First question from data request: 1. Does atypical perceptual processing track with reduced cognitive and social functioning? 2. Which social and cognitive functioning variables best predict atypical task performance? 3. Does accounting for cognitive/social functioning significantly reduce group differences?

So atypical perception is: 1. Ebbinghaus -> context\_sensitivity, misleading\_index, helpful\_index (don’t know what 2sd is yet) 2. Mooney -> Silverstein’s paper from 2021 seems largely focused on upright\_percent\_reported and inverted\_percent\_reported

Cognitive functioning is: 1. the Wide Range Achievement Test (WRAT) -> wrat\_standardscore 2. the Brief Assessment of Cognition in Schizophrenia-Symbol Coding (BACS) -> bacs\_total 3. Hopkins Verbal Learning Test-Revised (HVLT-R) -> hvlt

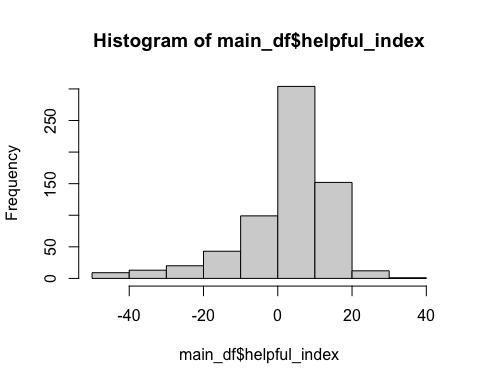
Social functioning is Global Functioning Scale: 1. Social and Role (GFS-S/R) -> gfs\_current, gfr\_current 2. Social Phobia Scale ->sps\_total

so let’s load data

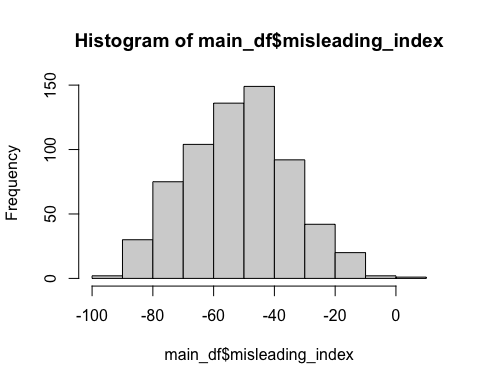
library(data.table)  
main\_dir = "/Users/victorpokorny/Library/CloudStorage/GoogleDrive-vpokorny123@gmail.com/My Drive/CAPR Ebbinghaus and Mooney/"  
load(paste0(main\_dir,'RData/fixed\_upright\_faces\_data.RData'))  
source(paste0(main\_dir,'R\_scripts/funcs.R')) # big group of functions

let’s look at hists and scatterplots of key variables starting with Ebbinghaus

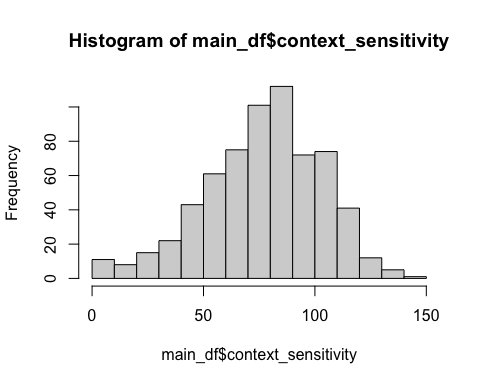
hist(main\_df$helpful\_index)



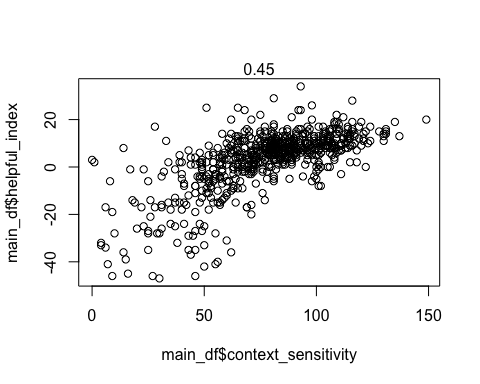
hist(main\_df$misleading\_index)



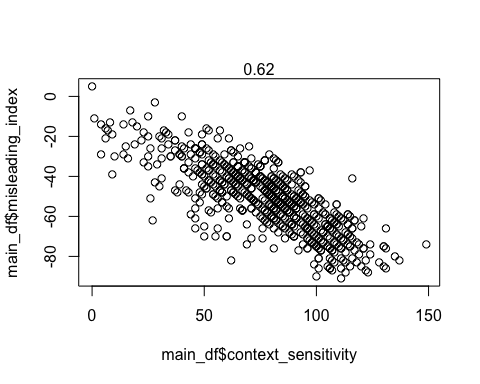
hist(main\_df$context\_sensitivity)



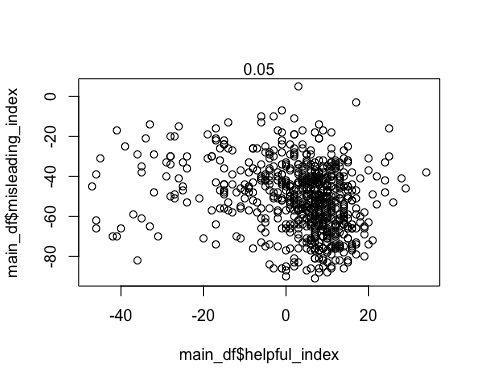
scatterplot(main\_df$context\_sensitivity, main\_df$helpful\_index)



scatterplot(main\_df$context\_sensitivity, main\_df$misleading\_index)



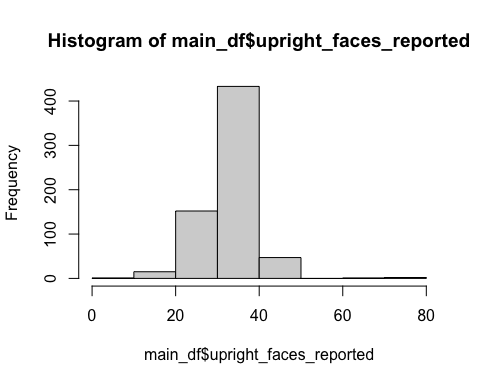
scatterplot(main\_df$helpful\_index, main\_df$misleading\_index)



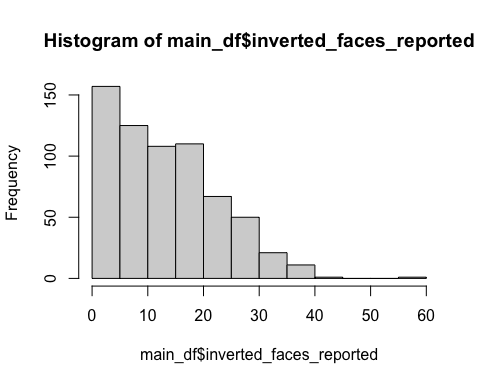
#ok so helpful and misleading aren't strongly correlated with each other, but both are   
#strongly correlate with the difference score (which makes sense)

next let’s look at Mooney

hist(main\_df$upright\_faces\_reported)



hist(main\_df$inverted\_faces\_reported)



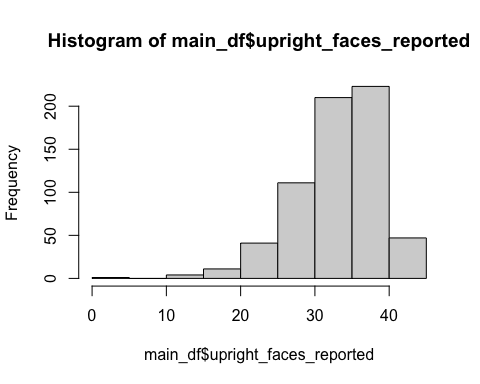
sum(!is.na(main\_df$inverted\_faces\_reported))

## [1] 651

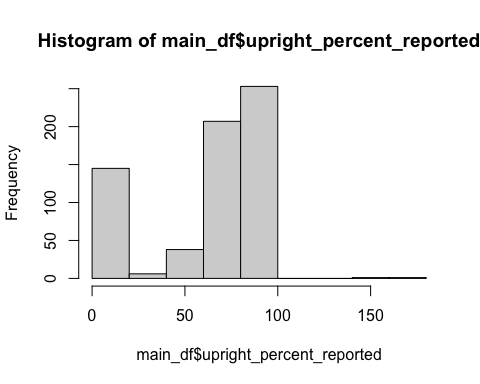
sum(!is.na(main\_df$upright\_faces\_reported))

## [1] 651

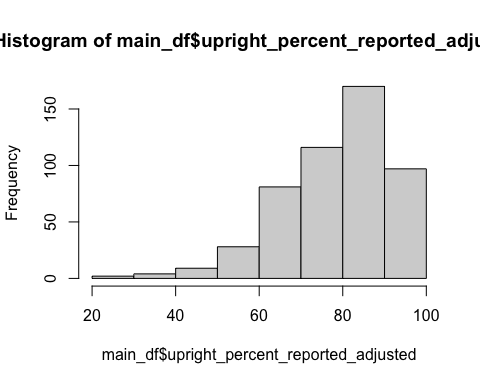
#shouldn't there be a max of 43 responses?  
#let's get rid of folks with more than 43 upright or inverted responses  
main\_df$upright\_faces\_reported[main\_df$upright\_faces\_reported>43] <- NA  
main\_df$inverted\_faces\_reported[main\_df$inverted\_faces\_reported>43] <- NA  
hist(main\_df$upright\_faces\_reported)



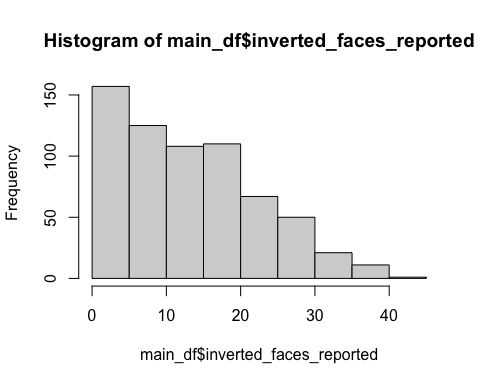
hist(main\_df$upright\_percent\_reported)



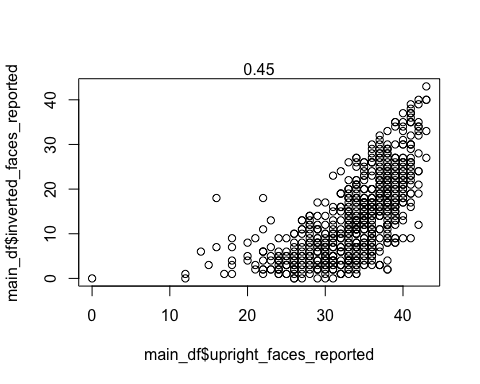
hist(main\_df$upright\_percent\_reported\_adjusted)



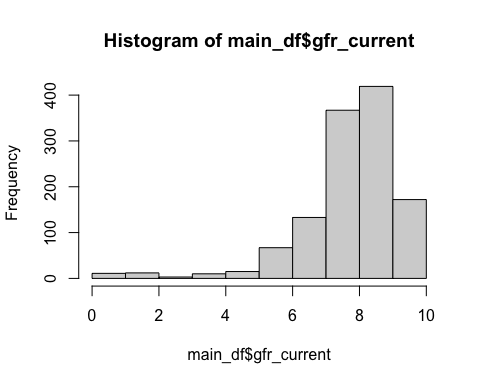
hist(main\_df$inverted\_faces\_reported)



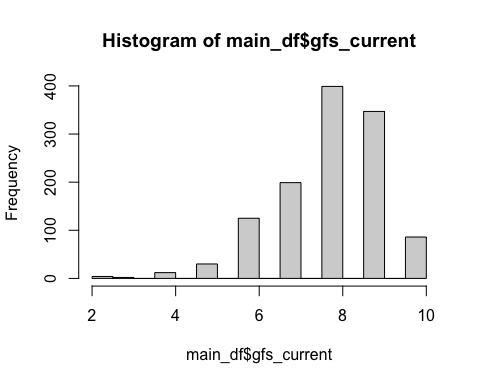
#phew so hopefully we've sufficiently cleaned these data  
scatterplot(main\_df$upright\_faces\_reported,main\_df$inverted\_faces\_reported)

 now let’s look at social variables…

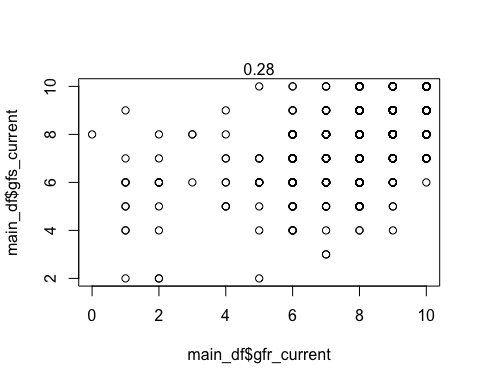
hist(main\_df$gfr\_current) # ceiling effect here as well



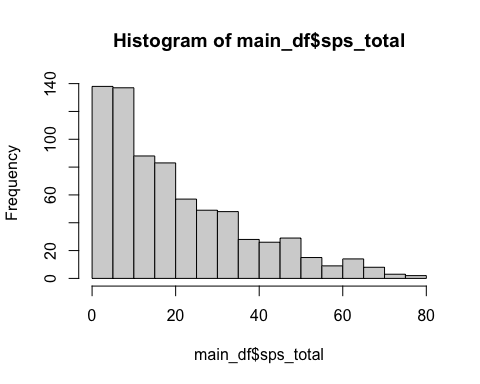
hist(main\_df$gfs\_current) # bit of a ceiling effect here it would seem



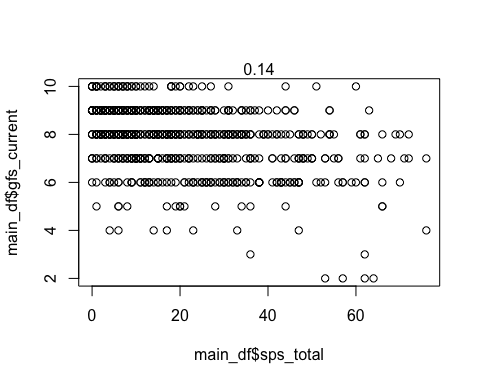
scatterplot(main\_df$gfr\_current,main\_df$gfs\_current ) #apparent range restriction here



#how about sps... very ugly  
hist(main\_df$sps\_total)



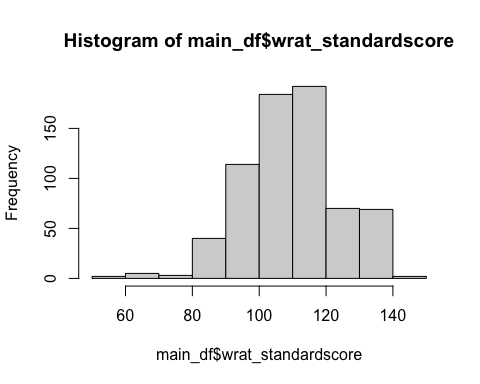
#looks quite a bit better!   
scatterplot(main\_df$sps\_total,main\_df$gfs\_current)



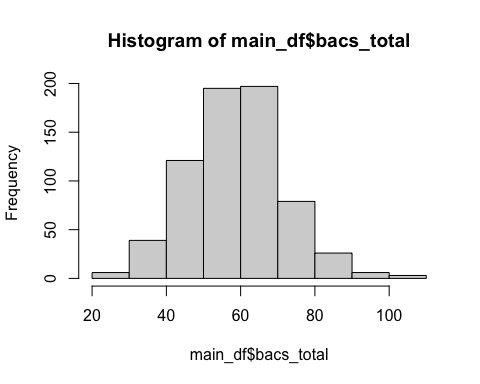
#log transform doesn't change correlation much

now let’s look at cog variables

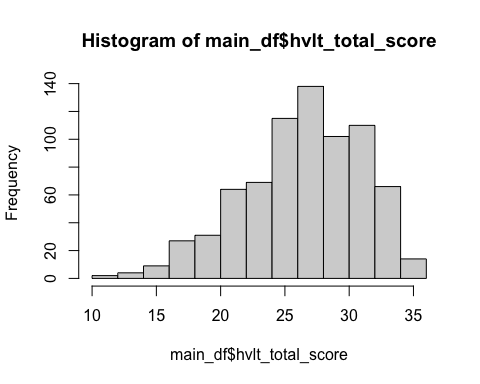
hist(main\_df$wrat\_standardscore) #looks fine



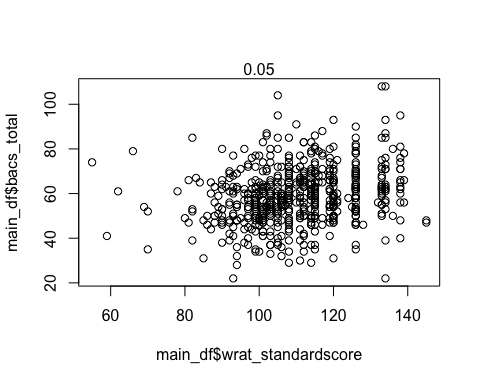
hist(main\_df$bacs\_total) #looks fine



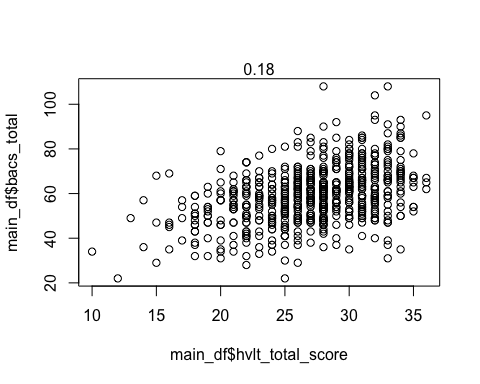
hist(main\_df$hvlt\_total\_score) #seems fine for now



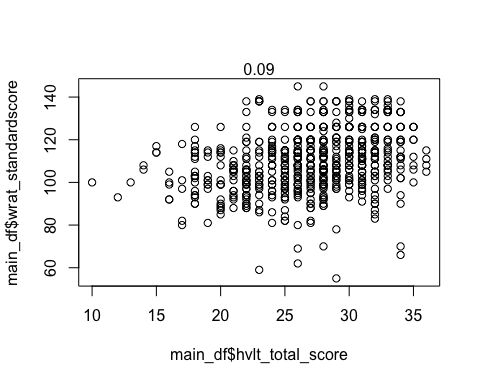
scatterplot(main\_df$wrat\_standardscore, main\_df$bacs\_total) #wow very little association



scatterplot(main\_df$hvlt\_total\_score, main\_df$bacs\_total) #better



scatterplot(main\_df$hvlt\_total\_score, main\_df$wrat\_standardscore) #not great



Finally we need to do a bit of recoding of some variables

main\_df$help\_index = main\_df$helpful\_index #something weird about this variable label  
main\_df$phenotype\_final<-ifelse(main\_df$phenotype=="hsc",  
 ifelse(main\_df$p1\_sev>1 |   
 main\_df$p2\_sev>1 |   
 main\_df$p3\_sev>1 |  
 main\_df$p4\_sev>1 |   
 main\_df$p5\_sev>1,  
 "hsc\_sub","hsc\_other"),main\_df$phenotype)  
  
main\_df$phenotype <- as.factor(main\_df$phenotype)  
main\_df$phenotype\_final <- factor(main\_df$phenotype\_final,   
 levels = c('hc','hsc\_other','hsc\_sub','chr') )  
#we want only baseline values   
main\_df <- main\_df[main\_df$visit == 'bl',]  
  
#there is a weird duplicate so let's drop for now  
main\_df<-main\_df[!duplicated(main\_df$src\_subject\_id),]  
  
#let's also drop folks with NA for their phenotype\_final (probably due to having NAs on SIPS)  
main\_df<-main\_df[!is.na(main\_df$phenotype\_final)]  
  
#let's make biological sex numeric and make it zeros and ones  
main\_df$demo\_biological\_sex <- as.numeric(main\_df$demo\_biological\_sex)-1  
  
#get rid of folks with negative context sensitivity  
main\_df$context\_sensitivity <- ifelse(main\_df$context\_sensitivity<0 |   
 is.na(main\_df$context\_sensitivity),  
 NA,  
 main\_df$context\_sensitivity)  
  
  
  
#let's also make some plotworthy variable names  
main\_df$`BACS Total`<-main\_df$bacs\_total  
main\_df$`WRAT Standard`<-main\_df$wrat\_standardscore  
main\_df$`HVLT Total`<-main\_df$hvlt\_total\_score  
main\_df$`Global Functioning: Role`<-main\_df$gfr\_current  
main\_df$`Global Functioning: Social`<-main\_df$gfs\_current  
main\_df$`Social Phobia Scale`<-main\_df$sps\_total  
main\_df$`Context Sensitivity`<-main\_df$context\_sensitivity\_all\_trials  
main\_df$`Misleading Index`<-main\_df$misleading\_index  
main\_df$`Inverted Faces Reported`<-main\_df$inverted\_faces\_reported  
main\_df$`Upright Faces Reported`<-main\_df$upright\_faces\_reported

let’s save out df and create new script since this one is getting long

#let's save as .RData file to keep meta-data  
save(main\_df,file = paste0(main\_dir,'RData/cleaned.RData'))