# Load necessary packages

```r
library(tidyverse)
library(data.table)
library(lavaan)
library(semPlot)
```

# MBT scoring function

```r
questions.vec <- paste('Q', rep(1:26), sep = '')
answers <- c(2, 4, 5, 3, 1, 3, 3, 4, 1, 5, 5, 3, 2, 2, 5, 1, 4, 2, 3, 3, 1, 2, 4, 1, 1, 5)

df.answers <- data.frame(t(data.frame(answers = answers)))
names(df.answers) <- questions.vec

Read.Score.MBT <- function(file, skip.vec = c(1)){
  dt <- fread(file, header = TRUE)[-skip.vec,]

  score.mat <- sweep(dt[, ..questions.vec], 2,
                     as.numeric(as.character(unlist(df.answers))), "==")

  df <- cbind(dt, 'score' = rowSums(score.mat))
  return(df)
}
```

# Read MBT files

```r
df.MBT.P1112.F2019pre <- Read.Score.MBT('C:/Users/Cole/Documents/MBT_DATA/Physics_1112_Pre_Test_Fall_20
                                        skip.vec = c(1, 2)) %>%
  filter((Finished == 1) & (Q82 == 1)) %>% # only keep students that finished and consented
  select(QA, QB, QC, QD, score) %>%
  `colnames<-`(c('last.name', 'first.name', 'net.id', 'student.id', 'MBT.score'))

df.MBT.P1112.F2019post <- Read.Score.MBT('C:/Users/Cole/Documents/MBT_DATA/Physics_1112_Post_Test_Fall_
                                     skip.vec = c(1, 2)) %>%
  filter((Finished == 1) & (Q82 == 1)) %>% # only keep students that finished and consented
  select(QA, QB, QC, QD, score) %>%
  `colnames<-`(c('last.name', 'first.name', 'net.id', 'student.id', 'MBT.score'))

df.MBT.P1112.S2019pre <- Read.Score.MBT('C:/Users/Cole/Documents/MBT_DATA/Physics_1112_Pre_Test_Spring_

df.MBT.P1112.S2019pre <- subset(df.MBT.P1112.S2019pre,
                                select = c('V5', 'Q55_1', 'Q49', 'Q51',
                                           'Q53', 'Q61', 'score')) %>%
  filter((V5 == 1) & (Q55_1 == 1)) %>% # only keep students that finished and consented
  select(Q49, Q51, Q53, Q61, score) %>%
  `colnames<-`(c('first.name', 'last.name', 'net.id', 'student.id', 'MBT.score'))
```

```
df.MBT.P1112.S2019post <- Read.Score.MBT('C:/Users/Cole/Documents/MBT_DATA/Physics_1112_Post_Test_Sprin

df.MBT.P1112.S2019post <- subset(df.MBT.P1112.S2019post,
                                 select = c('V5', 'Q55_1', 'Q49', 'Q51',
                                            'Q53', 'Q61', 'score')) %>%
  filter((V5 == 1) & (Q55_1 == 1)) %>% # only keep students that finished and consented
  select(Q49, Q51, Q53, Q61, score) %>%
  `colnames<-`(c('first.name', 'last.name', 'net.id', 'student.id', 'MBT.score'))
```

## Read PLIC files

```
df.PLIC.pre <- fread('C:/Users/Cole/Documents/PLIC_DATA/Collective_Surveys/Complete/Complete_Concat.csv
  filter(!is.na(PreScores)) %>%
  select(Class_ID, Q5a_x, Q5b_x, Q5c_x, PreScores)

df.PLIC.P1112.F2019pre <- df.PLIC.pre %>%
  filter(Class_ID == 'R_9EVBSZgwQyP6mWZ') %>%
  select(-Class_ID) %>%
  `colnames<-`(c('ID', 'last.name', 'first.name', 'PLIC.score'))

df.PLIC.P1112.S2019pre <- df.PLIC.pre %>%
  filter(Class_ID == 'R_RKRNIWFu1gZuSPf') %>%
  select(-Class_ID) %>%
  `colnames<-`(c('ID', 'last.name', 'first.name', 'PLIC.score'))

df.PLIC.post <- fread('C:/Users/Cole/Documents/PLIC_DATA/Collective_Surveys/Complete/Complete_Concat.cs
  filter(!is.na(PostScores)) %>%
  select(Class_ID, Q5a_y, Q5b_y, Q5c_y, PostScores)

df.PLIC.P1112.F2019post <- df.PLIC.post %>%
  filter(Class_ID == 'R_9EVBSZgwQyP6mWZ') %>%
  select(-Class_ID) %>%
  `colnames<-`(c('ID', 'last.name', 'first.name', 'PLIC.score'))

df.PLIC.P1112.S2019post <- df.PLIC.post %>%
  filter(Class_ID == 'R_RKRNIWFu1gZuSPf') %>%
  select(-Class_ID) %>%
  `colnames<-`(c('ID', 'last.name', 'first.name', 'PLIC.score'))
```

## Matching MBT and PLIC

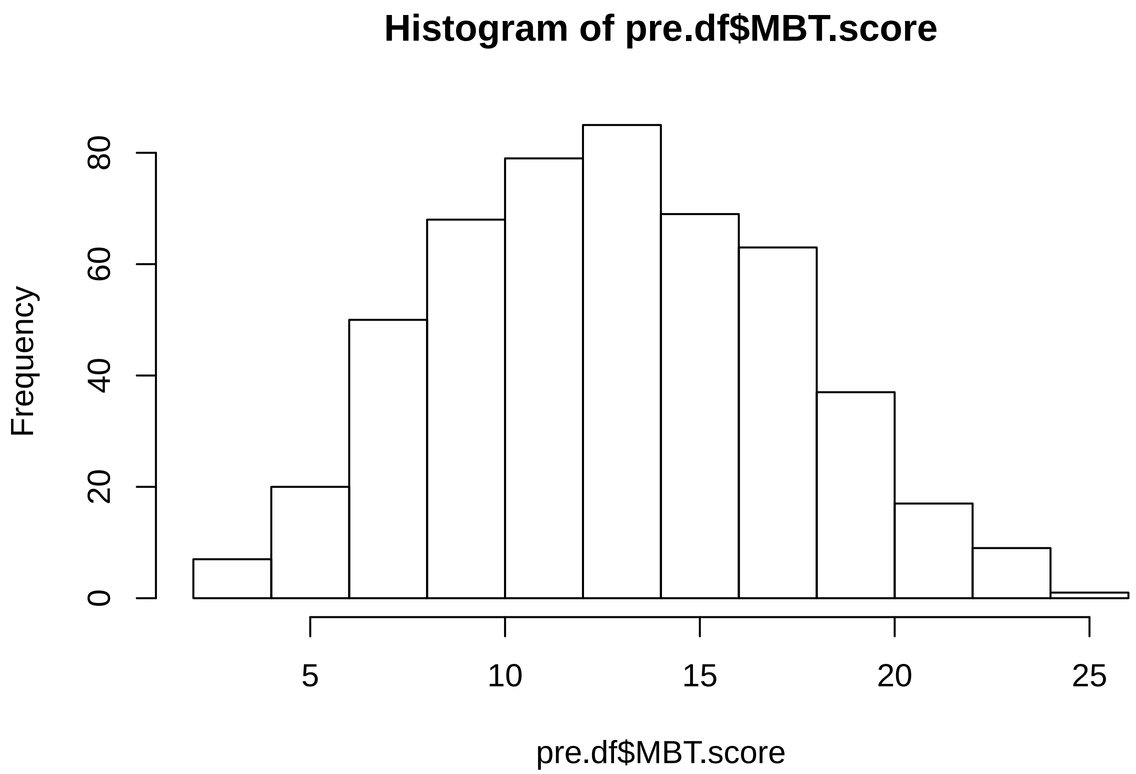
```
match.MBT.PLIC <- function(MBT.df, PLIC.df){
  netID.df <- inner_join(MBT.df, PLIC.df, by = c('net.id' = 'ID'))
  studentID.df <- inner_join(MBT.df, PLIC.df, by = c('student.id' = 'ID'))

match.df <- rbind(netID.df, studentID.df) %>%
  filter(!duplicated(student.id)) # %>%
  # select(MBT.score, PLIC.score)
}

F2019pre.df <- match.MBT.PLIC(df.MBT.P1112.F2019pre, df.PLIC.P1112.F2019pre)
S2019pre.df <- match.MBT.PLIC(df.MBT.P1112.S2019pre, df.PLIC.P1112.S2019pre)

pre.df <- rbind(F2019pre.df, S2019pre.df)

F2019post.df <- match.MBT.PLIC(df.MBT.P1112.F2019post, df.PLIC.P1112.F2019post)
S2019post.df <- match.MBT.PLIC(df.MBT.P1112.S2019post, df.PLIC.P1112.S2019post)

post.df <- rbind(F2019post.df, S2019post.df)
```

## Correlation and visualization (PRE)

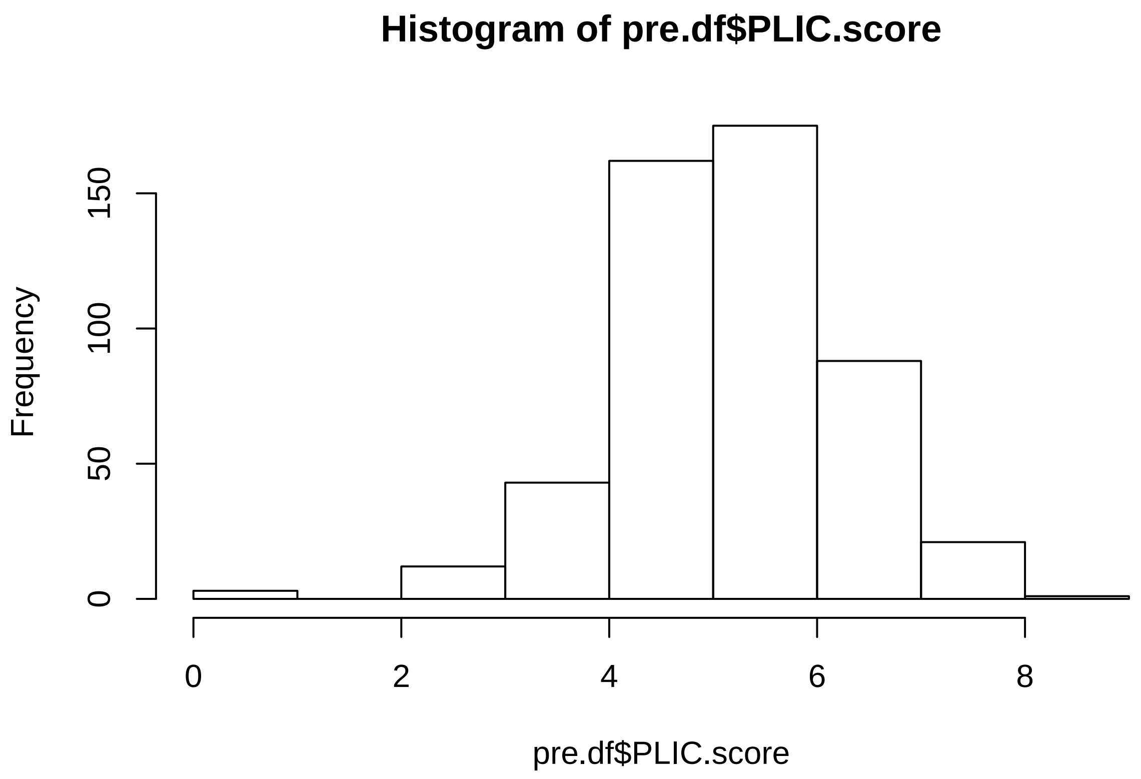
```
nrow(pre.df)
```

```
## [1] 505
```

```
hist(pre.df$MBT.score)
```



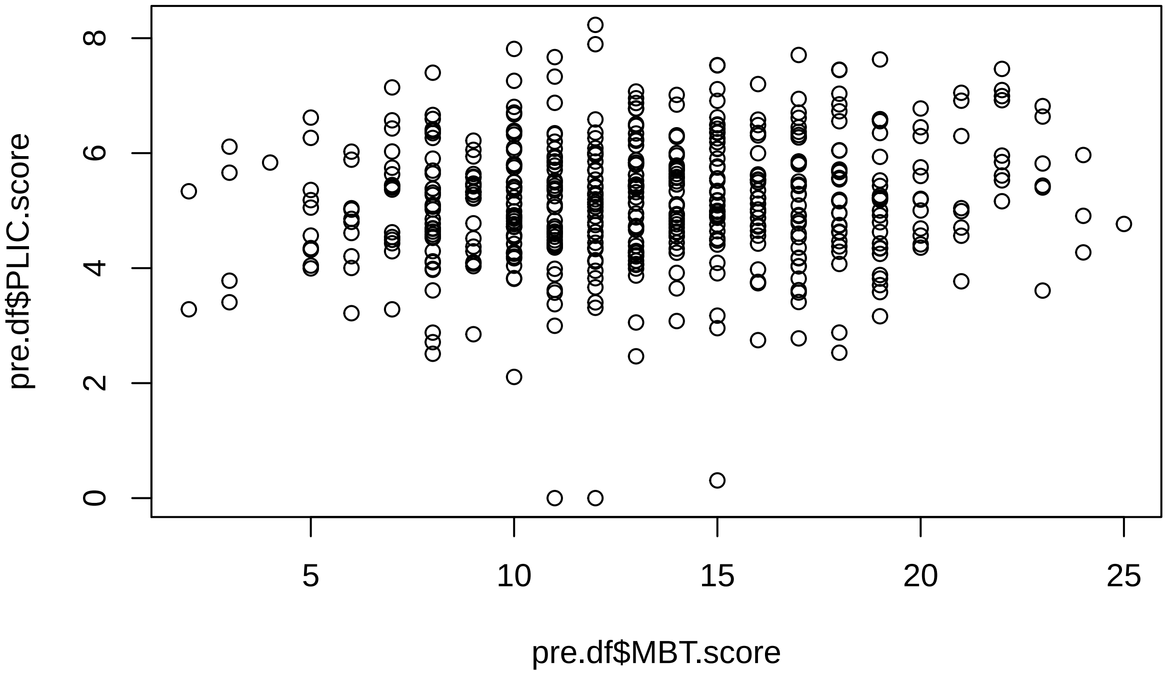
```
hist(pre.df$PLIC.score)
```



```
cor(pre.df$MBT.score, pre.df$PLIC.score)
```

```
## [1] 0.1144092
```

```
plot(pre.df$MBT.score, pre.df$PLIC.score)
```

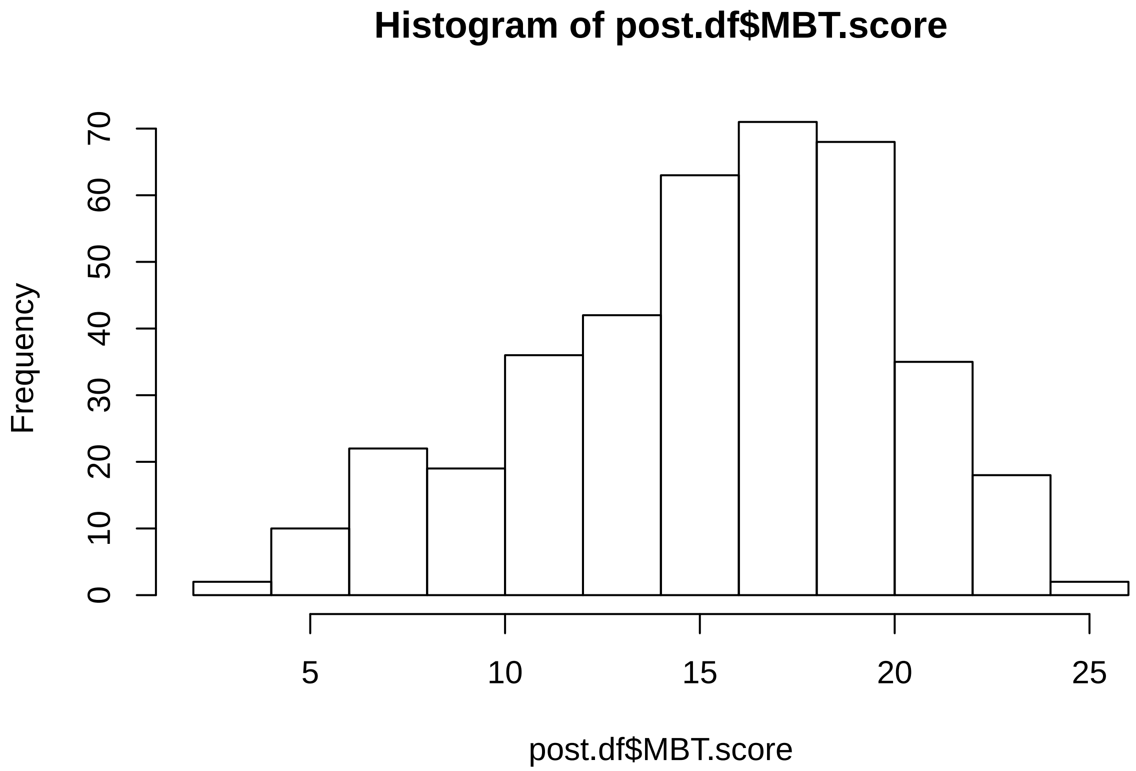


## Correlation and visualization (POST)

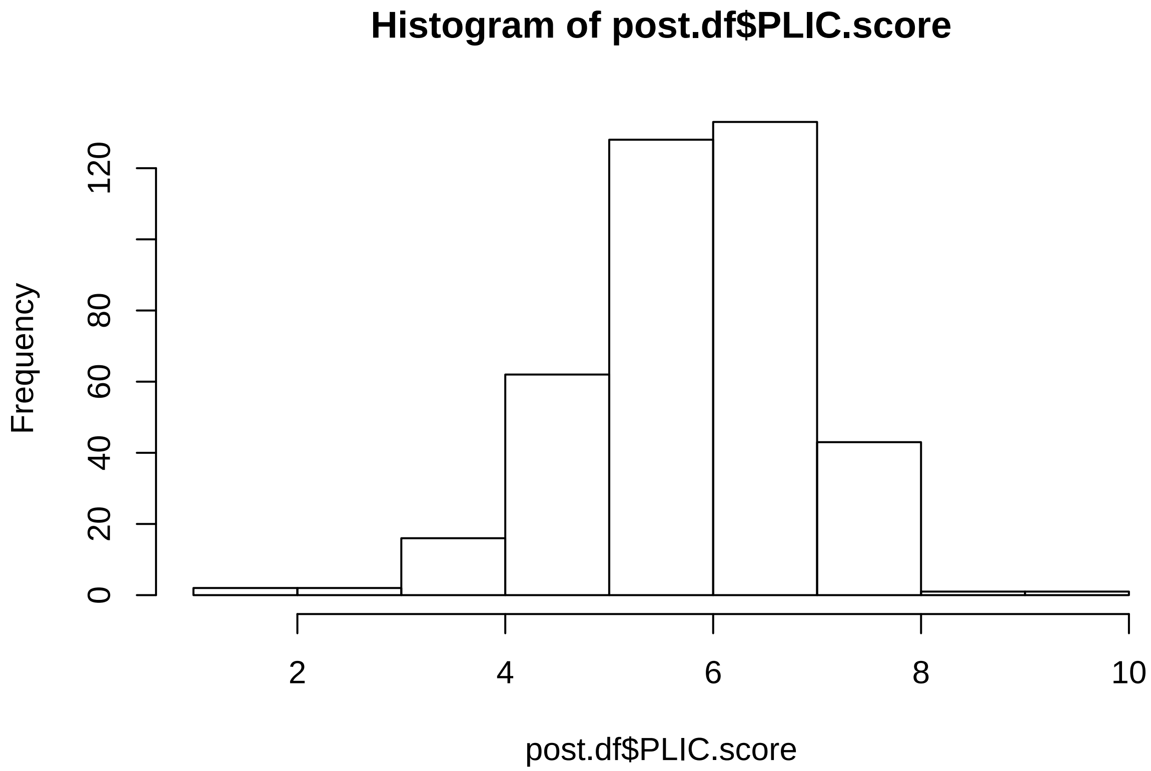
```
nrow(post.df)
```

```
## [1] 388
```

```
hist(post.df$MBT.score)
```



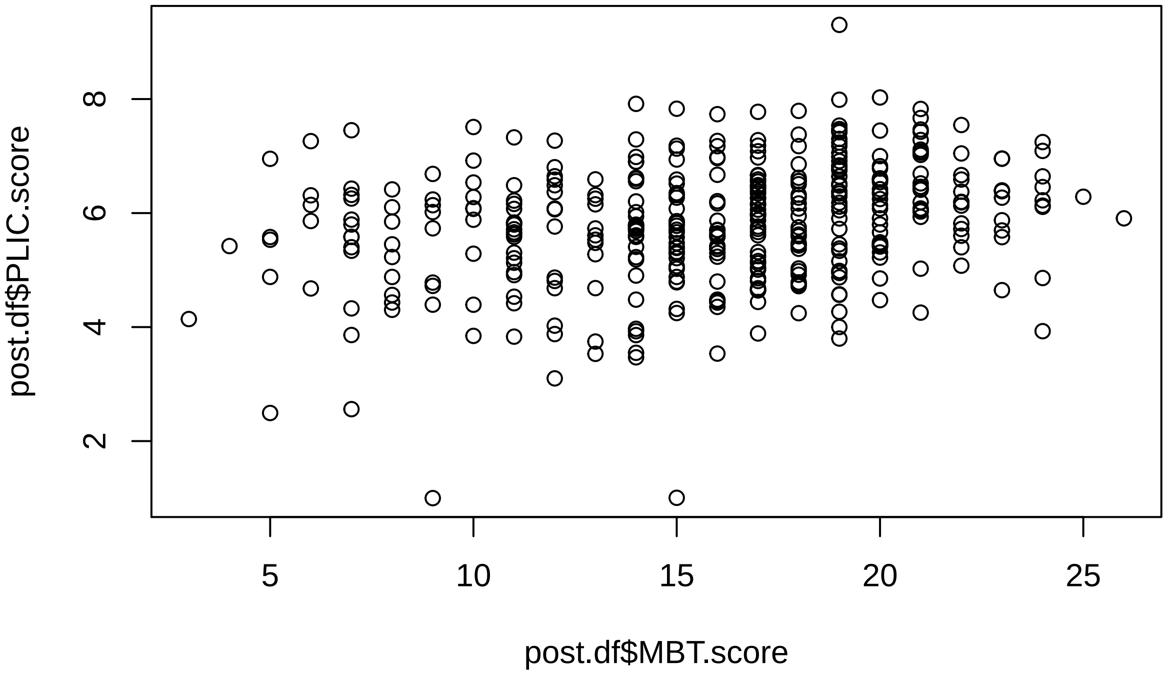
```r
hist(post.df$PLIC.score)
```



```
cor(post.df$MBT.score, post.df$PLIC.score)
```

```
## [1] 0.2585021
```

```
plot(post.df$MBT.score, post.df$PLIC.score)
```

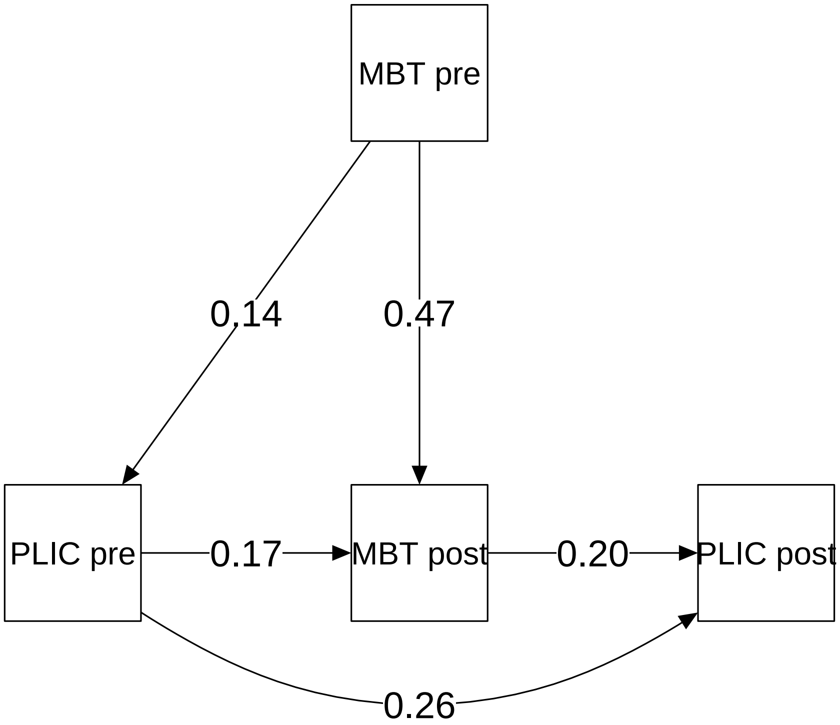


## SEM analysis

```
df <- inner_join(pre.df, post.df, by = 'student.id', suffix = c('.pre', '.post')) %>%
  filter(!duplicated(student.id)) %>%
  select(MBT.score.pre, MBT.score.post, PLIC.score.pre, PLIC.score.post)

mod <- '
  PLIC.score.pre ~ MBT.score.pre
  MBT.score.post ~ MBT.score.pre + PLIC.score.pre
  PLIC.score.post ~ PLIC.score.pre + MBT.score.post
'

fit <- sem(mod, df)
summary(fit, standardized = TRUE, fit.measures = TRUE, modindices = TRUE)
```

```
## lavaan 0.6-3 ended normally after 24 iterations
##
##   Optimization method                           NLMINB
##   Number of free parameters                          8
##
##   Number of observations                           351
##
##   Estimator                                         ML
##   Model Fit Test Statistic                       3.422
##   Degrees of freedom                                 1
##   P-value (Chi-square)                           0.064
##
## Model test baseline model:
##
##   Minimum Function Test Statistic              171.372
##   Degrees of freedom                                 6
##   P-value                                        0.000
##
## User model versus baseline model:
##
##   Comparative Fit Index (CFI)                    0.985
##   Tucker-Lewis Index (TLI)                       0.912
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)              -1975.730
##   Loglikelihood unrestricted model (H1)      -1974.019
##
##   Number of free parameters                          8
##   Akaike (AIC)                                3967.461
##   Bayesian (BIC)                              3998.347
##   Sample-size adjusted Bayesian (BIC)         3972.968
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.083
##   90 Percent Confidence Interval          0.000  0.187
##   P-value RMSEA <= 0.05                          0.183
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.025
##
## Parameter Estimates:
##
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##   Standard Errors                             Standard
##
## Regressions:
##                      Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   PLIC.score.pre ~
##     MBT.score.pre      0.037    0.014    2.707    0.007    0.037    0.143
##   MBT.score.post ~
##     MBT.score.pre      0.491    0.049   10.125    0.000    0.491    0.467
##     PLIC.score.pre     0.695    0.187    3.717    0.000    0.695    0.171
##   PLIC.score.post ~
##     PLIC.score.pre     0.249    0.049    5.126    0.000    0.249    0.262
##     MBT.score.post     0.047    0.012    3.900    0.000    0.047    0.200
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##    .PLIC.score.pre    1.167    0.088   13.248    0.000    1.167    0.980
##    .MBT.score.post   14.315    1.081   13.248    0.000   14.315    0.730
##    .PLIC.score.pst    0.931    0.070   13.248    0.000    0.931    0.867
##
## Modification Indices:
##
##               lhs op            rhs    mi    epc sepc.lv sepc.all
## 11  PLIC.score.pre ~~ PLIC.score.post 3.406 -0.816  -0.816   -0.783
## 12  MBT.score.post ~~ PLIC.score.post 3.406 -0.756  -0.756   -0.207
## 14  PLIC.score.pre  ~ PLIC.score.post 3.406 -0.876  -0.876   -0.832
## 15  MBT.score.post  ~ PLIC.score.post 3.406 -0.812  -0.812   -0.190
## 16 PLIC.score.post  ~   MBT.score.pre 3.406  0.026   0.026    0.105
## 19   MBT.score.pre  ~ PLIC.score.post 3.406  0.493   0.493    0.122
##    sepc.nox
## 11   -0.783
## 12   -0.207
## 14   -0.832
## 15   -0.190
## 16    0.025
## 19    0.122
```

```r
semPaths(fit, whatLabels = 'std', edge.color = 'black', edge.label.cex = 1.5, curve = 2, sizeMan = 12,
```


MBT pre
0.14
0.47
PLIC pre
0.17
MBT post
0.20
PLIC post
0.26