0.0168202226  0.0007539795  3.465321e-03 -0.0063525379
Swing..Jazz               0.0237974746  0.0025219192  0.0031955671 -1.345639e-03  0.0000228280
Rock.n.roll              -0.0078897963  0.0015069542 -0.0064946473  1.203826e-03  0.0029965458
Alternative              -0.0033663512  0.0177482762 -0.0052712675  1.347441e-05  0.0018618047
Latino                   -0.0075037947  0.0020541291 -0.0152040530  4.359701e-03 -0.0074823087
Techno..Trance            0.0014804368 -0.0001274511 -0.0063136498  2.775375e-02  0.0011192792
Opera                     0.0029213278 -0.0027696260  0.0018304965 -1.757861e-03  0.0299184116

 $ycoef
                                  [,1]          [,2]         [,3]          [,4]          [,5]         [,6]          [,7]          [,8]          [,9]        [,10]
 Flying                  -0.0006971671  0.0017340353  0.002429454 -0.0019398242  0.0052175377 -0.006778515  0.0094355976 -0.0111148937 -0.0311943277  0.003321462
 Storm                   -0.0092016618 -0.0117840142 -0.017731831 -0.0144148397  0.0114858320  0.014199101 -0.0225051428 -0.0063789024  0.0048661300  0.007047803
 Darkness                -0.0024080682 -0.0033119401  0.004570269 -0.0139576517  0.0033834255 -0.004878348  0.0233802058  0.0206991345 -0.0002632058 -0.011987055
 Heights                  0.0134110857  0.0054308465 -0.004187469  0.0169030678  0.0098037361  0.005443620 -0.0098275675  0.0170844798 -0.0030396365  0.002491095
 Spiders                 -0.0065415941  0.0006728127  0.007094494  0.0005596949 -0.0141244480  0.007921948  0.0013424462  0.0056190831 -0.0043699469  0.020948433
 Snakes                  -0.0105701381 -0.0006834546 -0.006331032  0.0143810734 -0.0047771231  0.014581354  0.0060714935 -0.0043139828  0.0005196203 -0.021558312
 Rats                    -0.0082993590  0.0204699519 -0.006921133 -0.0054778029  0.0040406960 -0.023432201 -0.0105652238  0.0020307489  0.0040394522  0.004179292
 Ageing                  -0.0038744740  0.0083297338  0.011807186  0.0037145073  0.0177605592  0.009730803  0.0054007997 -0.0084392385  0.0097830147  0.005361583
 Dangerous.dogs          -0.0024716613 -0.0242208973  0.010905214  0.0107709296 -0.0009966165 -0.015076273 -0.0007134369  0.0010892292  0.0003127758  0.001881576
 Fear.of.public.speaking  0.0017222543  0.0046355736  0.018527537 -0.0106226463 -0.0038069425  0.005652435 -0.0200334772  0.0008008703 -0.0029814729 -0.012723009

 $xcenter
                    Music Slow.songs.or.fast.songs                    Dance                     Folk                  Country          Classical.music                  Musical                      Pop                     Rock
                 4.759475                 3.294461                 3.069971                 2.258017                 2.112245                 2.981050                 2.759475                 3.440233                 3.787172
        Metal.or.Hardrock                     Punk              Hiphop..Rap              Reggae..Ska              Swing..Jazz              Rock.n.roll              Alternative                   Latino           Techno..Trance
                 2.355685                 2.451895                 2.889213                 2.774052                 2.758017                 3.161808                 2.887755                 2.806122                 2.298834
                    Opera
                 2.153061

 $ycenter
                  Flying                   Storm                Darkness                 Heights                 Spiders                  Snakes                    Rats                  Ageing          Dangerous.dogs
                1.992711                1.932945                2.272595                2.572886                2.842566                3.008746                2.389213                2.532070                3.002915
 Fear.of.public.speaking
                2.819242
```

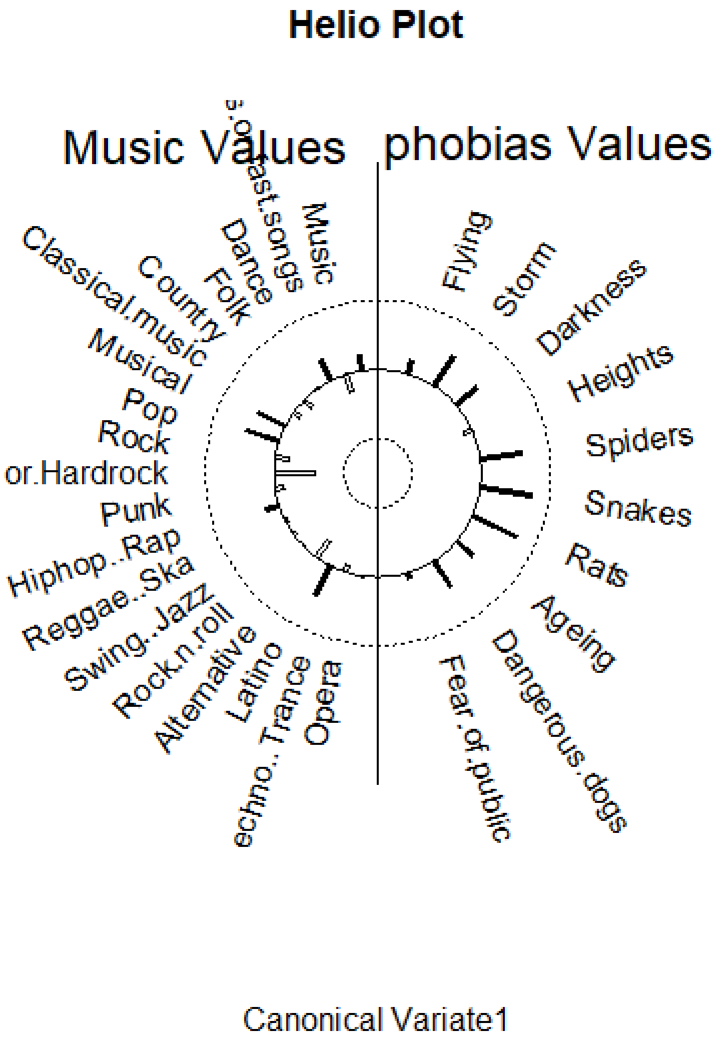
d. What can you conclude from the above analyses?

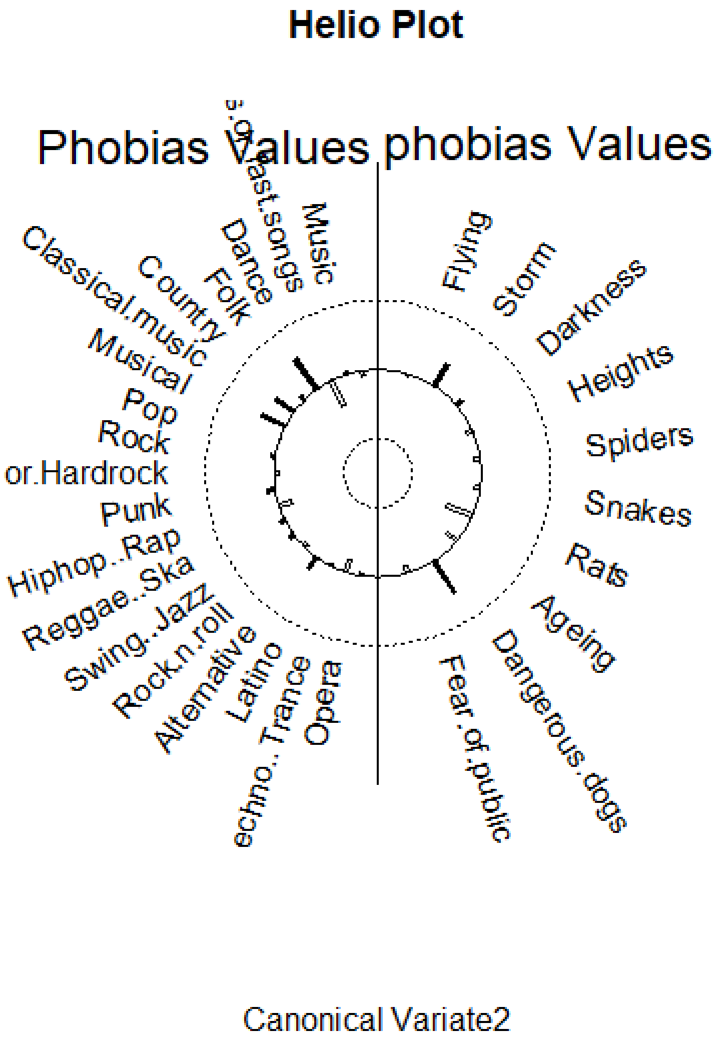
Based on the above output, all correlation values seem to be close/near to zero or really small numbers. This means that the variables do not seem to be very significant and correlation is very small.

2. Answer the following questions regarding the canonical variates.

a. Give the formula for the first canonical variate for the music and phobias variables.

b. Give the correlations between the first canonical variate for music and the phobias variables.


Helio Plot
Music Values
phobias Values
s.or.fast.songs
Music
Dance
Folk
Country
Classical.music
Musical
Pop
Rock
or.Hardrock
Punk
Hiphop..Rap
Reggae..Ska
Swing..Jazz
Rock.n.roll
Alternative
Latino
echno..Trance
Opera
Flying
Storm
Darkness
Heights
Spiders
Snakes
Rats
Ageing
Dangerous.dogs
Fear.of.public
Canonical Variate1


Helio Plot
Phobias Values
phobias Values
s.or.fast.songs
Music
Dance
Folk
Country
Classical.music
Musical
Pop
Rock
or.Hardrock
Punk
Hiphop..Rap
Reggae..Ska
Swing..Jazz
Rock.n.roll
Alternative
Latino
echno..Trance
Opera
Flying
Storm
Darkness
Heights
Spiders
Snakes
Rats
Ageing
Dangerous.dogs
Fear.of.public
Canonical Variate2

c. What can you conclude from the above analyses?

Based on CV1 you can see that Classical, musical, pop and latino music all have strong positive correlations while dance is the only strong negative correlation. For phobias, all had a strong positive correlation except fears of public and height.

Based on CV2 folk, classical, musical and pop that have the high positive correlations and dance is the only weak correlation. The phobias all had negative correlations except for storms, darkness and dangerous dogs.

```
#Libraries

library(Hmisc) #Describe Function

library(psych) #Multiple Functions for Statistics and Multivariate Analysis

library(GGally) #ggpairs Function

library(ggplot2) #ggplot2 Functions

library(vioplot) #Violin Plot Function

library(corrplot) #Plot Correlations

library(REdaS) #Bartlett's Test of Sphericity

library(psych) #PCA/FA functions
library(factoextra) #PCA Visualizations

library("FactoMineR") #PCA functions

library(ade4) #PCA Visualizations

library(foreign)

library(CCA)

library(yacca)

library(MASS)

##############################################################
################################



ccaWilks = function(set1, set2, cca)

{

  ev = ((1 - cca$cor^2))

  ev
```

```
n = dim(set1)[1]
p = length(set1)
q = length(set2)
k = min(p, q)
m = n - 3/2 - (p + q)/2
m

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)

}


################################################################
#################################
```

```
#Set Working Directory

setwd("C:/Users/jdoretti/Documents/DSC 424")



#Read in Datasets


responses <- read.csv(file="responses.csv", header=TRUE, sep=",")


#Check Sample Size and Number of Variables

dim(responses)

#1,010-Sample Size and 150 variables


#Show for first 6 rows of data

head(responses)
```

```
names(responses)


##############################################################
##################################


#Check for Missing Values (i.e. NAs)


#For All Variables

sum(is.na(responses))

#571 total missing values (571 cells with missing data)



#Treat Missing Values


#Listwise Deletion

responses2 <- na.omit(responses)
```

```
#Check new data has no missing data

sum(is.na(responses2))



##############################################################
###################################################


#Show Structure of Dataset

str(responses2, list.len=ncol(responses2))


#Show column Numbers

names(responses2)


#Categorical Variables (Var_num):  Smoking (74), Alcohol (75), Punctuality
(108), Lying (109), Internet.usuage (133), Gender (145),
#                  Left...right.handed (146), Education (147), Only.child(148),
Village.town (149), House...block.of.flats (150)


#Create new subsets of data (Numeric Variables Only)


responses3 <- responses2[,c(1:73,76,77:107,110:132,134:140,141:144)]


music <- responses2[,1:19]

movie <- responses2[,20:31]

hobbies_interests <- responses2[,32:63]

phobias <- responses2[,64:73]

health <- responses2[,76]

personality_views_opinions <- responses2[,c(77:107,110:132)]

spending <- responses2[,134:140]

demographics <- responses2[,141:144]
```

```
# This gives us the cannonical correlates, but no significance tests
c = cancor(music, phobias)
c

#Breakdown of the Correlations
matcor(music, phobias)

#Correlations between sepal and sepal (X)
#Correlations between petal and petal (Y)
cc_mm = cc(music, phobias)
cc_mm$cor

#Functions for CCA
ls(cc_mm)
```

```
#XCoef Correlations

cc_mm$xcoef


#YCoef Correlations

cc_mm$ycoef


#Calculate Scores

loadings_mm = comput(music, phobias, cc_mm)

ls(loadings_mm)


#Correlation X Scores

loadings_mm$corr.X.xscores


#Correlation Y Scores
```

```
loadings_mm$corr.Y.yscores

#Wilk's Lambda Test
wilks_mm = ccaWilks(music, phobias, cc_mm)
round(wilks_mm, 2)

# Now, let's calculate the standardized coefficients
s1 = diag(sqrt(diag(cov(music))))
s1 %*% cc_mm$xcoef

s2 = diag(sqrt(diag(cov(phobias))))
s2 %*% cc_mm$ycoef

# A basic visualization of the cannonical correlation
plt.cc(cc_mm)
```

```
###################################################################
##

# Now, let's try it with yacca

###################################################################
##



library(yacca)

c2 = cca(music,phobias)

summary(c2)



#CV1

helio.plot(c2, cv=1, x.name="Music Values",

          y.name="phobias Values")



#CV2
```

```
helio.plot(c2, cv=2, x.name="Phobias Values",

          y.name="phobias Values")


#Function Names

ls(c2)


# Perform a chi-square test on C2

c2

ls(c2)

c2$chisq

c2$df

summary(c2)

round(pchisq(c2$chisq, c2$df, lower.tail=F), 3)
```

EXTRA CREDIT (10 points) Perform a correspondence analysis on the Reading Level and Education Level Completed liking data in readers.csv. In this file you are provided with the table for the two sets of categories. In particular perform the following

E1-Some Primary

E2-Primary Completed

E3-Some Secondary\

E4-Secondary Completed

E5-Some tertiary

C1-Reading at a Glance

C2-Read Fairly Thoroughly

C3-Read Very Through

a) Create a mosaic plot of the two categorical variables.
b) Plot the results of the correspondence analysis
c) With each country, create a profile for the Reading Level. Which Reading Level are most highly and least highly represented? For each Education Level Completed, draw the scale for that Education Level completed and demonstrate that Reading Level profile on the graph.