CCA Analyses - Main Script

Session Info

I ran the code in a docker container, for better repoducibility :)

Code to docker image:

> docker pull bignardig/rocker3.6.2:yb64s

> docker run -d -p 8787:8787 -e PASSWORD=YOURPASSWORDHERE -v "/mnt/c/users/. . . /Paper 2 - Cumulative Risk/Analysis:/home/rstudio/" bignardig/rocker3.6.2:yb64s

Note that the above code works when running ubuntu on windows using WSL2. The /mnt/ part in the path specifies the windows file system.

I also did the following (in WSL2 ubuntu):

> docker exec -it CONTAINERID bash

Then i run the following:

> ulimit -s unlimited

Most of my CCA analysis code has been moved to its own (in development) R package on github (ccatools), linked below. The code will pull a specific version of that package, so it will be completely reproducible in the future!

Note that if your running the code from a fresh rocker container, you need to do the following:

> sudo apt-get update
> sudo apt-get install libxml2 sudo apt-get install zlib1g-dev sudo apt-get install libgsl-dev sudo apt-get install xclip

The last one is for clipr:: to work!

```
sessionInfo()
```

```
## R version 3.6.2 (2019-12-12)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Debian GNU/Linux 10 (buster)
##
## Matrix products: default
## BLAS/LAPACK: /usr/lib/x86_64-linux-gnu/libopenblasp-r0.3.5.so
##
## locale:
```

```
##  [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8        LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=C
##  [7] LC_PAPER=en_US.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## loaded via a namespace (and not attached):
##  [1] compiler_3.6.2  magrittr_1.5    tools_3.6.2     htmltools_0.4.0
##  [5] yaml_2.2.1      Rcpp_1.0.3      stringi_1.4.6   rmarkdown_2.1
##  [9] knitr_1.28      stringr_1.4.0   xfun_0.12       digest_0.6.25
## [13] rlang_0.4.4     evaluate_0.14
```

Load Data, Functions and Libraries Required

```
source("Script P2S0 - Load Analysis Environment.R")

# Load my own R (work in progress) R package for CCA analyses

if (!("ccatools" %in% installed.packages())) {
  devtools::install_github("giac01/ccatools@bd358f15fdec486a8cb06480fb0bf5dd66ada757") #The @ specifies
}


# library(usethis)
# library(testthat)
library(ccatools)
library(tidyverse)
library(patchwork)



REPEATS_BOOT = 10000
REPEATS_BOOT
```

Number of participants

```
DataSets = list(df0_imputed_group1,
                df0_imputed_group2,
                df0_imputed_BothGroups,
                df0_imputed_Teach_group1,
                df0_imputed_Teach_group2,
                df0_imputed_Teach_BothGroups)

sapply(DataSets,nrow) %>%
  `names<-`(c("Training", "Testing", "Combined",
              "Training Teach", "Testing Teach", "Combined Teach"))
```

```
##       Training        Testing       Combined Training Teach  Testing Teach
##           5160           5216          10376            2831           2834
## Combined Teach
##           5665
```

Variable Labels

```
cbind(all_var_labels2, all_var2) %>% kable()
```

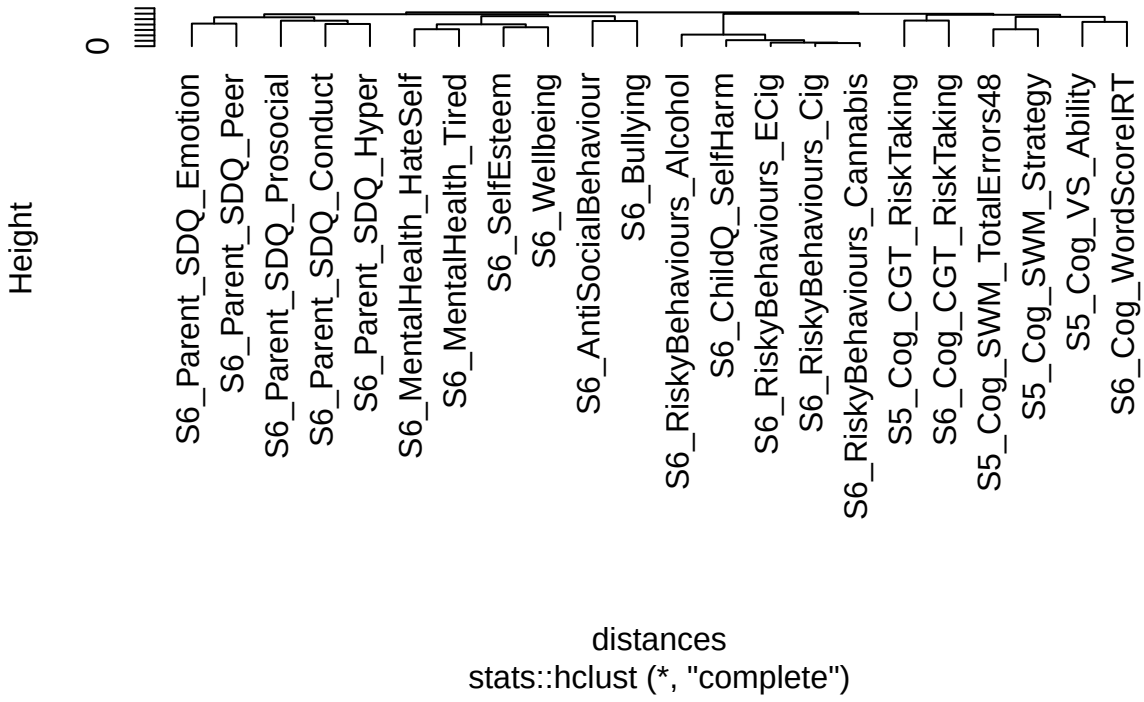
all_var_labels2	all_var2
PR1 SES-Education	S1_SES_EducationAvg
PR1 SES-Employment	S1_SES_AverageLastJob
PR1 SES-Income	S1_SES_EquivIncome_Norm
PR2 Two Carers at Home	S2_NumberParents
PR1 Num Siblings	S1_Fam_NSiblings
PR1 Preterm Birth	S1_Health_PretermBirth
PR1 Low Birth Weight	S1_Health_LowBirthWeight
PR1 Parent General Health	S1_Parent_GeneralHealth
PR1 Smoked During Pregnancy	S1_Health_SmokedThroughoutPregnancy
PR1 Breastfeeding	S1_Health_EverTriedBreastFeeding
PR1 Current Smoking	S1_Health_CurrentSmoking
PR1 Current Alcohol	S1_Parent_AlcoholCurrent
PR1 Alcohol During Pregnancy	S1_Parent_AlcoholPregnant
PR1 Planned Pregnancy	S1_Fam_PlannedPregnancy
PR1 Feelings About Pregnancy	S1_FeelingsPregnancy
S12 Air Pollution	S12_AirPollution
PR1 Parent Relationship Quality	S1_PWB_RelationshipQuality
PR1 Parent Self Esteem	S1_PWB_SelfEsteem
PR1 Parent Mental Health	S1_PWB_ParentMentalHealth
PR2 Parent Mental Health	S2_PWB_ParentMentalHealth
PR1 Parent Locus Of Control	S1_PWB_LocusOfControl
PR1 Neighbourhood Quality	S1_ParentNeighbourhoodRating
S1 Index Multiple Deprivation	S1_IMDDeprivation
PR1 Parent Literacy Problems	S1_ParentLiteracyProblems
PR1 Housing Quality	S1_HousingQuality
OR2 Housing Quality	S2_HomeRating
OR2 Parent Responsivity	S2_ParentRating
PR2 Home Learning Environment	S2_HelpWithLearning
PR2 Recreational Drug Use	S2_Parent_RecreationalDrugUse
PR2 Harsh Parenting	S2_HarshParenting
PR2 Monolingual-English Home	S2_Misc_EnglishMonolingualHome
TR5 SDQ Emotional Problems	S5_Teach_SDQ_Emotion
TR5 SDQ Conduct Problems	S5_Teach_SDQ_Conduct
TR5 SDQ Hyperactivity	S5_Teach_SDQ_Hyper
TR5 SDQ Peer Problems	S5_Teach_SDQ_Peer
TR5 SDQ Prosocial	S5_Teach_SDQ_Prosocial
TR5 Academic Ability	S5_TeacherRating
PR6 SDQ Emotional Problems	S6_Parent_SDQ_Emotion
PR6 SDQ Conduct Problems	S6_Parent_SDQ_Conduct
PR6 SDQ Hyperactivity	S6_Parent_SDQ_Hyper
PR6 SDQ Peer Problems	S6_Parent_SDQ_Peer

all_var_labels2	all_var2
PR6 SDQ Prosocial	S6_Parent_SDQ_Prosocial
CR6 Depression	S6_MentalHealth_HateSelf
CR6 Anhedonia	S6_MentalHealth_Tired
CR6 Self Esteem	S6_SelfEsteem
CR6 Wellbeing	S6_Wellbeing
CR6 Cigarette Use	S6_RiskyBehaviours_Cig
CR6 E-Cigarette Use	S6_RiskyBehaviours_ECig
CR6 Alcohol Use	S6_RiskyBehaviours_Alcohol
CR6 Cannabis Use	S6_RiskyBehaviours_Cannabis
CR6 Self Harm	S6_ChildQ_SelfHarm
CR5 Vocabulary	S5_Cog_VS_Ability
CR6 Vocabulary	S6_Cog_WordScoreIRT
CC5 Spatial WM Total Errors	S5_Cog_SWM_TotalErrors48
CC5 Spatial WM Strategy	S5_Cog_SWM_Strategy
CC5 CGT Risk Taking	S5_Cog_CGT_RiskTaking
CC6 CGT Risk Taking	S6_Cog_CGT_RiskTaking
CR6 Anti-Social Behaviour	S6_AntiSocialBehaviour
CR6 Bullying Others	S6_Bullying

Analysis 0 - Hierarchial cluster analysis of variables

To aid visualisation, we cluster the predictor and outcome variables into 4 (5 groups including teacher metrics)

```
distances = stats::dist(t(df0_imputed_BothGroups[,outcomes]), method="manhattan")
  hclust_results = stats::hclust(distances)
  Outcome_Order = hclust_results$order
  Outcome_Ordered = outcomes[Outcome_Order]
  Outcome_Labels_Ordered = outcomes_labels[Outcome_Order]
  plot(hclust_results, hang=-1, main = "Hierarchial Clustering of Outcome Variables")
```




```
  Outcome_Cluster = stats::cutree(hclust_results, 4)
# sort(Outcome_Cluster)
  Outcome_Cluster_Labels = Convert(Outcome_Cluster,1:4,c("Behavioural Problems","Mental Health", "Drug-

# For ease of presentation, teacher's responses are not clustered in the usual way, but just tagged on
  Outcome_Cluster2 = c(rep(0,6),Outcome_Cluster)
  Outcome_Ordered2 = c(outcomes2[1:6],Outcome_Ordered)
  Outcome_Labels_Ordered2 = c(outcomes_labels2[1:6],Outcome_Labels_Ordered)
  names(Outcome_Cluster2)[1:6] = outcomes2[1:6]
  Outcome_Cluster_Labels2 = Convert(Outcome_Cluster2,0:4,c("Teacher Ratings","Behavioural Problems","Me

# not used in results  - but out of curiorsity looked at clustering all outcomes with teacher-reports
#
#       hclust_results = hclust(dist(t(df0_imputed_Teach[,outcomes2]), method="manhattan"))
#       plot(hclust_results, hang=-1)

#Cluster Predictors
distances = dist(t(df0_imputed_group1[,envvar]), method="manhattan")
hclust_results = hclust(distances)
envvar_Order = hclust_results$order
envvar_Ordered = envvar[envvar_Order]
envvar_Labels_Ordered = envvar_labels[envvar_Order]
# plot(hclust_results, hang=-1)
```

Analysis 1

Quick Check - does my cca function give identical results to stats::cancor ? (the answer is yes)

R output

cancor_Xweights & cancor_Yweights are the normalised (in the linear algebra sense) raw coefficients for a cca model in df0_imputed_group1 !

```
cca_cancor = stats::cancor(x=scale(df0_imputed_group1[,envvar]), y=scale(df0_imputed_group1[,outcomes]))


  cancor_Xweights = apply(cca_cancor$xcoef,2, function(x) x/sqrt(sum(x^2)))
  cancor_Yweights = apply(cca_cancor$ycoef,2, function(x) x/sqrt(sum(x^2)))

cca_cancor$cor[1:10]
```

```
##  [1] 0.54093522 0.34748737 0.26853420 0.18871501 0.15796696 0.14148354
##  [7] 0.13281019 0.11711884 0.09967211 0.09383745
```

```
cancor_Xweights[1:5,1:5]
```

```
##                              [,1]        [,2]        [,3]        [,4]
## S1_SES_EducationAvg      0.68506861 -0.01489136 -0.35853172  0.21439176
## S1_SES_AverageLastJob    0.22601600 -0.10297593 -0.25349888  0.02808576
## S1_SES_EquivIncome_Norm  0.30417178 -0.12079440 -0.07360900 -0.14089545
## S2_NumberParents         0.11384118 -0.07533637  0.02533193 -0.09202954
## S1_Fam_NSiblings        -0.09387529  0.07476440  0.45723661  0.33689646
##                              [,5]
## S1_SES_EducationAvg      0.37026712
## S1_SES_AverageLastJob    0.23987077
## S1_SES_EquivIncome_Norm  0.02689835
## S2_NumberParents        -0.38972088
## S1_Fam_NSiblings         0.37944345
```

ccatools::.cca output

Result is the same apart from sign difference in loadings, which is why some negative canonical correlations exist when looking at correlations between variate scores.

```
cca_group1 = ccatools::.cca(X_FIT  = as.matrix(df0_imputed_group1[envvar]),
                            Y_FIT  = as.matrix(df0_imputed_group1[outcomes]))
cca_group1$cc_fit
```

```
##  [1] 0.54093522 0.34748737 0.26853420 0.18871501 0.15796696 0.14148354
##  [7] 0.13281019 0.11711884 0.09967211 0.09383745
```

```
#gb_test$cc_pred
cca_group1$xcoef[1:5,1:5]
```

```
##              [,1]        [,2]        [,3]        [,4]        [,5]
## [1,]  0.68506861 -0.01489136  0.35853172  0.21439176 -0.37026712
## [2,]  0.22601600 -0.10297593  0.25349888  0.02808576 -0.23987077
## [3,]  0.30417178 -0.12079440  0.07360900 -0.14089545 -0.02689835
## [4,]  0.11384118 -0.07533637 -0.02533193 -0.09202954  0.38972088
## [5,] -0.09387529  0.07476440 -0.45723661  0.33689646 -0.37944345
```

```
rm(cca_group1)
```

We can flip the sign loadings by rotated the weights to align to the R output:

```
cca_group1 = ccatools:::cca(
                      X_FIT  = as.matrix(df0_imputed_group1[envvar]),
                      Y_FIT  = as.matrix(df0_imputed_group1[outcomes]),
                      ProcrustX = cancor_Xweights[,1:22],
                      ProcrustY = cancor_Yweights[,1:22],
                      ncomp=22
              )
cca_group1$cc_fit
```

```
##  [1] 0.54093522 0.34748737 0.26853420 0.18871501 0.15796696 0.14148354
##  [7] 0.13281019 0.11711884 0.09967211 0.09383745 0.08991303 0.08452793
## [13] 0.07696123 0.06991164 0.06085872 0.05763233 0.04888958 0.04286058
## [19] 0.03597884 0.03040029 0.02372695 0.01900397
```

```
#gb_test$cc_pred
cca_group1$xcoef[1:5,1:5]
```

```
##              [,1]        [,2]        [,3]        [,4]        [,5]
## [1,]  0.68506861 -0.01489136 -0.35853172  0.21439176  0.37026712
## [2,]  0.22601600 -0.10297593 -0.25349888  0.02808576  0.23987077
## [3,]  0.30417178 -0.12079440 -0.07360900 -0.14089545  0.02689835
## [4,]  0.11384118 -0.07533637  0.02533193 -0.09202954 -0.38972088
## [5,] -0.09387529  0.07476440  0.45723661  0.33689646  0.37944345
```

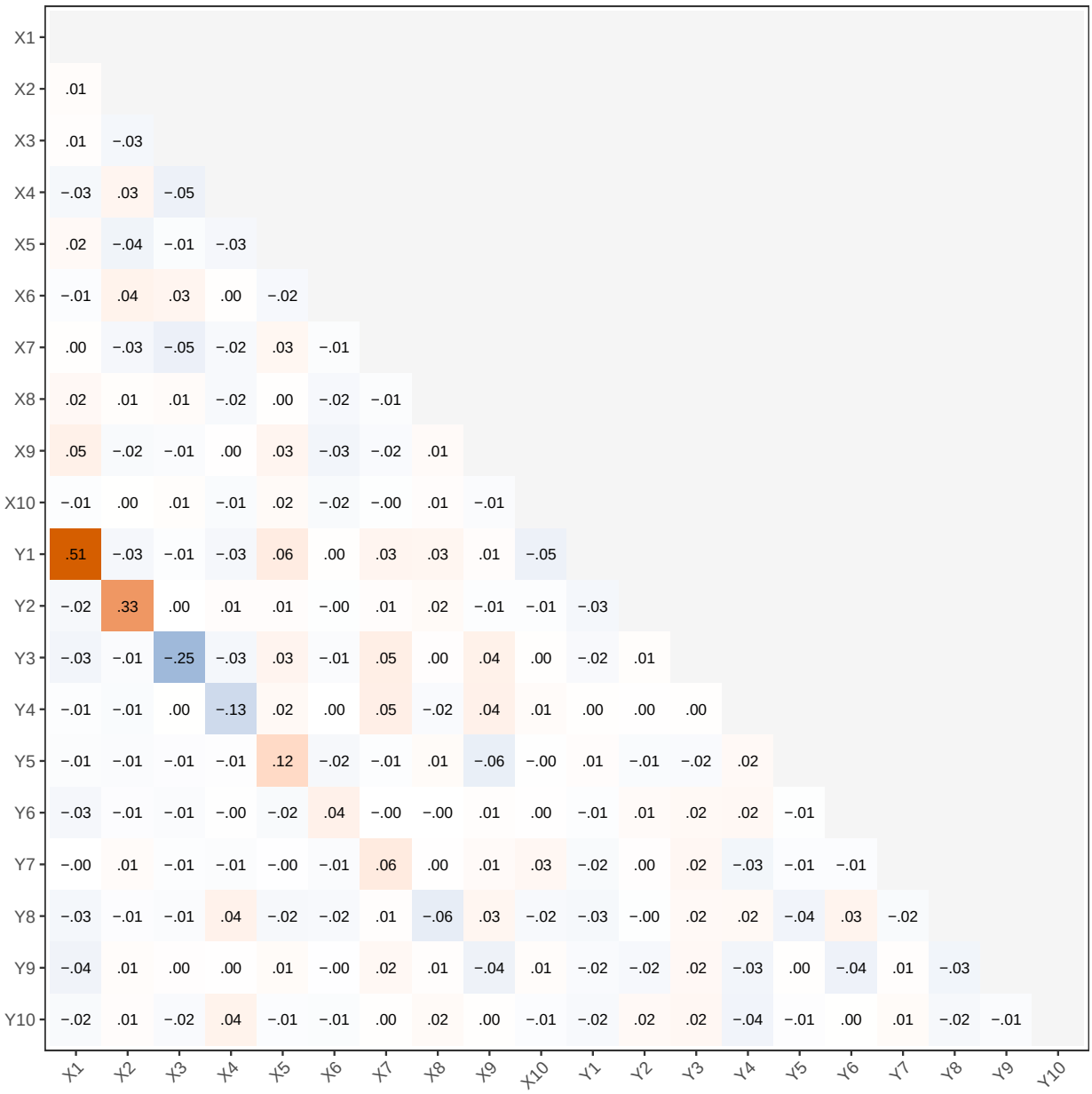
Run split half analyses

Fit Model in training dataset, and then generate canonical variates for testing dataset, and plot correlations between them.

Figure: Below diagonical we have correlations between canonical variates. X1-X10 are latent factors extracted from predictor variables, and Y1-Y10 are latent factors extracted from outcome variables.

```
#gb_CCA is defined in script P2S2!
NVARIATES = 10 # Number of variates to extract

cca_splithalf_noprocrustes = ccatools::cca_splithalf(
              X_FIT  = df0_imputed_group1[envvar],                    # Data
              Y_FIT  = df0_imputed_group1[outcomes],                  # Data
              X_PRED = df0_imputed_group2[envvar],                    # Data
              Y_PRED = df0_imputed_group2[outcomes],                  # Data
              ncomp = NVARIATES)                                      # Numb
```

```
ccatools::plotcor(cca_splithalf_noprocrustes$model_results$variates, abs_colour = FALSE, reportCI = FALS
```

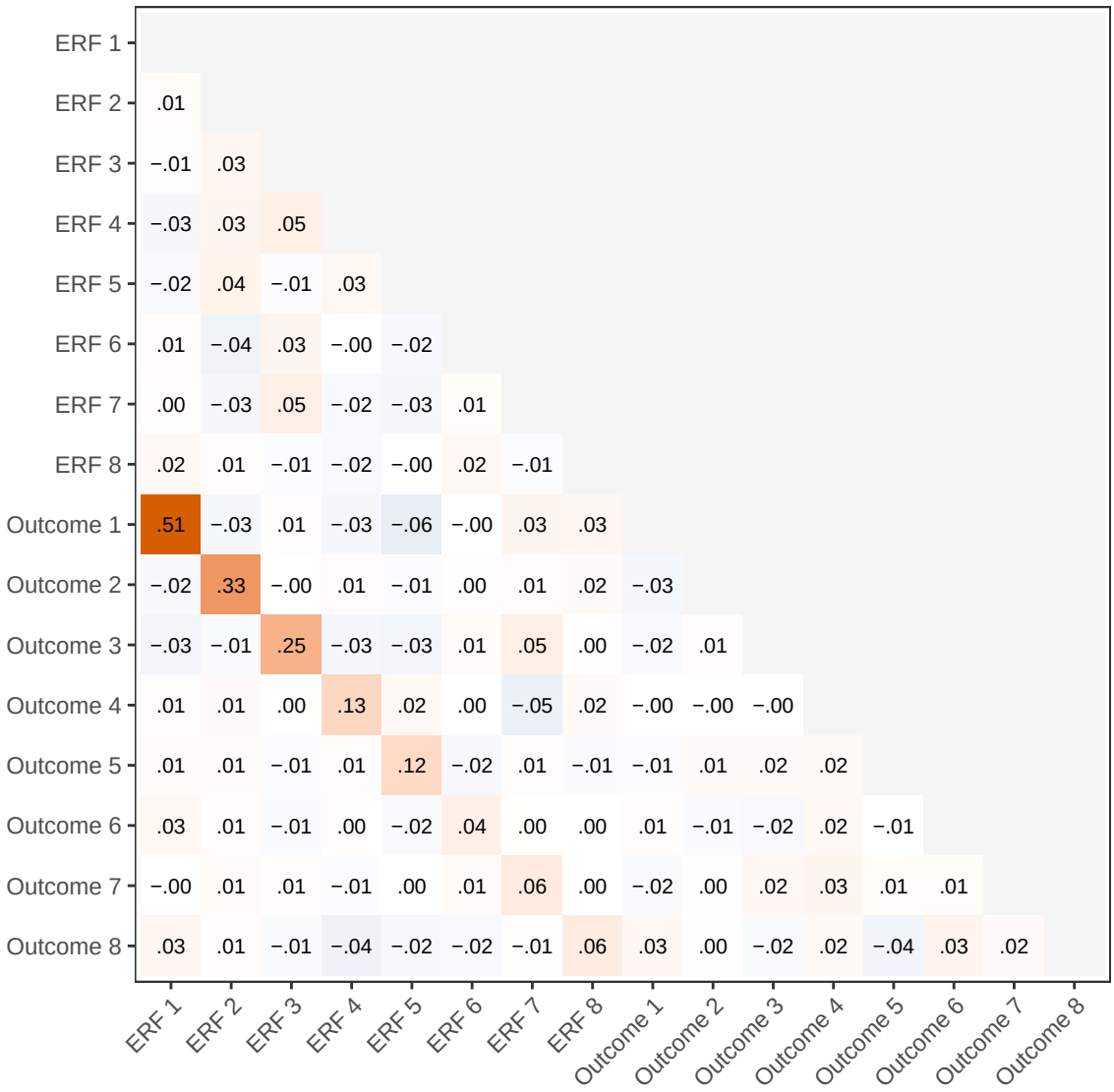


Because it gets confusing to interpret negative correlations (as most raw variables have been coded so that higher == better), we can flip the extracted CCA loadings to ensure that canonical correlations will be positive.

An easy way to do this given how the function is coded is to make sure the extracted loadings match the stats::cancor output which automatically corrects for this...

```
NVARIATES = 22
NVARIATES_Plot = 8
cca_splithalf_procrustes =
              ccatools::cca_splithalf(
              X_FIT  =      df0_imputed_group1[envvar],                    #
```

```
              Y_FIT  =      df0_imputed_group1[outcomes],                                  #
              X_PRED =      df0_imputed_group2[envvar],                                    #
              Y_PRED =      df0_imputed_group2[outcomes],                                  #
              ProcrustX =   cancor_Xweights[,1:NVARIATES],
              ProcrustY =   cancor_Yweights[,1:NVARIATES],
              ncomp  =      NVARIATES)                                                     #

Variates_Plot = cca_splithalf_procrustes$model_results$variates
  colnames(Variates_Plot) = c(paste0("ERF ",1:NVARIATES), paste0("Outcome ",1:NVARIATES))

# Correlation Plot of first 8 variates


Variates_Correlation_Test =
  Variates_Plot %>%
    dplyr::select(dplyr::matches("[ ][1-8]{1}$")) %>%
  ccatools::plotcor(., abs_colour = FALSE, reportCI = FALSE, includeN = FALSE)


Variates_Correlation_Test
```




```
save(cca_splithalf_procrustes, file=file.path("Output R Data", "cca_splithalf.Rdata"))
ggsave(file.path("Plots","Variates_Correlation_Test.pdf"), plot=Variates_Correlation_Test, device="pdf"
```

Plot Variate Scores

```
#
# Variates_Plot$id = 1:nrow(Variates_Plot)
# Variates_Plot_long = tidyr::pivot_longer(Variates_Plot, id)
library(patchwork)

Plot1=
Variates_Plot %>%
apply(.,2,scale) %>%
as.data.frame() %>%
ggplot(., aes(x=`ERF 1`, y=`Outcome 1`)) +
```

```
  geom_point(alpha=.4, shape=16) +
  jtools::theme_apa() + coord_fixed() + theme(aspect.ratio=1) +
  labs(x="1st CCA Risk Component",
       y="1st CCA Outcome Component")

Plot2=
Variates_Plot %>%
apply(.,2,scale) %>%
as.data.frame() %>%
ggplot(., aes(x=`ERF 2`, y=`Outcome 2`)) +
  geom_point(alpha=.4, shape=16) +
  jtools::theme_apa() + coord_fixed() + theme(aspect.ratio=1) +
  labs(x="2nd CCA Risk Component",
       y="2nd CCA Outcome Component")

Plot3=
Variates_Plot %>%
apply(.,2,scale) %>%
as.data.frame() %>%
ggplot(., aes(x=`ERF 3`, y=`Outcome 3`)) +
  geom_point(alpha=.4, shape=16) +
  jtools::theme_apa() + coord_fixed() + theme(aspect.ratio=1) +
  labs(x="3rd CCA Risk Component",
       y="3rd CCA Outcome Component")

plot_combined = Plot1 + Plot2 + Plot3

ggsave(file.path("Plots","CanonicalCorrelations.png"), plot=plot_combined, device="png", width=12, heigh

ggsave(file.path("Plots","CanonicalCorrelations1.png"), plot=Plot1, device="png", width=4.3, height=4.3
ggsave(file.path("Plots","CanonicalCorrelations2.png"), plot=Plot2, device="png", width=4.3, height=4.3
ggsave(file.path("Plots","CanonicalCorrelations3.png"), plot=Plot3, device="png", width=4.3, height=4.3
```

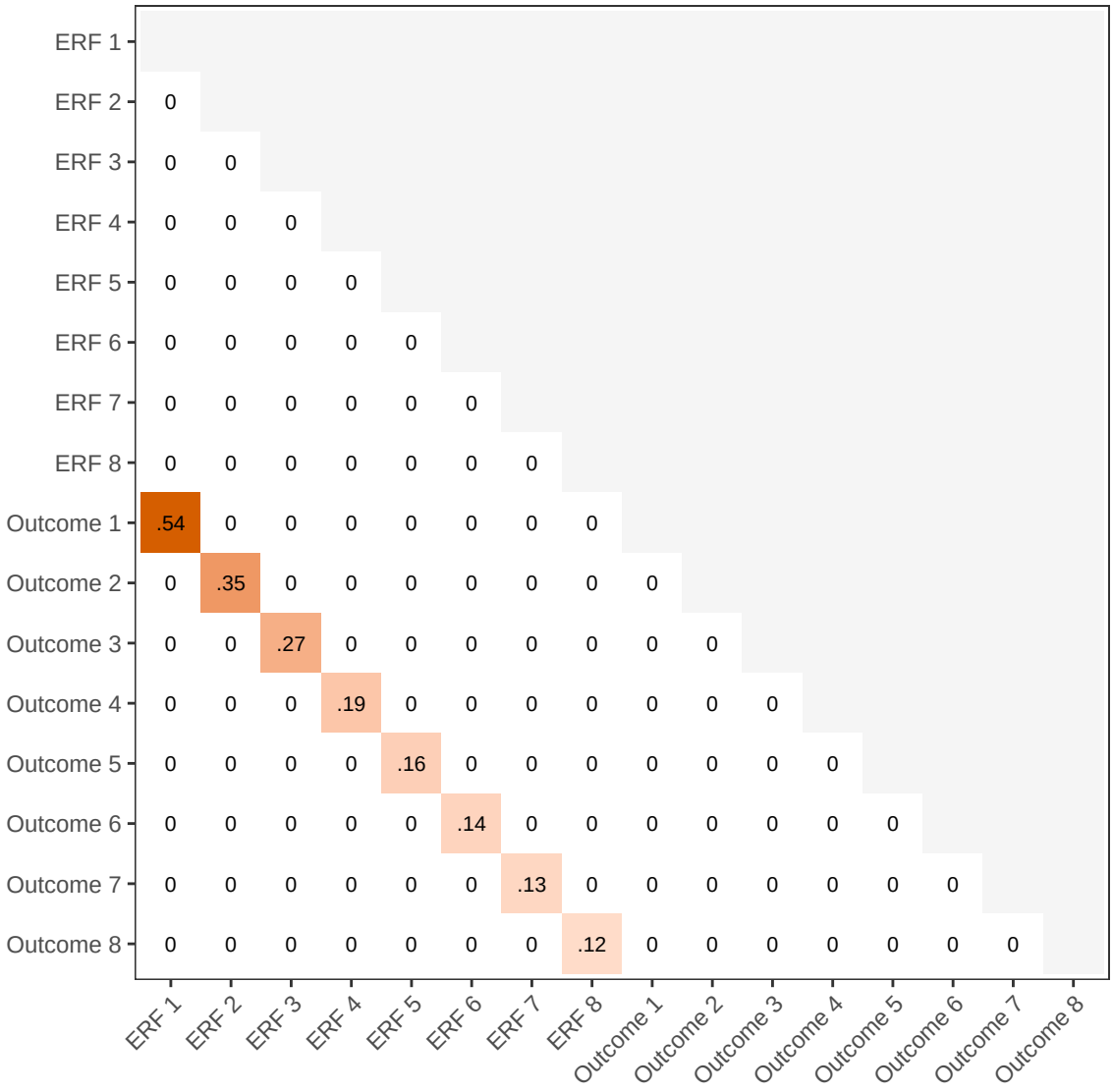
Latent Variable (Variate) score correlations within the training sample

I've plotted this here as it nicely illustrates how CCA works!

```
NVARIATES = 8

Variates_Plot = cca_group1$variates[,c(paste0("X",1:NVARIATES),paste0("Y",1:NVARIATES))]
  colnames(Variates_Plot) = c(paste0("ERF ",1:NVARIATES), paste0("Outcome ",1:NVARIATES))

Variates_Correlation_Train =

PlotCorrelationMatrix(Variates_Plot, includeN = FALSE, zerotext = "0", aligntext = FALSE) + ggplot2::the
        ggplot2::theme(axis.text.x = ggplot2::element_text(angle = 45, hjust = 1, vjust=1)) + ggplot2::s
        mid = "white", high = "#D55E00", midpoint = 0, na.value = "#F5F5F5", guide = "colourbar", aesthe
```

```
## Scale for 'fill' is already present. Adding another scale for 'fill', which
## will replace the existing scale.
```

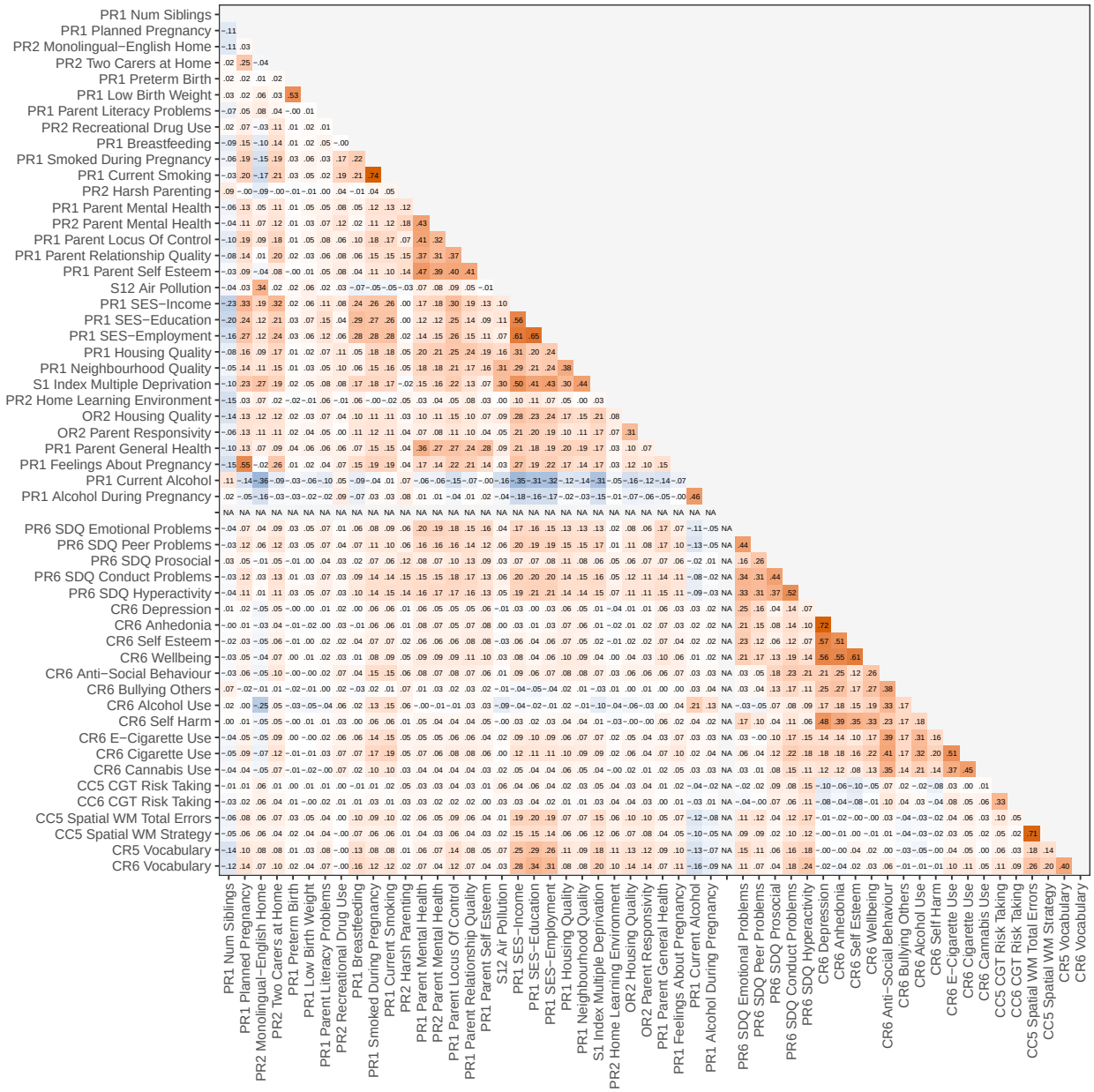
```
Variates_Correlation_Train
```




```
ggsave(file.path("Plots","Variates_Correlation_Train.pdf"), plot=Variates_Correlation_Train, device="pdf
```

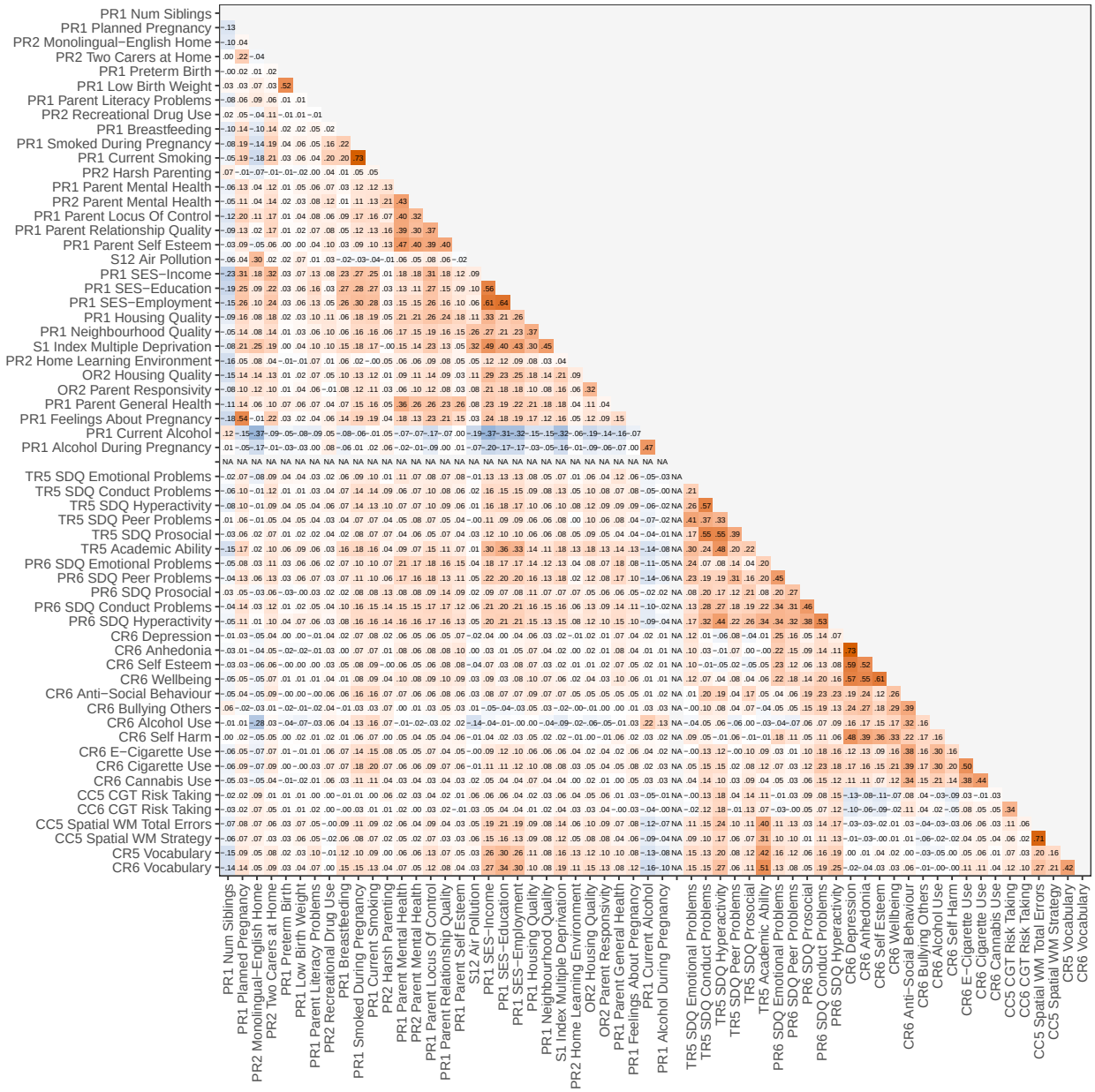
Correlation Matrix of all raw data

```
plot_data = cbind.data.frame(
                df0_imputed_BothGroups[envvar[envvar_Order]],
                rep(NA, nrow(df0_imputed_BothGroups)),
                df0_imputed_BothGroups[outcomes[Outcome_Order]])
colnames(plot_data) = c(envvar_labels[envvar_Order]," ",outcomes_labels[Outcome_Order])
gp = ccatools::plotcor(plot_data, abs_colour = FALSE, reportCI = FALSE, includeN = FALSE, textadjust =
gp
```




```
#Export File
ggsave(file.path("Plots","SupplementaryFigure1.pdf"), plot=gp, device="pdf", width=10, height=10)
ggsave(file.path("Plots","SupplementaryFigure1.png"), plot=gp, device="png", width=10, height=10, dpi=60

rm(gp,plot_data)

plot_data = cbind.data.frame(
               df0_imputed_Teach_BothGroups[envvar[envvar_Order]],
               rep(NA, nrow(df0_imputed_Teach_BothGroups)),
               df0_imputed_Teach_BothGroups[Outcome_Ordered2])
colnames(plot_data) = c(envvar_labels[envvar_Order],
                        " ",
                        outcomes_labels2[match(Outcome_Ordered2,outcomes2)])
gp = ccatools::plotcor(plot_data, abs_colour = FALSE, reportCI = FALSE, includeN = FALSE, textadjust =
```

```
gp
```

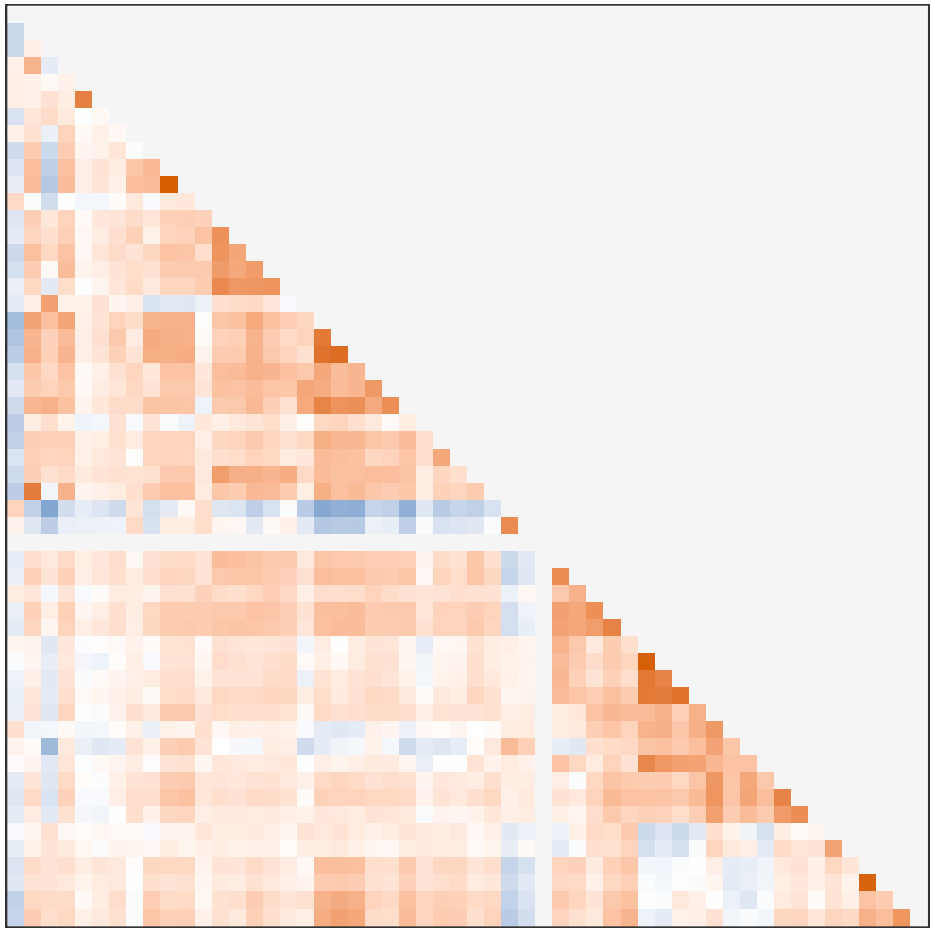



```
#Export File
ggsave(file.path("Plots","SupplementaryFigure2.pdf"), plot=gp, device="pdf", width=10, height=10)
ggsave(file.path("Plots","SupplementaryFigure2.png"), plot=gp, device="png", width=10, height=10, dpi=60

rm(gp, plot_data)
```

Correlation matrix with no numbers

```r
plot_data = cbind.data.frame(
                df0_imputed_BothGroups[envvar[envvar_Order]],
                rep(NA, nrow(df0_imputed_BothGroups)),
                df0_imputed_BothGroups[outcomes[Outcome_Order]])

# Function to transform data
sqwoosh = function(x, power = 2){
  is.neg = which(x<0)
  out = `^`(abs(x),power)
  out[is.neg] = out[is.neg]*-1
  return(out)
}

#Function which creates a correlation matrix plot of variables - without any labels
PlotCorrelationMatrix_Reduced = function(dat, Variables_Labels=NULL, textadjust=2){
      Variables = colnames(dat)
      if(is.null(Variables_Labels)){
        Variables_Labels = colnames(dat)
      }


      matrix_scores = dat
      Mat_Cor =gsub("^0","",gsub("^ +","",gsub("^-0","-", format(cor(matrix_scores, use="pairwise.compl
      Mat_Cor_fill = apply(cor(matrix_scores, use="pairwise.complete.obs"), 2,function(x) sqwoosh(x,.7))
      Mat_Cor_fill[lower.tri(Mat_Cor_fill,diag = TRUE)]=NA

      #Matrix on Ns per comparison - lower triag
      Mat_N = sapply(Variables, function(x)
        sapply(Variables, function(y)
          nrow(na.omit(data.frame(dat[,unique(c(x,y))]))))
        ))

      #Create Dataframe For Ggplot to Read
      PlotMat = Mat_Cor
      PlotMat[lower.tri(PlotMat, diag=TRUE)]=""
      colnames(PlotMat) = Variables_Labels ;  rownames(PlotMat) = Variables_Labels

      PlotMat = data.frame(reshape2::melt(PlotMat), stringsAsFactors = FALSE)
      head(PlotMat)

      PlotMat$value = (as.character(PlotMat$value))
      PlotMat$ValueFill = as.numeric(t(c(Mat_Cor_fill)))
      PlotMat$Var2 = factor(PlotMat$Var2, levels=rev(levels(PlotMat$Var2)))



      OutPlot =
        ggplot(data = PlotMat, aes(x=Var1, y=Var2,fill=ValueFill))+# + geom_point(aes(size=value^2,alph
        geom_tile() + labs(x=NULL, y=NULL) +
         jtools::theme_apa() +
        #scale_fill_brewer(palette=1,na.value="grey")+
        # scale_fill_continuous(na.value="white",low="#EDCB64", high="#B62A3D")+
        scale_fill_gradient2(low =  "#0072B2",
```

```
        mid = "white", high =  "#D55E00", midpoint = 0, na.value = "#F5F5F5",
        guide = "colourbar", aesthetics = "fill") +
        theme(legend.position = "#DAECED") +
        theme(axis.title=element_blank(),
                    axis.text=element_blank(),
                    axis.ticks=element_blank()) + coord_fixed()

      OutPlot

      return(OutPlot)
    }

# Plot Data
PlotCorrelationMatrix_Reduced(plot_data, textadjust = 1.3)
```



```
#Export File
ggsave(file.path("Plots","CorMat_BothGroups_NONUM.pdf"), device="pdf", width=5, height=5)
rm(plot_data)
```

Canonical Correlation point-estimates, CIs, p-values

```
format(cca_splithalf_procrustes$combined_cc, digits=0, nsmall=5)
```

```
##        cc1     cc2     cc3     cc4     cc5     cc6     cc7     cc8     cc9
## cc 0.51397 0.33248 0.24582 0.12658 0.11659 0.04246 0.06378 0.06104 0.04439
## LB 0.49372 0.30812 0.22015 0.09978 0.08974 0.01534 0.03671 0.03395 0.01728
## UB 0.53367 0.35640 0.27115 0.15319 0.14328 0.06952 0.09076 0.08803 0.07145
## p  0.00000 0.00000 0.00000 0.00000 0.00000 0.00216 0.00000 0.00001 0.00134
##        cc10     cc11     cc12     cc13     cc14     cc15     cc16     cc17
## cc -0.00925  0.00791 -0.00202  0.01395 -0.00516  0.00801  0.00716 -0.00051
## LB -0.03638 -0.01924 -0.02915 -0.01319 -0.03230 -0.01914 -0.01998 -0.02765
## UB  0.01789  0.03504  0.02512  0.04108  0.02198  0.03514  0.03430  0.02663
## p   0.50410  0.56800  0.88421  0.31374  0.70930  0.56321  0.60493  0.97078
##        cc18     cc19     cc20     cc21     cc22
## cc -0.00635 -0.01133 -0.02692  0.00136  0.00803
## LB -0.03349 -0.03846 -0.05402 -0.02578 -0.01911
## UB  0.02079  0.01581  0.00022  0.02850  0.03516
## p   0.64636  0.41328  0.05192  0.92159  0.56207
```

```
# NVARIATES = 15
#
# cca_splithalf_CI =
#                gb_CCA_splithalf(
#                X_FIT  = df0_imputed_group1[envvar],                                    # Da
#                Y_FIT  = df0_imputed_group1[outcomes],                                  # Da
#                ProcrustX = cancor_Xweights[,1:NVARIATES],                              # Ro
#                ProcrustY = cancor_Yweights[,1:NVARIATES],
#                X_PRED = df0_imputed_group2[envvar],                                    # Da
#                Y_PRED = df0_imputed_group2[outcomes],
#                ncomp=NVARIATES)
#
#
# # Internal check for giacomo - check that rotating the results JUST impacts the sign of the canonical
# cca_splithalf_CI_old =
#                gb_CCA_splithalf(
#                X_FIT  = df0_imputed_group1[envvar],                                    # Da
#                Y_FIT  = df0_imputed_group1[outcomes],                                  # Da
#                ProcrustX = NULL,                                    # Target matrices used to rota
#                ProcrustY = NULL,
#                X_PRED = df0_imputed_group2[envvar],                                    # Da
#                Y_PRED = df0_imputed_group2[outcomes],
#                ncomp=NVARIATES)
#



# save(cca_splithalf_CI, file=file.path("Output R Data", "cca_splithalf_CI.Rdata"))
# round(cca_splithalf_CI$combined_cc*100)/100
```

Analysis 2 - CCA loadings and bootstrap estimates of consistency

Run Model

```
set.seed(10)
timeA = base::Sys.time()

NVARIATES = 4
cca_boot = ccatools::coef_boot(
              X_FIT  = df0_imputed_group1[envvar],                    # Data
              Y_FIT  = df0_imputed_group1[outcomes],                  # Data
              ProcrustX = cancor_Xweights[,1:NVARIATES],             # Targe
              ProcrustY = cancor_Yweights[,1:NVARIATES],
              ncomp  = NVARIATES,
              Nboot  =  REPEATS_BOOT                                 # Number o
              )

save(cca_boot, file=file.path("Output R Data", "cca_boot.Rdata"))

timeB = base::Sys.time()

timeB-timeA
```

print CCA weights

CCA confidence intervals

```
cca_boot$ycoef_Quantiles
```

```
## [[1]]
##                                    X2.5.         X50.      X97.5.     original
## S6_Parent_SDQ_Emotion         0.0791953834  0.161445836  0.23315645  0.170836147
## S6_Parent_SDQ_Conduct         0.1729547754  0.253167971  0.33305008  0.247938224
## S6_Parent_SDQ_Hyper           0.1317472146  0.208680779  0.28389697  0.208496779
## S6_Parent_SDQ_Peer            0.2560803761  0.334643098  0.40724417  0.345994016
## S6_Parent_SDQ_Prosocial      -0.1269315667 -0.054808157  0.01930529 -0.063909889
## S6_MentalHealth_HateSelf     -0.3221510112 -0.229079839 -0.13351441 -0.229898257
## S6_MentalHealth_Tired         0.0004015874  0.091433129  0.17787987  0.094247710
## S6_SelfEsteem                -0.0003824570  0.082411846  0.16435522  0.085684295
## S6_Wellbeing                 -0.0117810975  0.076698844  0.16386347  0.077469865
## S6_RiskyBehaviours_Cig        0.0917962405  0.172640980  0.25574290  0.175395860
## S6_RiskyBehaviours_ECig       0.0548146174  0.127155562  0.20008534  0.124423270
## S6_RiskyBehaviours_Alcohol   -0.3187073739 -0.275295374 -0.23011877 -0.279910713
## S6_RiskyBehaviours_Cannabis  -0.1415600594 -0.073686980 -0.00280292 -0.068922965
## S6_ChildQ_SelfHarm           -0.0786341833 -0.008873290  0.05974562 -0.007862621
## S5_Cog_VS_Ability             0.2877362872  0.361587787  0.43465120  0.368264166
## S6_Cog_WordScoreIRT           0.4094581925  0.480089642  0.54499001  0.490172754
## S5_Cog_SWM_TotalErrors48      0.1612398656  0.248461948  0.33133964  0.254359046
## S5_Cog_SWM_Strategy          -0.0911328897 -0.008265321  0.07795808 -0.006917372
## S5_Cog_CGT_RiskTaking        -0.0549918147  0.009088661  0.07429858  0.003023441
## S6_Cog_CGT_RiskTaking        -0.0542168390  0.015138020  0.09170788  0.011688229
## S6_AntiSocialBehaviour        0.1477681395  0.223418208  0.29970840  0.233478175
```

```
## S6_Bullying                 -0.2526079523 -0.192021239 -0.12772395 -0.205660197
##
## [[2]]
##                                    X2.5.         X50.       X97.5.     original
## S6_Parent_SDQ_Emotion       -0.059536737  0.045078891  0.149783309  0.054585615
## S6_Parent_SDQ_Conduct       -0.190992463 -0.081921913  0.029820983 -0.082993292
## S6_Parent_SDQ_Hyper         -0.259615973 -0.155558181 -0.051331461 -0.156251881
## S6_Parent_SDQ_Peer          -0.158440174 -0.053849712  0.050281264 -0.053590036
## S6_Parent_SDQ_Prosocial     -0.125643851 -0.022134819  0.083297390 -0.021872201
## S6_MentalHealth_HateSelf    -0.137702630 -0.003557115  0.131367072 -0.001954563
## S6_MentalHealth_Tired       -0.004966187  0.125799855  0.249307902  0.131772191
## S6_SelfEsteem               -0.238489068 -0.117786189  0.005308454 -0.125613483
## S6_Wellbeing                -0.193788777 -0.073204002  0.046835916 -0.075071708
## S6_RiskyBehaviours_Cig      -0.289661039 -0.175731885 -0.059259295 -0.185322568
## S6_RiskyBehaviours_ECig     -0.110115188 -0.004489759  0.097638515 -0.007766819
## S6_RiskyBehaviours_Alcohol  -0.941484161 -0.878274649 -0.809942082 -0.902107057
## S6_RiskyBehaviours_Cannabis -0.109846394 -0.016493360  0.077029796 -0.020879497
## S6_ChildQ_SelfHarm          -0.168296105 -0.069003475  0.033913679 -0.070396928
## S5_Cog_VS_Ability           -0.042222419  0.059852859  0.163039848  0.059596819
## S6_Cog_WordScoreIRT         -0.008037468  0.086577762  0.178147020  0.088042547
## S5_Cog_SWM_TotalErrors48    -0.226754573 -0.110848360  0.009704123 -0.110858411
## S5_Cog_SWM_Strategy          0.004963851  0.122973182  0.235333085  0.121178620
## S5_Cog_CGT_RiskTaking       -0.035161225  0.056707784  0.147309032  0.062416549
## S6_Cog_CGT_RiskTaking        0.006780436  0.109687472  0.214343514  0.105697366
## S6_AntiSocialBehaviour      -0.139766102 -0.031842813  0.077019850 -0.042362130
## S6_Bullying                  0.024354737  0.114598220  0.203050970  0.121870106
##
## [[3]]
##                                   X2.5.         X50.      X97.5.     original
## S6_Parent_SDQ_Emotion        0.23893959  0.398822649  0.54323210  0.429115279
## S6_Parent_SDQ_Conduct       -0.01928395  0.140458291  0.29493580  0.153686491
## S6_Parent_SDQ_Hyper          0.08833895  0.244957430  0.39456891  0.263604945
## S6_Parent_SDQ_Peer          -0.05255922  0.100732106  0.25650082  0.116516156
## S6_Parent_SDQ_Prosocial      0.05649658  0.205763516  0.34854300  0.220840899
## S6_MentalHealth_HateSelf    -0.18667472 -0.005458822  0.18679488 -0.001486685
## S6_MentalHealth_Tired       -0.08530918  0.093312653  0.26569146  0.096726835
## S6_SelfEsteem               -0.27787181 -0.118698520  0.04460875 -0.124189441
## S6_Wellbeing                 0.01744288  0.194515034  0.35780978  0.205518635
## S6_RiskyBehaviours_Cig      -0.20515443 -0.032236150  0.13983764 -0.032284018
## S6_RiskyBehaviours_ECig     -0.15093316 -0.007158565  0.13597704 -0.007308379
## S6_RiskyBehaviours_Alcohol  -0.29997113 -0.221347732 -0.13651073 -0.240389925
## S6_RiskyBehaviours_Cannabis -0.18679285 -0.035977257  0.10898238 -0.039842981
## S6_ChildQ_SelfHarm          -0.19395422 -0.047701123  0.09667982 -0.051898699
## S5_Cog_VS_Ability           -0.40418855 -0.253077249 -0.10102304 -0.272416400
## S6_Cog_WordScoreIRT         -0.63542493 -0.510686431 -0.36924529 -0.546181526
## S5_Cog_SWM_TotalErrors48    -0.35351643 -0.190928530 -0.01183662 -0.201343716
## S5_Cog_SWM_Strategy         -0.14971730  0.027313083  0.19383587  0.024721541
## S5_Cog_CGT_RiskTaking        0.09398527  0.227448320  0.35348713  0.242701290
## S6_Cog_CGT_RiskTaking       -0.23087406 -0.081131005  0.07418486 -0.086478792
## S6_AntiSocialBehaviour      -0.16640064 -0.007342150  0.14875275 -0.007972262
## S6_Bullying                  0.08714090  0.207310814  0.32652972  0.223833997
##
## [[4]]
##                                   X2.5.        X50.     X97.5.    original
```

```
## S6_Parent_SDQ_Emotion       -0.51266858 -0.23476831 0.14311499 -0.29271611
## S6_Parent_SDQ_Conduct       -0.10234797  0.21349831 0.44944381  0.29083662
## S6_Parent_SDQ_Hyper         -0.20267847  0.07087669 0.32510141  0.11043283
## S6_Parent_SDQ_Peer          -0.54603818 -0.25965905 0.09179878 -0.32242961
## S6_Parent_SDQ_Prosocial     -0.02860861  0.29669787 0.53340049  0.38139162
## S6_MentalHealth_HateSelf    -0.25873941  0.02452172 0.30176654  0.02067216
## S6_MentalHealth_Tired       -0.41394367 -0.13589636 0.16127395 -0.17754757
## S6_SelfEsteem               -0.30561584 -0.05290440 0.22217827 -0.06262221
## S6_Wellbeing                -0.35690201 -0.08940026 0.20755545 -0.11569691
## S6_RiskyBehaviours_Cig      -0.36452337 -0.02333786 0.31679387 -0.02128605
## S6_RiskyBehaviours_ECig     -0.11213239  0.11724473 0.31901865  0.15777497
## S6_RiskyBehaviours_Alcohol  -0.09488338  0.09404139 0.25406779  0.10632332
## S6_RiskyBehaviours_Cannabis -0.37638540 -0.14379785 0.13411831 -0.18325692
## S6_ChildQ_SelfHarm          -0.27703230 -0.04186120 0.19014581 -0.05616743
## S5_Cog_VS_Ability           -0.26540499  0.02529446 0.28320308  0.05004041
## S6_Cog_WordScoreIRT         -0.26266023  0.09671199 0.31213874  0.12679043
## S5_Cog_SWM_TotalErrors48    -0.33575547 -0.06426789 0.19532867 -0.07606634
## S5_Cog_SWM_Strategy         -0.22194240  0.03203523 0.29257138  0.03897452
## S5_Cog_CGT_RiskTaking       -0.06329136  0.16397880 0.35625205  0.21466078
## S6_Cog_CGT_RiskTaking       -0.28461630  0.11761848 0.45056721  0.15461800
## S6_AntiSocialBehaviour      -0.56577393 -0.29889967 0.16486152 -0.35660506
## S6_Bullying                  0.06857640  0.37151724 0.54394869  0.47162596
```

```
xcoef_Quantiles = cca_boot$xcoef_Quantiles

# Shortening row names for easier printing!
for (i in 1:length(xcoef_Quantiles)){
  rownames(xcoef_Quantiles[[i]]) = substr(rownames(xcoef_Quantiles[[i]]),1,26)
}
xcoef_Quantiles
```

```
## [[1]]
##                                    X2.5.        X50.        X97.5.     original
## S1_SES_EducationAvg          0.572408763  0.66226846  0.7434995253  0.685068612
## S1_SES_AverageLastJob        0.100838235  0.21490894  0.3349750925  0.226015999
## S1_SES_EquivIncome_Norm      0.179349434  0.29508929  0.4119966756  0.304171782
## S2_NumberParents             0.001359759  0.10645273  0.2005536758  0.113841183
## S1_Fam_NSiblings            -0.176191051 -0.08994324  0.0008976994 -0.093875287
## S1_Health_PretermBirth      -0.105862738 -0.00554946  0.0939732900 -0.007643201
## S1_Health_LowBirthWeight     0.061608566  0.15892994  0.2545541556  0.162339659
## S1_Parent_GeneralHealth      0.026803690  0.11703987  0.2051017121  0.114949196
## S1_Health_SmokedThroughout -0.163632578 -0.03492102  0.0916146670 -0.023351440
## S1_Health_EverTriedBreastF  0.126324482  0.21328092  0.2997424822  0.223303981
## S1_Health_CurrentSmoking   -0.062413766  0.05999811  0.1794945443  0.057106218
## S1_Parent_AlcoholCurrent   -0.270099374 -0.17804594 -0.0837765699 -0.186931379
## S1_Parent_AlcoholPregnant  -0.168363664 -0.06918296  0.0247905436 -0.065627090
## S1_Fam_PlannedPregnancy    -0.070424593  0.03050880  0.1311252768  0.034779678
## S1_FeelingsPregnancy       -0.050603051  0.04779048  0.1486767323  0.049813212
## S12_AirPollution           -0.114693231 -0.02371565  0.0684664161 -0.024273328
## S1_PWB_RelationshipQuality  0.035298480  0.13098861  0.2253114054  0.137126346
## S1_PWB_SelfEsteem          -0.082600952  0.01482254  0.1144769272  0.014879270
## S1_PWB_ParentMentalHealth   0.061842456  0.16290559  0.2642764937  0.166198379
## S2_PWB_ParentMentalHealth  -0.020001341  0.07397793  0.1681258764  0.068812359
## S1_PWB_LocusOfControl      -0.009299600  0.08769237  0.1885407821  0.092486717
```

```
## S1_ParentNeighbourhoodRati -0.061658266  0.03649131  0.1327741522  0.036895747
## S1_IMDDeprivation           0.017498588  0.12141391  0.2201585059  0.130509473
## S1_ParentLiteracyProblems   0.005122818  0.08949481  0.1733926641  0.095292374
## S1_HousingQuality          -0.048811781  0.04867655  0.1417187216  0.050491064
## S2_HomeRating               0.093632694  0.18398466  0.2707727088  0.192943752
## S2_ParentRating             0.081039591  0.17172576  0.2615873656  0.183844506
## S2_HelpWithLearning         0.055324399  0.14327683  0.2255922854  0.155989095
## S2_Parent_RecreationalDrug -0.136904035 -0.04508643  0.0467497416 -0.042562774
## S2_HarshParenting           0.086300797  0.17704136  0.2617812869  0.192027637
## S2_Misc_EnglishMonolingual -0.036403474  0.05683273  0.1477819631  0.066878524
##
## [[2]]
##                                    X2.5.        X50.       X97.5.     original
## S1_SES_EducationAvg        -0.128044649 -0.01603668  0.096847059 -0.014891358
## S1_SES_AverageLastJob      -0.238174041 -0.10118753  0.038680164 -0.102975930
## S1_SES_EquivIncome_Norm    -0.246342670 -0.11570611  0.022721777 -0.120794396
## S2_NumberParents           -0.185019452 -0.06520482  0.049648051 -0.075336369
## S1_Fam_NSiblings           -0.036770414  0.06455472  0.165833650  0.074764403
## S1_Health_PretermBirth     -0.127642293 -0.01842802  0.090420170 -0.022310716
## S1_Health_LowBirthWeight    0.028210506  0.13528174  0.238058617  0.145515304
## S1_Parent_GeneralHealth    -0.212907848 -0.11067026 -0.009048049 -0.118111098
## S1_Health_SmokedThroughout -0.371295563 -0.22003773 -0.065017750 -0.238059046
## S1_Health_EverTriedBreastF -0.070259849  0.03690186  0.140864120  0.036349799
## S1_Health_CurrentSmoking   -0.426489008 -0.28357342 -0.130863996 -0.299338444
## S1_Parent_AlcoholCurrent   -0.567086427 -0.45714772 -0.337417397 -0.483123183
## S1_Parent_AlcoholPregnant  -0.199561704 -0.08181444  0.032213547 -0.090064518
## S1_Fam_PlannedPregnancy    -0.140642731 -0.02574543  0.089789683 -0.029554991
## S1_FeelingsPregnancy       -0.214227734 -0.09593788  0.023114018 -0.098638894
## S12_AirPollution           -0.115940507 -0.00394529  0.111167385 -0.001189445
## S1_PWB_RelationshipQuality -0.217616056 -0.10673041  0.004502304 -0.110438510
## S1_PWB_SelfEsteem          -0.086224631  0.03284933  0.149291552  0.034103750
## S1_PWB_ParentMentalHealth  -0.107927693  0.01093703  0.132999774  0.012372742
## S2_PWB_ParentMentalHealth  -0.094473242  0.01572369  0.125545279  0.015787774
## S1_PWB_LocusOfControl      -0.098955844  0.01848780  0.137026711  0.019671719
## S1_ParentNeighbourhoodRati -0.266909236 -0.15774735 -0.043694021 -0.166691441
## S1_IMDDeprivation           0.065743767  0.19288478  0.318560371  0.199642179
## S1_ParentLiteracyProblems  -0.074317621  0.01770722  0.110033585  0.019944122
## S1_HousingQuality          -0.174810373 -0.06061844  0.051238549 -0.065053034
## S2_HomeRating              -0.011486857  0.09483286  0.199252037  0.098606934
## S2_ParentRating            -0.047279234  0.05714650  0.163060914  0.063092219
## S2_HelpWithLearning        -0.009149618  0.08978540  0.188327145  0.096399240
## S2_Parent_RecreationalDrug -0.222635687 -0.12139946 -0.018374387 -0.128898690
## S2_HarshParenting          -0.248422316 -0.15962167 -0.065835788 -0.164636821
## S2_Misc_EnglishMonolingual  0.473364540  0.58160265  0.683085237  0.612237836
##
## [[3]]
##                                    X2.5.        X50.       X97.5.    original
## S1_SES_EducationAvg        -0.472340521 -0.32550839 -0.167586087 -0.35853172
## S1_SES_AverageLastJob      -0.402816140 -0.22789879 -0.039537288 -0.25349888
## S1_SES_EquivIncome_Norm    -0.241626966 -0.06452137  0.106967202 -0.07360900
## S2_NumberParents           -0.134538174  0.01988409  0.169054813  0.02533193
## S1_Fam_NSiblings            0.262685302  0.41623174  0.550868898  0.45723661
## S1_Health_PretermBirth     -0.213987290 -0.07972261  0.063472379 -0.08514105
## S1_Health_LowBirthWeight   -0.021636611  0.12377381  0.264890349  0.13567844
```

```
## S1_Parent_GeneralHealth     0.098004889  0.23384831  0.368121111  0.26061068
## S1_Health_SmokedThroughout -0.396176582 -0.21597667 -0.012946940 -0.22567024
## S1_Health_EverTriedBreastF -0.276033690 -0.13882543 -0.003134135 -0.15326171
## S1_Health_CurrentSmoking   -0.092107656  0.09286268  0.261838876  0.09627971
## S1_Parent_AlcoholCurrent   -0.213376920 -0.07638778  0.064206622 -0.08460886
## S1_Parent_AlcoholPregnant  -0.122592707  0.02117097  0.165224509  0.02707102
## S1_Fam_PlannedPregnancy    -0.229716889 -0.08116786  0.071342899 -0.08948564
## S1_FeelingsPregnancy       -0.053412714  0.09600258  0.243792772  0.10523076
## S12_AirPollution           -0.018213200  0.13578647  0.280783886  0.14852235
## S1_PWB_RelationshipQuality  0.003433363  0.15073569  0.291429086  0.16577377
## S1_PWB_SelfEsteem          -0.053022055  0.09815936  0.253970438  0.11085604
## S1_PWB_ParentMentalHealth  -0.060370545  0.09832842  0.262424433  0.11454977
## S2_PWB_ParentMentalHealth   0.161686042  0.31037542  0.450520115  0.34452367
## S1_PWB_LocusOfControl      -0.042961490  0.10462509  0.250958285  0.11520058
## S1_ParentNeighbourhoodRati  0.015577275  0.16150379  0.302151366  0.17892016
## S1_IMDDeprivation          -0.123265103  0.03716431  0.195861086  0.04300445
## S1_ParentLiteracyProblems   0.009033838  0.13282777  0.250752142  0.14660308
## S1_HousingQuality          -0.036459803  0.12091716  0.271479050  0.13483196
## S2_HomeRating              -0.059835587  0.07353337  0.207895002  0.08311946
## S2_ParentRating            -0.057120804  0.07731422  0.211022985  0.08784169
## S2_HelpWithLearning        -0.150787684 -0.01687480  0.116665684 -0.01769417
## S2_Parent_RecreationalDrug -0.177117818 -0.04799657  0.078289353 -0.05321347
## S2_HarshParenting           0.133140196  0.26224559  0.385331996  0.29446029
## S2_Misc_EnglishMonolingual -0.101062343  0.04433377  0.195215718  0.05389745
##
## [[4]]
##                                  X2.5.         X50.     X97.5.       original
## S1_SES_EducationAvg        -0.10787674  0.183068940 0.41469831  2.143918e-01
## S1_SES_AverageLastJob      -0.24426093  0.021623339 0.28537457  2.808576e-02
## S1_SES_EquivIncome_Norm    -0.31715635 -0.097541278 0.14079511 -1.408955e-01
## S2_NumberParents           -0.37712151 -0.075942446 0.28473922 -9.202954e-02
## S1_Fam_NSiblings           -0.15461097  0.271165161 0.53129955  3.368965e-01
## S1_Health_PretermBirth     -0.28463629 -0.113789221 0.07854343 -1.500778e-01
## S1_Health_LowBirthWeight   -0.19443350  0.006618713 0.21556275  9.735256e-05
## S1_Parent_GeneralHealth    -0.40178848 -0.228248292 0.01062629 -3.021013e-01
## S1_Health_SmokedThroughout -0.26923411  0.170807849 0.47111428  2.056646e-01
## S1_Health_EverTriedBreastF -0.14740077  0.026741386 0.20176505  2.875666e-02
## S1_Health_CurrentSmoking   -0.41198261 -0.170308145 0.11615288 -2.221749e-01
## S1_Parent_AlcoholCurrent   -0.20967189 -0.026712925 0.16056694 -2.127168e-02
## S1_Parent_AlcoholPregnant  -0.15260309  0.119923150 0.33998677  1.548531e-01
## S1_Fam_PlannedPregnancy    -0.21587581  0.004542533 0.21944587  6.347353e-03
## S1_FeelingsPregnancy       -0.13374985  0.061562525 0.24807656  7.579588e-02
## S12_AirPollution           -0.23106898 -0.003887192 0.23358438 -4.171928e-03
## S1_PWB_RelationshipQuality -0.13996548  0.102331297 0.31138318  1.264980e-01
## S1_PWB_SelfEsteem          -0.29090842 -0.093284204 0.12557920 -1.249178e-01
## S1_PWB_ParentMentalHealth  -0.29457981 -0.040656032 0.22421429 -5.485501e-02
## S2_PWB_ParentMentalHealth  -0.39052746 -0.193081193 0.07109354 -2.543603e-01
## S1_PWB_LocusOfControl      -0.13307705  0.063663364 0.25484513  8.295647e-02
## S1_ParentNeighbourhoodRati -0.24676789 -0.055655069 0.14599928 -8.010655e-02
## S1_IMDDeprivation          -0.28317829 -0.054604410 0.19842419 -7.767574e-02
## S1_ParentLiteracyProblems  -0.05266072  0.123305143 0.28763496  1.584552e-01
## S1_HousingQuality          -0.25763100 -0.018332981 0.21917235 -2.759235e-02
## S2_HomeRating              -0.19701556 -0.013697462 0.19272512 -2.145202e-02
## S2_ParentRating            -0.04447793  0.171104806 0.35092805  2.148708e-01
```

```
## S2_HelpWithLearning        -0.01263580  0.263411926 0.45683566  3.308621e-01
## S2_Parent_RecreationalDrug -0.17190442  0.031770468 0.20958284  4.095100e-02
## S2_HarshParenting           0.06288279  0.406156965 0.57845124  5.156473e-01
## S2_Misc_EnglishMonolingual -0.20038748  0.068971399 0.31348415  8.205245e-02
```

CCA Weights (training dataset)

```
XWeights = cca_group1$xcoef
rownames(XWeights) = substr(envvar_labels,1,26)
colnames(XWeights) = paste0("LV ",1:ncol(XWeights))

XWeights[,1:4]
```

```
##                                  LV 1         LV 2        LV 3          LV 4
## PR1 SES-Education           0.685068612 -0.014891358 -0.35853172  2.143918e-01
## PR1 SES-Employment          0.226015999 -0.102975930 -0.25349888  2.808576e-02
## PR1 SES-Income              0.304171782 -0.120794396 -0.07360900 -1.408955e-01
## PR2 Two Carers at Home      0.113841183 -0.075336369  0.02533193 -9.202954e-02
## PR1 Num Siblings           -0.093875287  0.074764403  0.45723661  3.368965e-01
## PR1 Preterm Birth          -0.007643201 -0.022310716 -0.08514105 -1.500778e-01
## PR1 Low Birth Weight        0.162339659  0.145515304  0.13567844  9.735256e-05
## PR1 Parent General Health   0.114949196 -0.118111098  0.26061068 -3.021013e-01
## PR1 Smoked During Pregnanc -0.023351440 -0.238059046 -0.22567024  2.056646e-01
## PR1 Breastfeeding           0.223303981  0.036349799 -0.15326171  2.875666e-02
## PR1 Current Smoking         0.057106218 -0.299338444  0.09627971 -2.221749e-01
## PR1 Current Alcohol        -0.186931379 -0.483123183 -0.08460886 -2.127168e-02
## PR1 Alcohol During Pregnan -0.065627090 -0.090064518  0.02707102  1.548531e-01
## PR1 Planned Pregnancy       0.034779678 -0.029554991 -0.08948564  6.347353e-03
## PR1 Feelings About Pregnan  0.049813212 -0.098638894  0.10523076  7.579588e-02
## S12 Air Pollution          -0.024273328 -0.001189445  0.14852235 -4.171928e-03
## PR1 Parent Relationship Qu  0.137126346 -0.110438510  0.16577377  1.264980e-01
## PR1 Parent Self Esteem      0.014879270  0.034103750  0.11085604 -1.249178e-01
## PR1 Parent Mental Health    0.166198379  0.012372742  0.11454977 -5.485501e-02
## PR2 Parent Mental Health    0.068812359  0.015787774  0.34452367 -2.543603e-01
## PR1 Parent Locus Of Contro  0.092486717  0.019671719  0.11520058  8.295647e-02
## PR1 Neighbourhood Quality   0.036895747 -0.166691441  0.17892016 -8.010655e-02
## S1 Index Multiple Deprivat  0.130509473  0.199642179  0.04300445 -7.767574e-02
## PR1 Parent Literacy Proble  0.095292374  0.019944122  0.14660308  1.584552e-01
## PR1 Housing Quality         0.050491064 -0.065053034  0.13483196 -2.759235e-02
## OR2 Housing Quality         0.192943752  0.098606934  0.08311946 -2.145202e-02
## OR2 Parent Responsivity     0.183844506  0.063092219  0.08784169  2.148708e-01
## PR2 Home Learning Environm  0.155989095  0.096399240 -0.01769417  3.308621e-01
## PR2 Recreational Drug Use  -0.042562774 -0.128898690 -0.05321347  4.095100e-02
## PR2 Harsh Parenting         0.192027637 -0.164636821  0.29446029  5.156473e-01
## PR2 Monolingual-English Ho  0.066878524  0.612237836  0.05389745  8.205245e-02
```

CCA Weights (testing dataset)

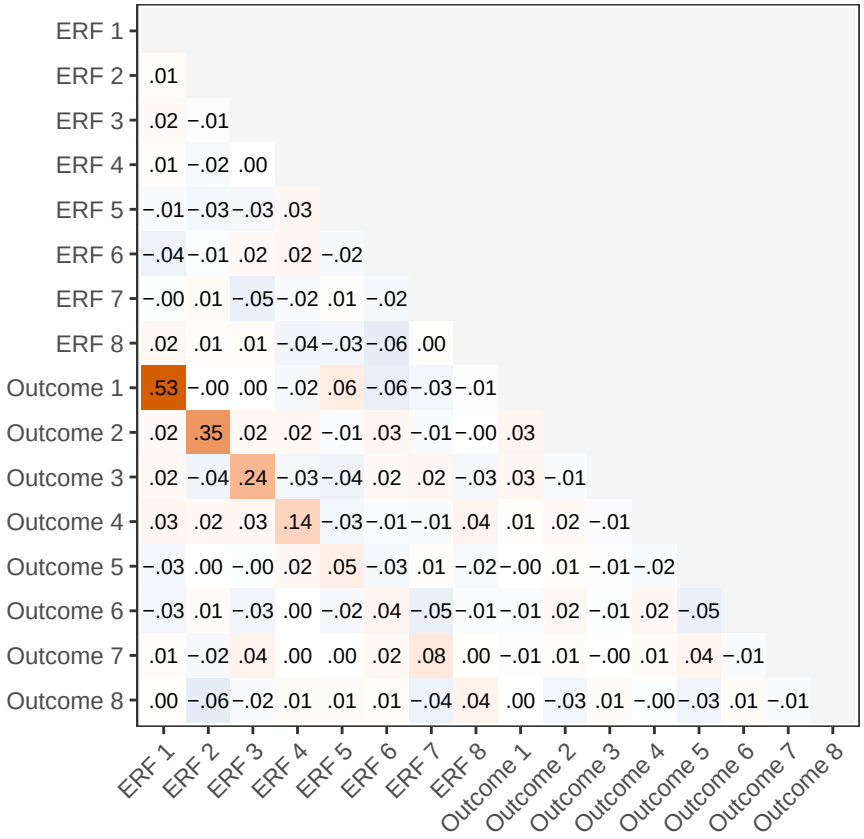
```
YWeights = cca_group1$ycoef
rownames(YWeights) = outcomes_labels
colnames(YWeights) = paste0("LV ",1:ncol(YWeights))
YWeights[,1:4]
```

```
##                                  LV 1         LV 2         LV 3        LV 4
## PR6 SDQ Emotional Problems   0.170836147  0.054585615  0.429115279 -0.29271611
## PR6 SDQ Conduct Problems     0.247938224 -0.082993292  0.153686491  0.29083662
## PR6 SDQ Hyperactivity        0.208496779 -0.156251881  0.263604945  0.11043283
## PR6 SDQ Peer Problems        0.345994016 -0.053590036  0.116516156 -0.32242961
## PR6 SDQ Prosocial           -0.063909889 -0.021872201  0.220840899  0.38139162
## CR6 Depression              -0.229898257 -0.001954563 -0.001486685  0.02067216
## CR6 Anhedonia                0.094247710  0.131772191  0.096726835 -0.17754757
## CR6 Self Esteem              0.085684295 -0.125613483 -0.124189441 -0.06262221
## CR6 Wellbeing                0.077469865 -0.075071708  0.205518635 -0.11569691
## CR6 Cigarette Use            0.175395860 -0.185322568 -0.032284018 -0.02128605
## CR6 E-Cigarette Use          0.124423270 -0.007766819 -0.007308379  0.15777497
## CR6 Alcohol Use             -0.279910713 -0.902107057 -0.240389925  0.10632332
## CR6 Cannabis Use            -0.068922965 -0.020879497 -0.039842981 -0.18325692
## CR6 Self Harm               -0.007862621 -0.070396928 -0.051898699 -0.05616743
## CR5 Vocabulary               0.368264166  0.059596819 -0.272416400  0.05004041
## CR6 Vocabulary               0.490172754  0.088042547 -0.546181526  0.12679043
## CC5 Spatial WM Total Errors  0.254359046 -0.110858411 -0.201343716 -0.07606634
## CC5 Spatial WM Strategy     -0.006917372  0.121178620  0.024721541  0.03897452
## CC5 CGT Risk Taking          0.003023441  0.062416549  0.242701290  0.21466078
## CC6 CGT Risk Taking          0.011688229  0.105697366 -0.086478792  0.15461800
## CR6 Anti-Social Behaviour    0.233478175 -0.042362130 -0.007972262 -0.35660506
## CR6 Bullying Others         -0.205660197  0.121870106  0.223833997  0.47162596
```

Run same analyses on teacher dataset

```
NVARIATES = 22


cca_cancor_teach = stats::cancor(x=scale(df0_imputed_Teach_group1[,envvar]),
                                 y=scale(df0_imputed_Teach_group1[,outcomes2])
                                 )

  cancor_Xweights_teach = apply(cca_cancor_teach$xcoef,2, function(x) x/sqrt(sum(x^2)))
  cancor_Yweights_teach = apply(cca_cancor_teach$ycoef,2, function(x) x/sqrt(sum(x^2)))

# Run Split Half
cca_splithalf_teacher = ccatools::cca_splithalf(
              X_FIT  =      df0_imputed_Teach_group1[,envvar],
              Y_FIT  =      df0_imputed_Teach_group1[,outcomes2],
              X_PRED =      df0_imputed_Teach_group2[,envvar],
              Y_PRED =      df0_imputed_Teach_group2[,outcomes2],
              ProcrustX = cancor_Xweights_teach[,1:NVARIATES],                    #
              ProcrustY = cancor_Yweights_teach[,1:NVARIATES],
              ncomp  =      NVARIATES)                                            #


Variates_Plot =
cca_splithalf_teacher$model_results$variates %>%
    dplyr::select(dplyr::matches("[X|Y][1-8]{1}$")) %>%
    `colnames<-` (c(paste0("ERF ",1:8), paste0("Outcome ",1:8))) %>%
  ccatools::plotcor(., abs_colour = FALSE, reportCI = FALSE, includeN = FALSE)
```

```
Variates_Plot
```




```
save(cca_splithalf_teacher, file=file.path("Output R Data", "cca_splithalf_teacher.Rdata"))
```

CCA Weights (teacher dataset)

```
XWeights = cca_splithalf_teacher$model_results$xcoef
rownames(XWeights) = substr(envvar_labels,1,26)
colnames(XWeights) = paste0("LV ",1:ncol(XWeights))
XWeights[,1:4]
```

```
##                                  LV 1         LV 2         LV 3        LV 4
## PR1 SES-Education           0.672627524  0.008093837  0.492141241 -0.08947625
## PR1 SES-Employment          0.205304910  0.074105775  0.173554907 -0.07376188
## PR1 SES-Income              0.372275169  0.037349417 -0.018181958  0.10729922
## PR2 Two Carers at Home      0.016099415 -0.059961549 -0.156168787  0.13561997
## PR1 Num Siblings           -0.178342357 -0.123946245 -0.450683794 -0.07657572
## PR1 Preterm Birth          -0.008020715  0.011819193  0.106143667  0.10997840
## PR1 Low Birth Weight        0.211466024 -0.161010339 -0.055308494  0.02511204
## PR1 Parent General Health   0.124788394  0.071994458 -0.167997997  0.33862799
## PR1 Smoked During Pregnanc  0.055759851  0.171925156 -0.147421496 -0.21434501
```

```
## PR1 Breastfeeding           0.140746339  0.014714906  0.125355674  0.08135814
## PR1 Current Smoking         0.135223245  0.360090548  0.006893521  0.05093574
## PR1 Current Alcohol        -0.085594136  0.380561075  0.075572404 -0.11388637
## PR1 Alcohol During Pregnan -0.067019979  0.148908478 -0.073519914 -0.20905108
## PR1 Planned Pregnancy       0.127050080  0.087360499 -0.146185782  0.10826224
## PR1 Feelings About Pregnan  0.030065046  0.069561182  0.021795510 -0.04059866
## S12 Air Pollution          -0.128936325 -0.219785973 -0.036297016  0.01183216
## PR1 Parent Relationship Qu  0.071119261  0.001074683 -0.131901542 -0.15919739
## PR1 Parent Self Esteem      0.073184323 -0.017952428 -0.170292080  0.08159802
## PR1 Parent Mental Health    0.079224188 -0.068468461 -0.036125409  0.24325341
## PR2 Parent Mental Health    0.046930437 -0.123367352 -0.176711563  0.29455830
## PR1 Parent Locus Of Contro  0.041538902 -0.058873965 -0.139104597 -0.11128325
## PR1 Neighbourhood Quality  -0.023718406 -0.008531137 -0.057828061 -0.05437604
## S1 Index Multiple Deprivat  0.103592169 -0.069264057 -0.204295425  0.04917687
## PR1 Parent Literacy Proble  0.069335314  0.012133460  0.019257619 -0.03177670
## PR1 Housing Quality         0.142952504  0.053317690 -0.116220447 -0.01212424
## OR2 Housing Quality         0.162832393 -0.091451464 -0.056297848 -0.09167278
## OR2 Parent Responsivity     0.129681790 -0.046190684 -0.044462503 -0.30607992
## PR2 Home Learning Environm  0.207831704 -0.039575443  0.099814163 -0.31310127
## PR2 Recreational Drug Use  -0.089623060  0.152868617 -0.016281016 -0.10586126
## PR2 Harsh Parenting         0.173501917  0.171896556 -0.459630774 -0.43420387
## PR2 Monolingual-English Ho -0.076847304 -0.677560257 -0.040556768 -0.31833055
```

```
YWeights = cca_splithalf_teacher$model_results$ycoef
rownames(YWeights) = substr(outcomes_labels2,1,26)
colnames(YWeights) = paste0("LV ",1:ncol(YWeights))
YWeights[,1:4]
```

```
##                                LV 1         LV 2         LV 3         LV 4
## TR5 SDQ Emotional Problems  0.048676004  0.157618301 -0.164550666  0.430867076
## TR5 SDQ Conduct Problems    0.193927459  0.137721311 -0.137709209 -0.055301827
## TR5 SDQ Hyperactivity      -0.099802326  0.097504256 -0.091538538 -0.070345147
## TR5 SDQ Peer Problems      -0.049515439  0.030378217  0.005860666  0.020486485
## TR5 SDQ Prosocial           0.012414111 -0.146641976  0.045900844 -0.036372237
## TR5 Academic Ability        0.449274283  0.158243943  0.353128684 -0.196902067
## PR6 SDQ Emotional Problems  0.145832198 -0.091008214 -0.267953567  0.279221883
## PR6 SDQ Conduct Problems    0.233802876 -0.016412393 -0.401467242 -0.079452162
## PR6 SDQ Hyperactivity       0.185927729 -0.015822824 -0.118254730 -0.021735139
## PR6 SDQ Peer Problems       0.314076834 -0.155777795 -0.235587372  0.350322199
## PR6 SDQ Prosocial          -0.079478384  0.079461469 -0.236445176 -0.387260040
## CR6 Depression             -0.173029384  0.021038625 -0.140535181 -0.084180502
## CR6 Anhedonia               0.053952231 -0.118133997  0.099133422  0.203427633
## CR6 Self Esteem             0.085807383  0.105795127 -0.032439348 -0.069606168
## CR6 Wellbeing               0.032270069  0.008702615 -0.016387699  0.058158573
## CR6 Cigarette Use           0.131181501  0.147229129 -0.232425065 -0.095689114
## CR6 E-Cigarette Use         0.151970235  0.003509431  0.086589846 -0.168617293
## CR6 Alcohol Use            -0.134895547  0.872638887  0.082262181  0.163575281
## CR6 Cannabis Use           -0.035131463 -0.003068358  0.108187081 -0.042823280
## CR6 Self Harm               0.004713132  0.031204656  0.042865851  0.099128036
## CR5 Vocabulary              0.336947406  0.027345822  0.379815964  0.061701557
## CR6 Vocabulary              0.365274159 -0.119727196  0.340860496 -0.004907661
## CC5 Spatial WM Total Error  0.218582960 -0.033515971 -0.035342909  0.118088405
## CC5 Spatial WM Strategy    -0.043214286 -0.111819685  0.165891830 -0.216502873
## CC5 CGT Risk Taking        -0.016381518 -0.073605455 -0.165366182 -0.286600887
```

```
## CC6 CGT Risk Taking        -0.009198998 -0.107931995 -0.042989502 -0.099239842
## CR6 Anti-Social Behaviour   0.262414512 -0.022494598  0.040220509  0.043458034
## CR6 Bullying Others        -0.251630381 -0.060924159 -0.166187250 -0.342285335
```

Split half cca results (teacher dataset)

```
format(cca_splithalf_teacher$combined_cc, digits=0, nsmall=5)
```

```
##        cc1     cc2     cc3     cc4     cc5      cc6     cc7     cc8     cc9
## cc 0.53144 0.34537 0.23818 0.13874 0.05236  0.03602 0.07505 0.03985 0.04361
## LB 0.50449 0.31253 0.20314 0.10244 0.01557 -0.00080 0.03834 0.00304 0.00680
## UB 0.55735 0.37739 0.27261 0.17466 0.08901  0.07274 0.11157 0.07656 0.08030
## p  0.00000 0.00000 0.00000 0.00000 0.00530  0.05521 0.00006 0.03389 0.02026
##        cc10     cc11     cc12     cc13     cc14     cc15     cc16     cc17
## cc  0.02132  0.01834 -0.01835 -0.02119  0.00512  0.02829  0.01848  0.02542
## LB -0.01551 -0.01849 -0.05514 -0.05796 -0.03171 -0.00854 -0.01835 -0.01141
## UB  0.05810  0.05513  0.01848  0.01565  0.04193  0.06504  0.05526  0.06219
## p   0.25651  0.32896  0.32870  0.25956  0.78532  0.13220  0.32540  0.17602
##        cc18     cc19     cc20     cc21     cc22
## cc  0.01282  0.00807 -0.02180  0.03093  0.01472
## LB -0.02401 -0.02876 -0.05857 -0.00589 -0.02211
## UB  0.04962  0.04487  0.01504  0.06768  0.05151
## p   0.49495  0.66777  0.24607  0.09967  0.43334
```

Bootstrap resampling on teaching data

```
set.seed(101)
timeA = base::Sys.time()

NVARIATES = 4                                                                    #This is
cca_boot_teach = ccatools::coef_boot(
              X_FIT  = df0_imputed_Teach_BothGroups[envvar],
              Y_FIT  = df0_imputed_Teach_BothGroups[outcomes2],
              ProcrustX = cancor_Xweights_teach[,1:NVARIATES],                          #
              ProcrustY = cancor_Yweights_teach[,1:NVARIATES],
              ncomp  = NVARIATES,
              Nboot  =  REPEATS_BOOT                                             # Number o
              )

save(cca_boot_teach, file=file.path("Output R Data", "cca_boot_teach.Rdata"))


timeB = base::Sys.time()

timeB-timeA
```