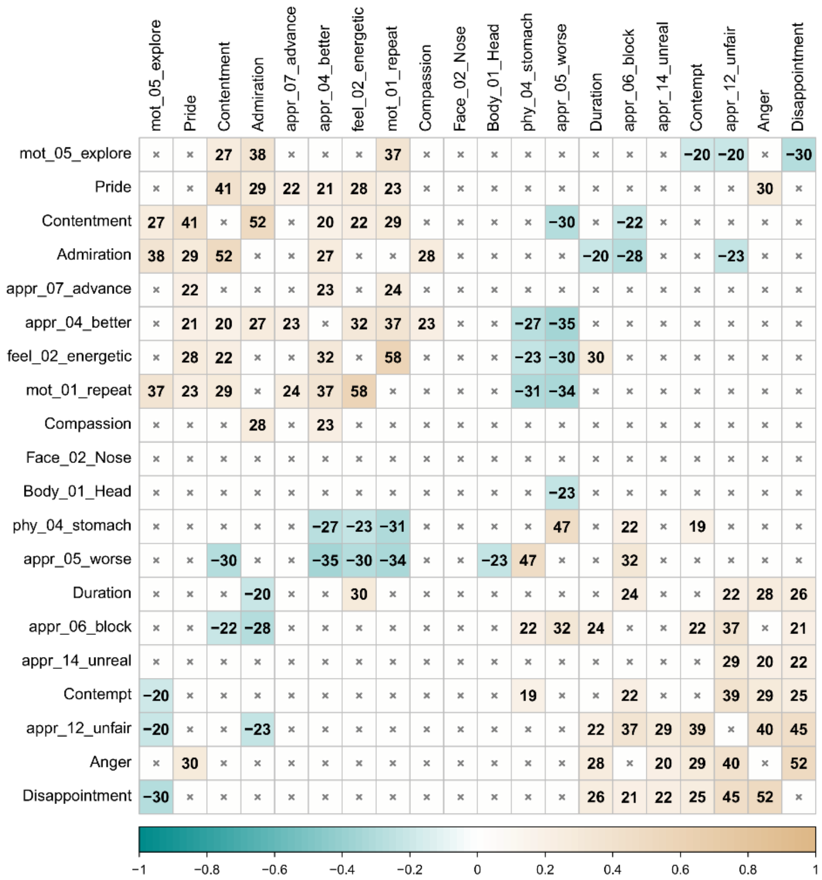
STATISTICAL SUPPORT
SWISS CENTER FOR AFFECTIVE SCIENCES

Correlation plots

E-mail distributed on 23-08-2022

Dear all,

For this week's support e-mail, I would like to promote the use of correlation plots for data exploration, specifically with the R package `corrplot`. The package vignettes are an excellent reference with many variations to copy-paste. However, for inexperienced users, you may have trouble adapting `corrplot` to data sets with categorical variables, missing values, and/or 0 variances.



	mot_05_explore	Pride	Contentment	Admiration	appr_07_advance	appr_04_better	feel_02_energetic	mot_01_repeat	Compassion	Face_02_Nose	Body_01_Head	phy_04_stomach	appr_05_worse	Duration	appr_06_block	appr_14_unreal	Contempt	appr_12_unfair	Anger	Disappointment
mot_05_explore	×	×	27	38	×	×	×	37	×	×	×	×	×	×	×	×	−20	−20	×	−30
Pride	×	×	41	29	22	21	28	23	×	×	×	×	×	×	×	×	×	×	30	×
Contentment	27	41	×	52	×	20	22	29	×	×	×	×	−30	×	−22	×	×	×	×	×
Admiration	38	29	52	×	×	27	×	×	28	×	×	×	×	−20	−28	×	×	−23	×	×
appr_07_advance	×	22	×	×	×	23	×	24	×	×	×	×	×	×	×	×	×	×	×	×
appr_04_better	×	21	20	27	23	×	32	37	23	×	×	−27	−35	×	×	×	×	×	×	×
feel_02_energetic	×	28	22	×	×	32	×	58	×	×	×	−23	−30	30	×	×	×	×	×	×
mot_01_repeat	37	23	29	×	24	37	58	×	×	×	×	−31	−34	×	×	×	×	×	×	×
Compassion	×	×	×	28	×	23	×	×	×	×	×	×	×	×	×	×	×	×	×	×
Face_02_Nose	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
Body_01_Head	×	×	×	×	×	×	×	×	×	×	×	×	−23	×	×	×	×	×	×	×
phy_04_stomach	×	×	×	×	×	−27	−23	−31	×	×	×	×	47	×	22	×	19	×	×	×
appr_05_worse	×	×	−30	×	×	−35	−30	−34	×	×	−23	47	×	×	32	×	×	×	×	×
Duration	×	×	×	−20	×	×	30	×	×	×	×	×	×	×	24	×	×	22	28	26
appr_06_block	×	×	−22	−28	×	×	×	×	×	×	×	22	32	24	×	×	22	37	×	21
appr_14_unreal	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	29	20	22
Contempt	−20	×	×	×	×	×	×	×	×	×	×	19	×	×	22	×	×	39	29	25
appr_12_unfair	−20	×	×	−23	×	×	×	×	×	×	×	×	×	22	37	29	39	×	40	45
Anger	×	30	×	×	×	×	×	×	×	×	×	×	×	28	×	20	29	40	×	52
Disappointment	−30	×	×	×	×	×	×	×	×	×	×	×	×	26	21	22	25	45	52	×

For this reason I wrote a wrapper script (`cpplus.r`) which handles such data automatically for quick plotting, and which also includes some helpful visual defaults (see attached example). Colors and numbers are overlaid, non-significant correlations are suppressed (at a specified alpha level), and variables can be reordered optionally with hierarchical clustering to reveal interesting variable groupings. The color scheme preserves contrast for color blindness.

The function can be run as follows:

```
library(corrplot)
library(Hmisc)
library(psych)
source("C:/myfilepath/.../cpplus.r")[1]
cpplus(mydata, reorder=TRUE, alpha=0.005, method="pearson")
```

With mydata one of your data sets that you have loaded in R (typically a data frame object). The default in `cpplus` is for reordering to be switched on, and for a significance level of 0.005. Of course you can always go back to the original `corrplot` function if specific tweaking is required. In addition, `cpplus` supports three different types of correlations with a `method` argument: Pearson (`"pearson"`), Spearman (`"spearman"`), and distance correlation (`"dcor"`), the latter which is a non-linear measure of association between 0 and 1. For Pearson and Spearman correlation, there is also an option to plot partial correlations instead, using the `partial=TRUE` argument.

Correlation mosaics like these can be enormously helpful to make sense of association patterns. As well, they can be adapted to visualize regression coefficients, causal models, and confusion matrices based on MCC coefficients.

Best,
Ben

--
Ben Meuleman, Ph.D.
Statistician
Swiss Center for Affective Sciences
University of Geneva | Campus Biotech
Chemin des Mines 9 | CH-1202 Genève
ben.meuleman@unige.ch | +41 (0)22 379 09 79

[1] Alternatively, copy the code at the bottom of this document and run it directly in R

```
###################################
## CORRPLOT PLUS FUNCTION
###################################

library(psych)
library(corrplot)
library(Hmisc)
library(energy)
library(foreach)
library(ppcor)
dummify <- function(D) {
as.data.frame(lapply(data.frame(D),dummy.code)) }
mdcor <- function(V) {
      nms <- colnames(V)
      d <- ncol(V)
      out <- foreach(i=1:d,.combine="c") %:%
foreach(j=1:d,.combine="c") %do% {
          nm <- complete.cases(V[,c(i,j)])
          dcor(V[nm,i],V[nm,j])
        }
      out <- matrix(out,d,d)
      rownames(out) <- nms
      colnames(out) <- nms
      out
}

cpplus <-
function(V,reorder=TRUE,alpha=0.005,method="pearson",partial=FALSE,...
) {

  i <- unlist(lapply(data.frame(V),is.numeric))
  if(sum(i)==0) {
    catdat <- dummify(cbind(V)[,!i])
    alldat <- catdat
  } else {
    numdat <- cbind(V[,i])
      if(sum(i)==ncol(cbind(V))) {
        alldat <- numdat
      } else {
        catdat <- dummify(cbind(V)[,!i])
        alldat <- data.frame(numdat,catdat)
      }
  }
  alldat <- as.data.frame(alldat)
  allcor <- cor(alldat,use="pairwise.complete.obs")
  pmat <- rcorr(as.matrix(alldat),type="pearson")$P
  diag(pmat) <- 1
  pmat[is.na(pmat)] <- 1
  if(partial==TRUE) {
    allcor <- pcor(alldat)$estimate
    pmat <- pcor(alldat)$p.value
    diag(pmat) <- 1
    pmat[is.na(pmat)] <- 1
  }
  if(method=="spearman" ) {
    allcor <-
cor(alldat,use="pairwise.complete.obs",method="spearman")
    pmat <- rcorr(as.matrix(alldat),type="pearson")$P
    diag(pmat) <- 1
    pmat[is.na(pmat)] <- 1
    if(partial==TRUE) {
      allcor <- pcor(alldat,method="spearman")$estimate
      pmat <- pcor(alldat)$p.value
      diag(pmat) <- 1
      pmat[is.na(pmat)] <- 1
    }
  }
  if(method=="dcor") {
    allcor <- mdcor(alldat)
    pmat <- matrix(0,nrow(allcor),ncol(allcor))
    cat("\n","p-values not available\n")
  }

  cat("\n",sum(is.na(allcor[lower.tri(allcor)])),"missing correlations
set to zero\n\n")
  flush.console()
  allcor[is.na(allcor)] <- 0
  diag(allcor) <- 1

  allcor2 <- ifelse(pmat>alpha,0,allcor)
  rownames(allcor2) <- colnames(alldat)
  colnames(allcor2) <- colnames(alldat)
  rownames(pmat) <- colnames(alldat)
  colnames(pmat) <- colnames(alldat)
  allcor2col <-  ifelse(pmat>alpha,adjustcolor("white"),"black")

  colpal <- colorRampPalette(c("darkcyan","white","burlywood"))

  if(reorder==FALSE) {
    corrplot(allcor2,method="color",tl.col="black",p.mat=pmat,
sig.level=alpha,insig="pch",pch=4,pch.cex=0.6,pch.col="grey50",

addCoefasPercent=TRUE,addgrid.col="grey",col=colpal(100),addCoef.col=a
llcor2col,cl.pos="b")
  }
  if(reorder==TRUE) {
    horder <-
corrplot(allcor2,method="color",col="white",order="hclust",hclust.meth
od="ward.D")
    horder <- rownames(horder$corr)
    corrplot(allcor2,method="color",tl.col="black",p.mat=pmat,
sig.level=alpha,insig="pch",pch=4,pch.cex=0.6,pch.col="grey50",

addCoefasPercent=TRUE,addgrid.col="grey",col=colpal(100),addCoef.col=a
llcor2col[horder,horder],cl.pos="b",
order="hclust",hclust.method="ward.D")
  }
  invisible(alldat)
}
```