```
library(memisc)
library(reshape2)
library(plyr)
library(dplyr)
library(ggplot2)

df <- as.data.set(spss.system.file("./Data/College coping data (complete).sav"))
df <- as.data.frame(df)

#reshape to long form
df$ID <- rownames(df)
dfl <- melt(df, id = c("ID", "class", "ethnicity", "gender", "age"))

#just s, b and m
dfl.bm <- dfl[which(grepl("s|b|m", dfl$variable)), ]

#Create vartype variable indicating the type of variable by pulling the first letter
#from the variable name
dfl.bm$vartype <- gsub("^(\\w{1})\\d+$", "\\1", dfl.bm$variable)

#create condensed variables by summing over the three variable types respectively
df.bm.sum <- dfl.bm %>%
  group_by(ID, class, ethnicity, gender, age, vartype) %>%
  summarise(total = sum(value, na.rm = TRUE))#Some NA's in the data, you should know what these mean

#reshape to wide format
df.sbm.w <- reshape(as.data.frame(df.bm.sum),
                   timevar = "vartype",
                   times = c("s", "b", "m"),
                   idvar = c("ID", "class", "ethnicity", "gender", "age"),
                   direction = "wide")

#Simple plot of the relationship between B and M
ggplot(df.sbm.w, aes(x = total.s, y = total.b)) +
  geom_point() +
  geom_smooth(method = "lm")

# correlation of B and M
cbm <- cor(df.sbm.w$total.s, df.sbm.w$total.b, use = "pairwise.complete.obs")

# Basic model with interaction
m1 <- lm(total.m ~ total.s + total.b + total.s:total.b, data = df.sbm.w)
#Create summary of the model for evaluation
sm1 <- summary(m1)
# View the summary
print(sm1) #nothing significant

#View histogram of B to see where potential cutpoints might be
hist(df.bm.w$total.b)

#looks pretty normal, can use quartiles to create 4 categories
cuts <- quantile(x = df.sbm.w$total.b, probs = c(0.33, 0.66))

#remember, you can learn more about a function by typing a ? before the function:
?quantile

#Create a new variable with the cuts using logic
df.sbm.w$bCat <- with(df.sbm.w, ifelse(total.b < cuts[1], "Low",
                       ifelse(total.b >= cuts[1] & total.b < cuts[2], "Moderate",
                              ifelse(total.b >= cuts[2], "High", NA))))

#set the factor levels so we can control which category is the "reference"
df.sbm.w$bCat <- factor(df.sbm.w$bCat, levels = c("Low", "Moderate", "High"))

#New model with dichotomized  B variable
m2 <- lm(total.m ~ total.s + bCat + total.s:bCat, data = df.sbm.w)
#Create summary of the model for evaluation
sm2 <- summary(m2)
# View the summary
print(sm2) #the highest category of B has a significant difference in mean S than the lowest
 # the second highest is 'almost' significant ;)

#visualize the results
ggplot(df.bm.w, aes(x = bCat, y = total.s)) +
  geom_boxplot()

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

#Creating factors for Academic Engagment variable
df$skills <- with(df, 0.64*m143 + 0.59*m144 + 0.57*m145 + 0.55*m146 + 0.53*m147 + 0.53*m148 + 0.53*m149 + 0.51*m150 +
0.47*m151)

df$emot <- with(df, 0.86*m152 + 0.86*m153 + 0.54*m154 + 0.46*m155 + 0.43*m156)

df$part <- with(df, 0.82*m157 + 0.64*m158 + 0.57*m159 + 0.55*m160 + 0.50*m161 + 0.45*m162)
```

```
df$perf <- with(df, 0.77*m163 + 0.68*m164 + 0.64*m165)

#creating the other variables as part of the data field
df$total.s <- rowSums(df[, which(grepl("^s\\d{1,2}" , names(df)))], na.rm = TRUE)
df$total.b <- rowSums(df[, which(grepl("^b\\d{1,2}" , names(df)))], na.rm = TRUE)



#Simple plot of the relationship between s and skills
ggplot(df, aes(x = total.b, y = perf)) +
  geom_point() +
  geom_smooth(method = "lm")

# correlation of variables with academic eng factors
cbskills <- cor(df$total.b, df$skills, use = "pairwise.complete.obs")

# Factored (academic engagement) model with interaction
m4 <- lm(emot ~ total.s + total.b + total.s:total.b, data = df)
#Create summary of the model for evaluation
sm4 <- summary(m4)
# View the summary
print(sm4) #nothing significant

m3 <- lm(skills ~ total.s + total.b + total.s:total.b, data = df)
#Create summary of the model for evaluation
sm3 <- summary(m3)
# View the summary
print(sm3) #nothing significant

m5 <- lm(part ~ total.s + total.b + total.s:total.b, data = df)
#Create summary of the model for evaluation
sm5 <- summary(m5)
# View the summary
print(sm5) #nothing significant

m6 <- lm(perf ~ total.s + total.b + total.s:total.b, data = df)
#Create summary of the model for evaluation
sm6 <- summary(m6)
# View the summary
print(sm6) #nothing significant
```