Formal Exploratory Analysis

Anand Rajan

Data Cleaning

```
## Rows: 592 Columns: 2453

## -- Column specification --------------------------------------------------------
## Delimiter: "\t"
## chr    (1): BACIDATE
## dbl (2452): M2ID, BACRAGE, BACBYR, BACRSEX, BACA1, BACA2, BACAS6A, BACAS6B, ...

##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

## Joining, by = "M2ID"
## Joining, by = "M2ID"
## Joining, by = "M2ID"
```

After completion of data cleaning there were 1108 observations with 20 variables. A fairly substantial data set to run analyses on. Now lets dive into some exploratory analyses starting with looking at our outcome variable cognition.

Exploratory

To start, we will be looking out our outcome variable cognition. We can evaluate cognition through three different variables. We have composite cognition z scores, then cognition z scores for episodic memory and executive functioning. We will look at the distributions of all three outcome variables to potentially inform us of what sort of analysese we can run. Moreover we will look at univariate statistics for each of the cognition scores.

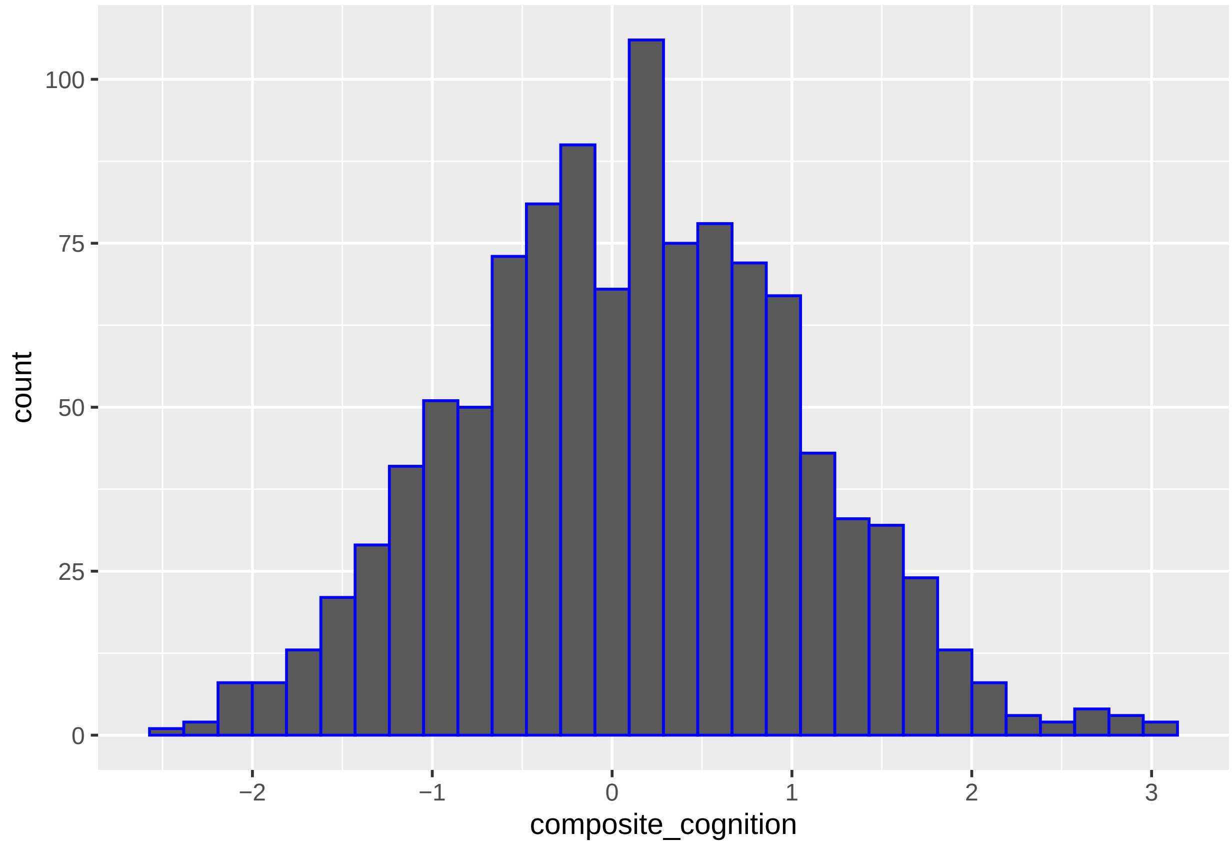
```
shapiro.test(midus$composite_cognition)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  midus$composite_cognition
## W = 0.99823, p-value = 0.3093
```

```
cognition_distribution <- ggplot(midus,aes(x=composite_cognition)) +
                        geom_histogram(color = "blue")

cognition_distribution
```

```
## Don't know how to automatically pick scale for object of type haven_labelled/vctrs_vctr/double. Defa
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



The histogram indicates a fairly normal distribution of composite cognition z scores. Furthermore normality was tested via the shapiro-wilk test, and at the 0.05 significance level we conclude the distribution of scores is approximately normal.

```
midus %>%
  summarize(
    mean_composite = mean(composite_cognition),
    median_composite=median(composite_cognition),
    sd = sd(composite_cognition),
    min = min(composite_cognition),
    max = max(composite_cognition)
  ) %>%
  knitr::kable()
```

mean_composite	median_composite	sd	min	max
0.1083309	0.1216147	0.9348151	-2.474524	3.051121

The table above provides univariate statistics regarding composite cognition scores in the data set.

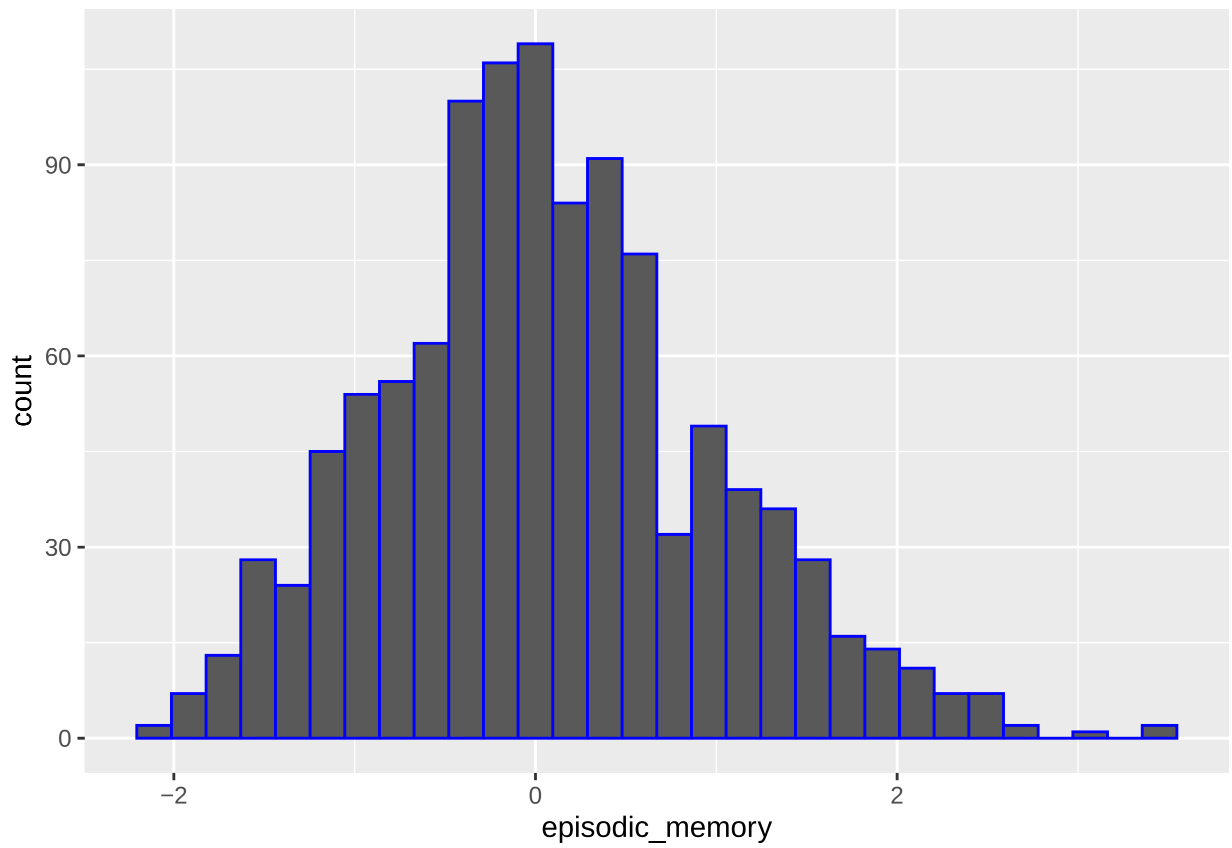
Let us now look at episodic memory.

```
episodic_memory_distribution <-ggplot(midus,aes(x=episodic_memory)) + geom_histogram(color="blue")
shapiro.test(midus$episodic_memory)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  midus$episodic_memory
## W = 0.98897, p-value = 2.327e-07
```

```
episodic_memory_distribution
```

```
## Don't know how to automatically pick scale for object of type haven_labelled/vctrs_vctr/double. Defa

## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



The histogram indicates the distribution of episodic memory z scores is right skewed. Normality was tested via the shapiro-wilk test, and at the 0.05 significance level, we conclude the distribution of scores is NOT normal.

```
midus %>%
  summarize(
    mean_episodic = mean(episodic_memory),
    median_episodic=median(episodic_memory),
    sd = sd(episodic_memory),
```

```
    min = min(episodic_memory),
    max = max(episodic_memory)
  ) %>%
  knitr::kable()
```

mean_episodic	median_episodic	sd	min	max
0.0679781	0.0128389	0.9150803	-2.155253	3.405208

The table above provides univariate statistics on episodic memory z scores.

Let's now take a look at executive functioning.

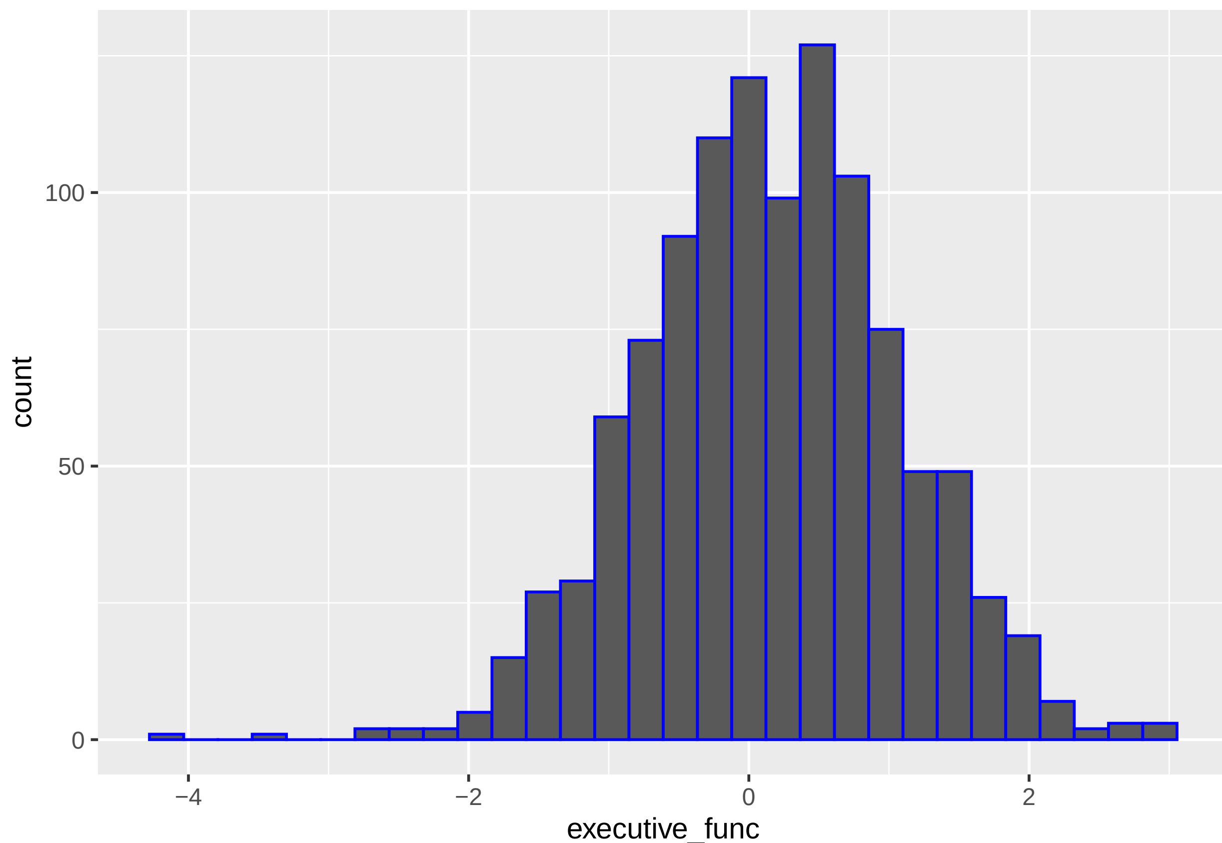
```
exec_functioning_distribution <- ggplot(midus,aes(x=executive_func)) + geom_histogram(color = "blue")
shapiro.test(midus$executive_func)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  midus$executive_func
## W = 0.99617, p-value = 0.007999
```

```
exec_functioning_distribution
```

```
## Don't know how to automatically pick scale for object of type haven_labelled/vctrs_vctr/double. Defau
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



The histogram indicates slight skewness, but mainly the existence of possible outliers. The shapiro wilks test also indicates non-normality, but I suspect this is due to outliers.

```
midus %>%
  summarize(
    mean_exec = mean(executive_func),
    median_exec=median(executive_func),
    sd = sd(executive_func),
    min = min(executive_func),
    max = max(executive_func)
  ) %>%
  knitr::kable()
```

mean_exec	median_exec	sd	min	max
0.1539497	0.1634781	0.9078578	-4.184306	2.905415

The table above provides univariate statistics on executive function z-scores.

Childhood Trauma

Childhood Trauma is broken down into 6 components scores. These components are emotional abuse,emotional neglect, minimization/denial,physical abuse, physical neglect, and sexual abuse. For this study, we will be categorizing trauma into two strata, Threat and Deprivation. Threat is the sum of the scores for emotional abuse, physical abuse and sexual abuse. While deprivation is the sum of emotional neglect, minimization/denial, and parental neglect. As you can see the division of the two strata make sense as one threat in the purview of childhood trauma refers to abuse, while deprivation refers to neglect. Now lets dive into each category.

```
midus %>%
  summarize(
    mean_threat = mean(threat),
    median_threat=median(threat),
    sd = sd(threat),
    min = min(threat),
    max = max(threat)
  ) %>%
  knitr::kable()
```

mean_threat	median_threat	sd	min	max
21.50309	18	9.067415	15	75

The table above provides univariate statistics for childhood threat scores.

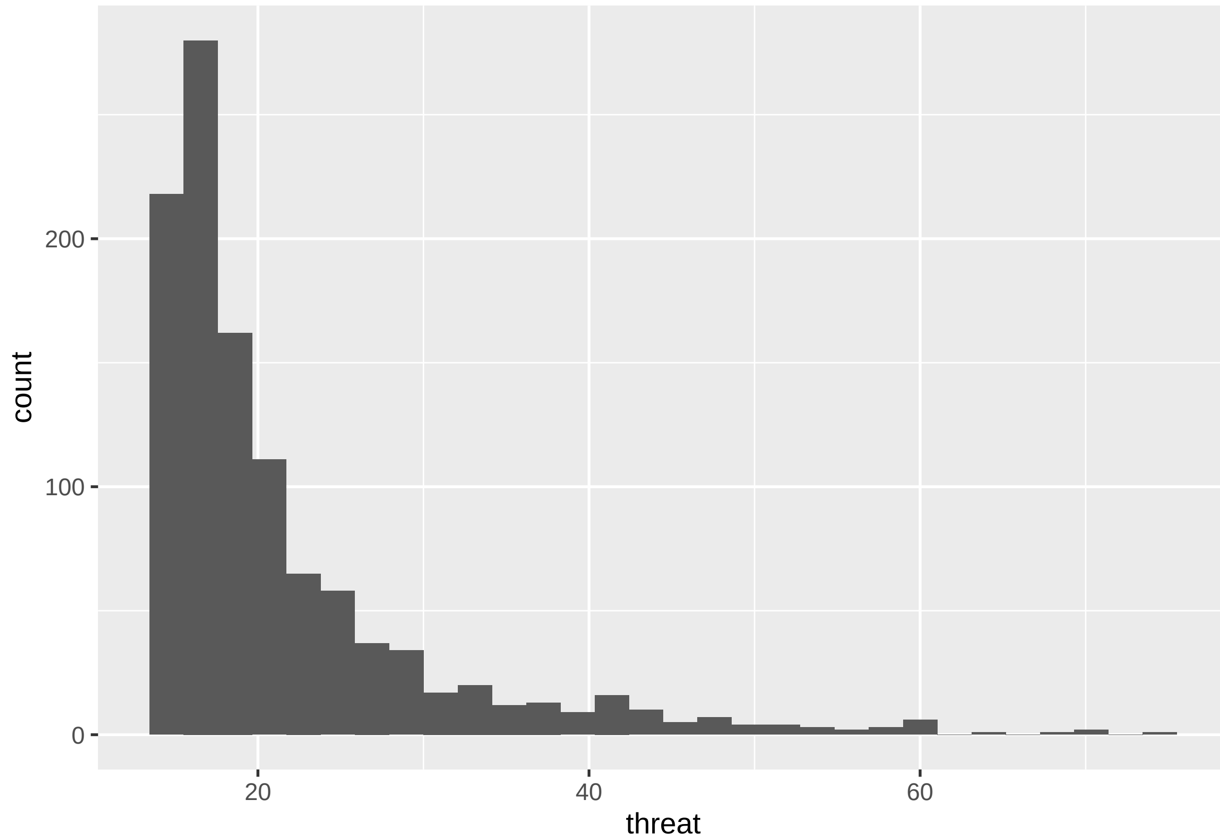
```
midus %>%
  summarize(
    mean_deprivation = mean(deprivation),
    median_deprivaton=median(deprivation),
    sd = sd(deprivation),
    min = min(deprivation),
    max = max(deprivation)
  )
```

```
##   mean_deprivation median_deprivaton       sd min max
## 1         17.10527                15 6.279626  10  43
```

The table above provides univariate statistics regarding deprivation scores.

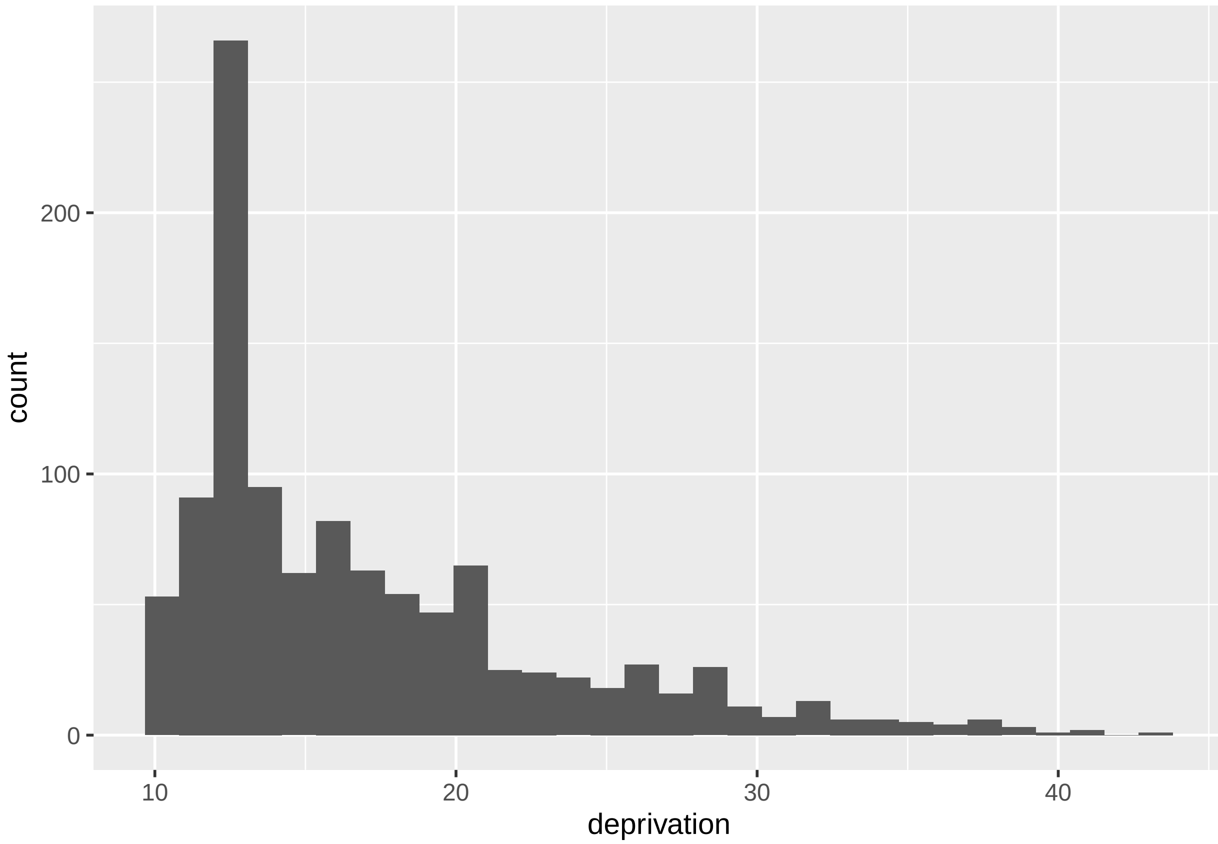
```
ggplot(midus,aes(x=threat)) +geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
ggplot(midus,aes(x=deprivation)) +geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

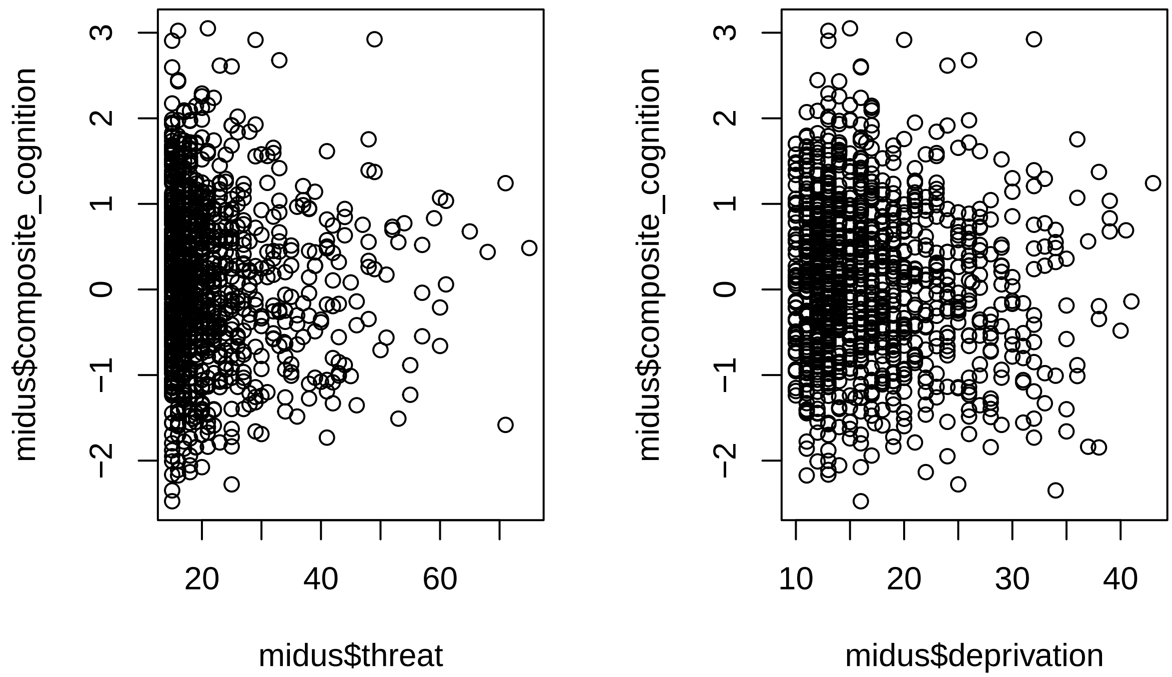


Unadjusted Analysis

Now that we have looked at both cognition and childhood trauma individually as variables, lets explore the relationship between these two variables.

Composite Cognition

```
par(mfrow=c(1,2))
plot(midus$threat,midus$composite_cognition)
plot(midus$deprivation,midus$composite_cognition)
```



The graphs above do not indicate any significant li

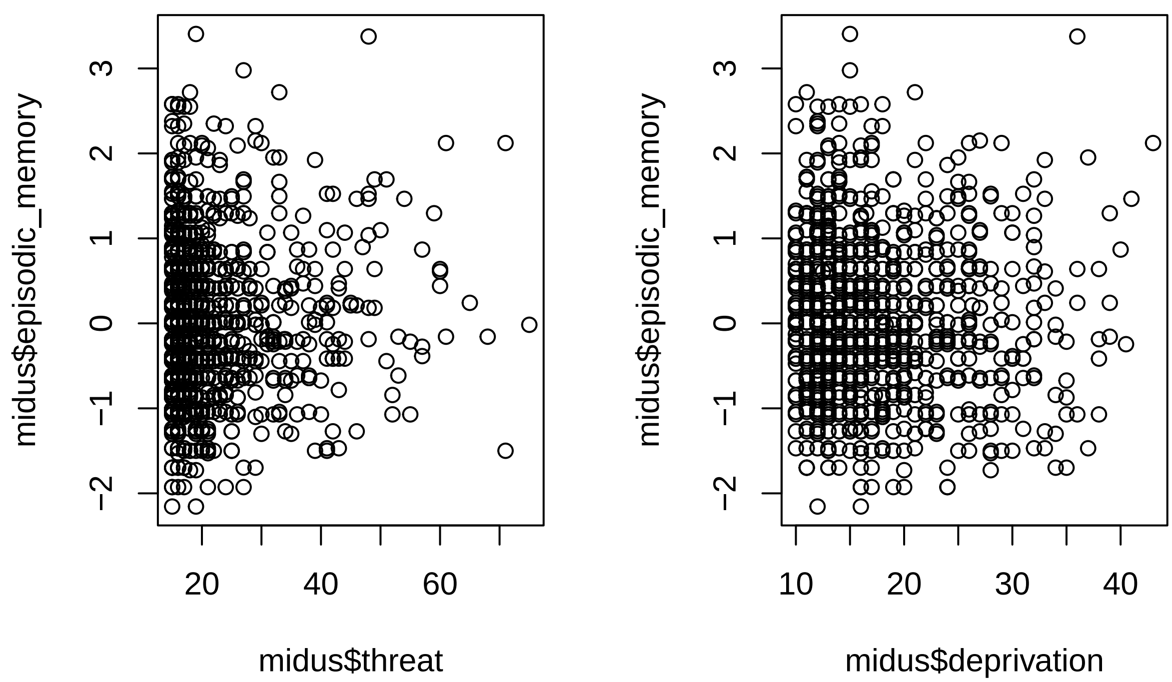
```
fit1 <- lmer(composite_cognition ~ threat + deprivation + (1|family_id), data=midus)

summary(fit1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: composite_cognition ~ threat + deprivation + (1 | family_id)
##    Data: midus
##
## REML criterion at convergence: 2550.7
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.97770 -0.48149 -0.01154  0.46162  2.31229
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  family_id (Intercept) 0.4383   0.6621
##  Residual              0.3761   0.6133
## Number of obs: 979, groups:  family_id, 844
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) 0.290394   0.090226   3.219
## threat      0.004134   0.004619   0.895
```

```
## deprivation -0.011905   0.006359  -1.872
##
## Correlation of Fixed Effects:
##             (Intr) threat
## threat      -0.292
## deprivation -0.476 -0.664
```

Episodic Memory

```
par(mfrow=c(1,2))
plot(midus$threat,midus$episodic_memory)
plot(midus$deprivation,midus$episodic_memory)
```



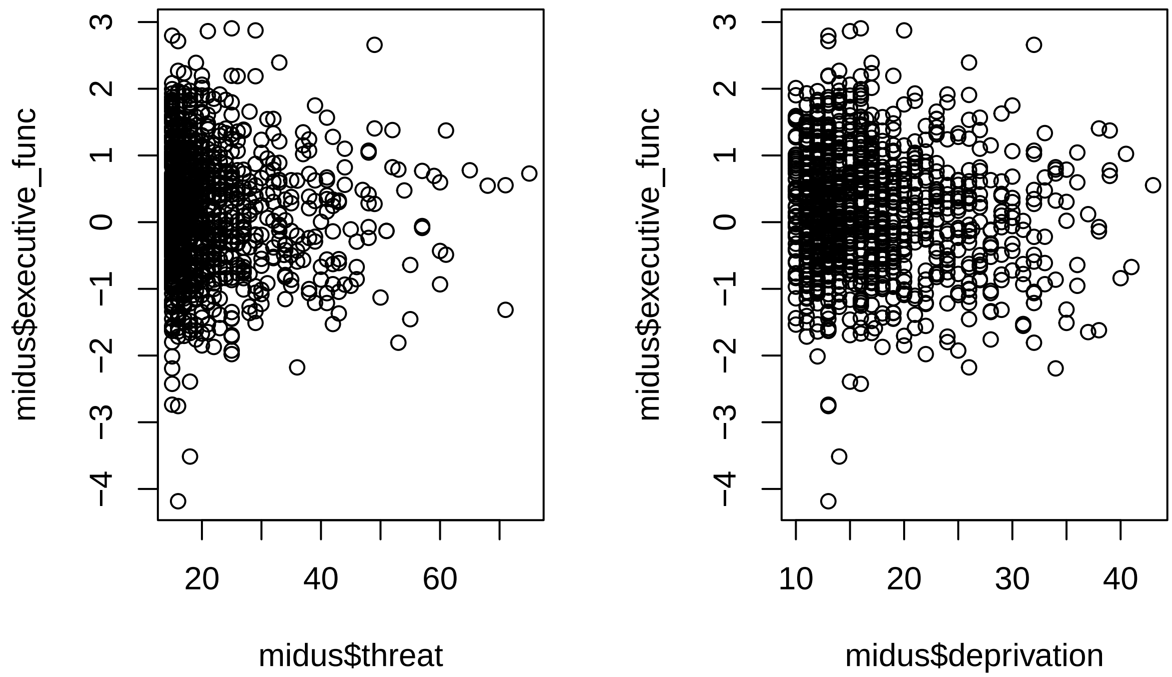

```
fit2 <- lmer(episodic_memory ~ threat + +deprivation + (1|family_id), data=midus)

summary(fit2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: episodic_memory ~ threat + +deprivation + (1 | family_id)
##    Data: midus
##
## REML criterion at convergence: 2585.6
##
```

```
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.2860 -0.5780 -0.0774  0.5541  3.4103
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  family_id (Intercept) 0.2263   0.4757
##  Residual              0.5877   0.7666
## Number of obs: 979, groups:  family_id, 844
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  0.073687   0.090306   0.816
## threat       0.006174   0.004670   1.322
## deprivation -0.006671   0.006483  -1.029
##
## Correlation of Fixed Effects:
##             (Intr) threat
## threat      -0.278
## deprivation -0.482 -0.671
```

Executive Functioning

```
par(mfrow=c(1,2))
plot(midus$threat,midus$executive_func)
plot(midus$deprivation,midus$executive_func)
```




```r
fit3 <- lmer(executive_func ~ threat + +deprivation + (1|family_id), data=midus)

summary(fit3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: executive_func ~ threat + +deprivation + (1 | family_id)
##    Data: midus
##
## REML criterion at convergence: 2454.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.2594 -0.4471  0.0001  0.4322  2.4856
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  family_id (Intercept) 0.4458   0.6677
##  Residual              0.3025   0.5500
## Number of obs: 979, groups:  family_id, 844
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  0.345884   0.086155   4.015
## threat       0.005031   0.004397   1.144
## deprivation -0.013564   0.006034  -2.248
```

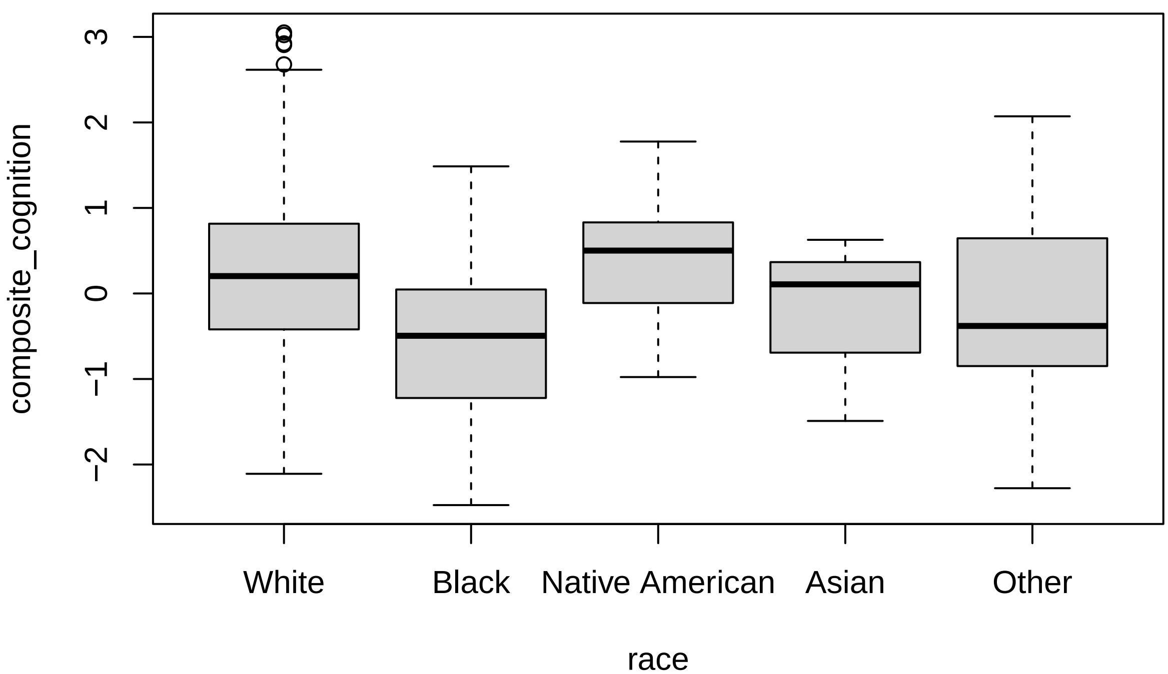
```
##
## Correlation of Fixed Effects:
##             (Intr) threat
## threat      -0.296
## deprivation -0.474 -0.661
```

Evaluating Confounding

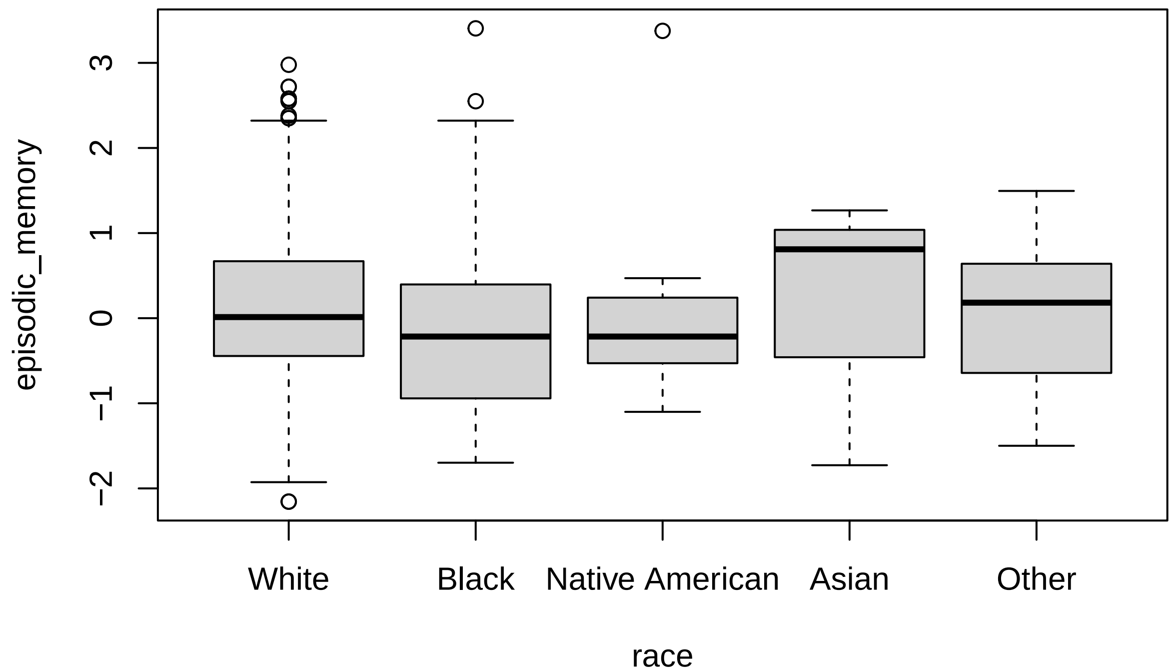
Now we will be getting into the meat of our analysis. First it should noted that the observations are in fact non-independent. Why is this the case? Well observations were not randomly sampled. We see sampling of participants from the same family, hence the inclusion of family ID in our data set. Cluster sampling was utilized. Therefore participants from the same family likely have similar levels of childhood trauma experience and similar cognition. Now given this, much of the analysis we could question the validity of the analysis run prior, but the purpose of the analysis before is to get an initial look at the data we are working with. the linear models fitted were merely to detect if there is any un-adjusted association.

Now we will delve into the real analysis. First lets start by looking at confounding. The variables we will be looking at is race, education, age, and sex. To evaluate confounding we will build linear mixed models evaluating the covariates association to both the outcome(composite_cognition,episodic_memory,executive_func)

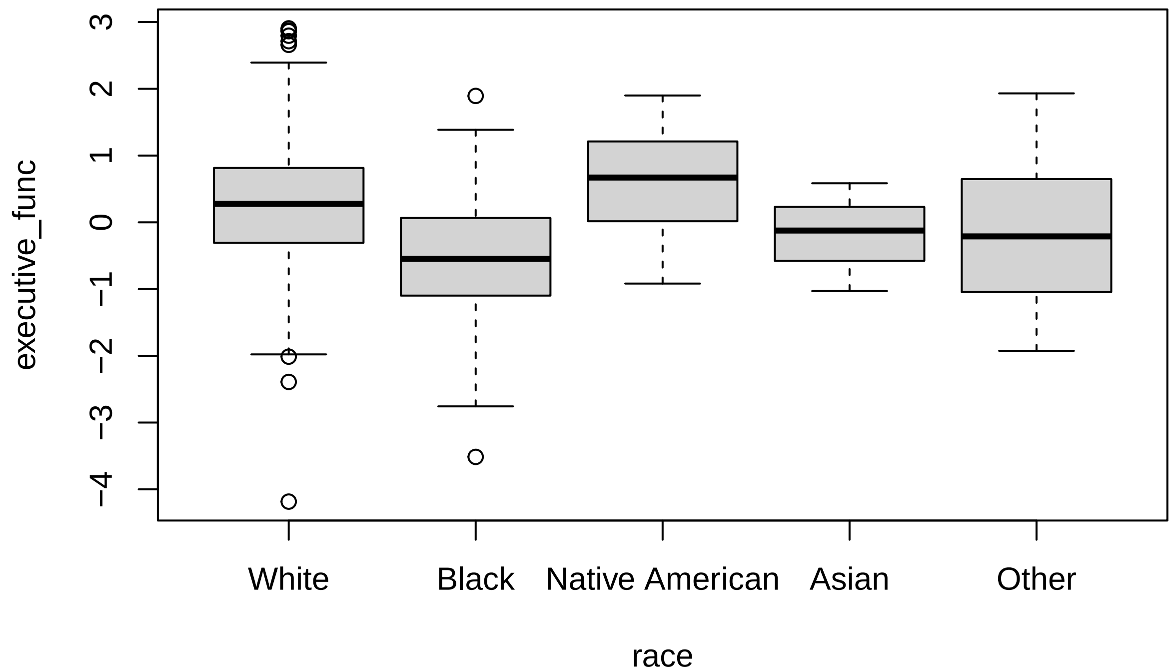
```
boxplot(composite_cognition ~ race, data=midus)
```




```
boxplot(episodic_memory ~ race, data=midus)
```




```
boxplot(executive_func ~ race, data=midus)
```



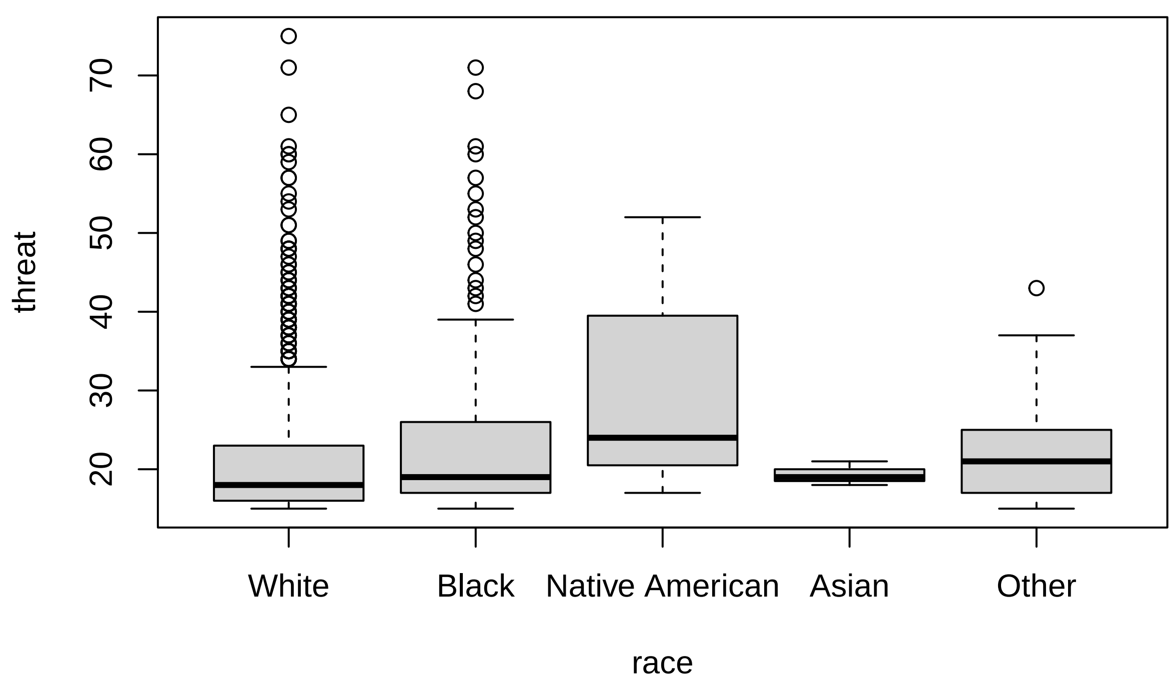

```
race_lme <- lmer(composite_cognition~ race + (1|family_id), data=midus)
summary(race_lme)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: composite_cognition ~ race + (1 | family_id)
##    Data: midus
##
## REML criterion at convergence: 2519.1
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.92707 -0.49788 -0.01158  0.46608  2.26216
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  family_id (Intercept) 0.4342   0.6589
##  Residual              0.3681   0.6067
## Number of obs: 979, groups:  family_id, 844
##
## Fixed effects:
##                     Estimate Std. Error t value
## (Intercept)          0.20166    0.03121   6.461
## raceBlack           -0.65717    0.18329  -3.585
## raceNative American  0.40882    0.25123   1.627
## raceAsian           -0.45386    0.51808  -0.876
```

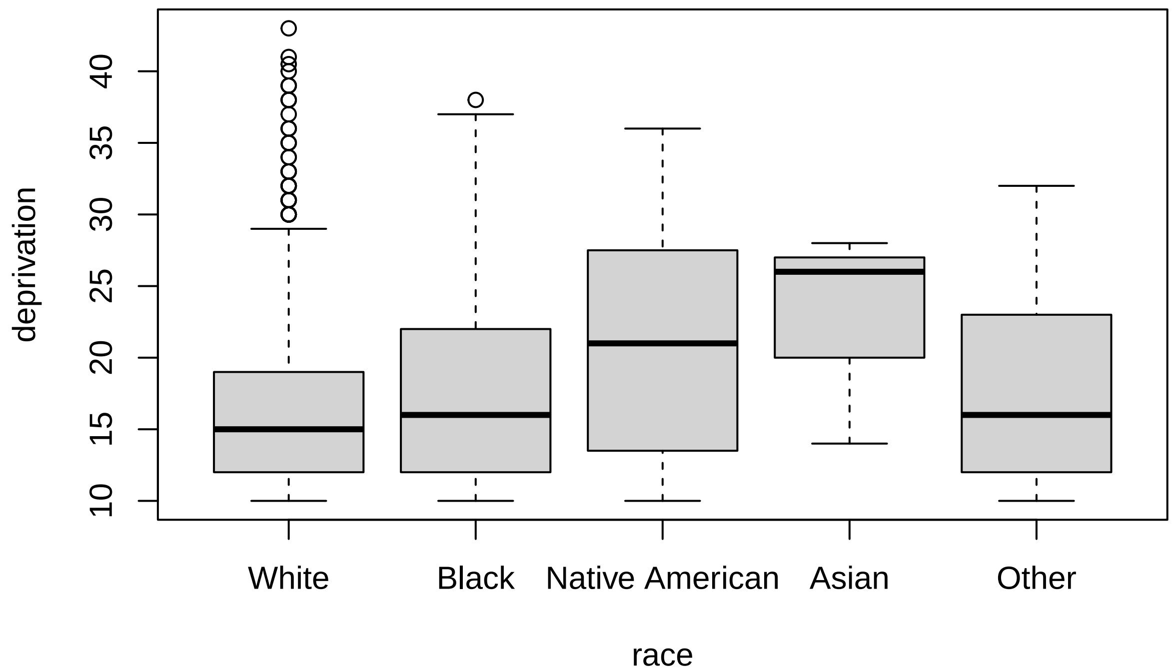
```
## raceOther           -0.41553    0.18378  -2.261
##
## Correlation of Fixed Effects:
##             (Intr) rcBlck rcNtvA racAsn
## raceBlack   -0.170
## racNtvAmrcn -0.107  0.018
## raceAsian   -0.060  0.010  0.006
## raceOther   -0.164  0.028  0.018  0.010
```

To evaluate confounding we first fit a boxplot. From the box plot we see median composite cognition score is different across races. We see that the highest composite cognition score is among those who identified as native american, the lowest median composite cognition score is among those who identified as black. To further to see if there is an association between race and composite cognition, we it a linear model. To note, the observations are not considered independent, but I just want to see if there is an association while disregarding grouping. . To evaluate the specific parameters, lets introduce a mixed effects model to have greater interpretability. To begin, the mixed effect model tells us that 54.35% of the variance is not explained by the fixed effects. This indicates the importance of including grouping by family as an effect. From the model we see that on average, composite cognition scores tend to be lower compared to those who identify as white with the exception of native americans. So we we conclude race is infact associated with composite cognition. Lets look at race vs our exposure(threat and deprivation).

```
boxplot(threat ~ race, data=midus)
```




```
boxplot(deprivation ~ race, data=midus)
```



The box plots indicate that median deprivation and threat scores are different across races.

Thus based on our results we conclude that race is a confounding factor to the association between childhood trauma and cognition.

Sex

```
t1 <-
midus %>%
  group_by(sex) %>%
  summarize(
    mean_composite=mean(composite_cognition))

t2 <-
midus %>%
  group_by(sex) %>%
  summarize(
    mean_episodic=mean(episodic_memory))

t3 <-
midus %>%
  group_by(sex) %>%
  summarize(
    mean_exec=mean(executive_func))
```

sex	mean_composite	sex	mean_episodic	sex	mean_exec
Male	0.0911089	Male	-0.2322856	Male	0.2422354
Female	0.1217412	Female	0.3017861	Female	0.0852037

```
knitr::kable(list(t1,t2,t3))
```

From the tables constructed we see there are significant differences in cognition scores across gender.

```
c1 <-
midus %>%
  na.omit() %>%
  group_by(sex) %>%
  summarize(
    mean_deprivation=mean(deprivation))

c2 <-
midus %>%
  na.omit() %>%
  group_by(sex) %>%
  summarize(
    mean_threat=mean(threat))


summary(lmer(deprivation ~ sex + (1|family_id), data=midus))
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: deprivation ~ sex + (1 | family_id)
##    Data: midus
##
## REML criterion at convergence: 6261.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.0087 -0.5055 -0.1966  0.3525  2.8383
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  family_id (Intercept) 20.48    4.526
##  Residual              16.46    4.057
## Number of obs: 979, groups:  family_id, 844
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  16.7903     0.2951   56.89
## sexFemale     0.3649     0.3922    0.93
##
## Correlation of Fixed Effects:
##           (Intr)
## sexFemale -0.721
```

sex	mean_deprivation
Male	16.60928
Female	17.24202

sex	mean_threat
Male	19.74231
Female	22.37262

```
summary(lmer(threat ~ sex + (1|family_id), data=midus))
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: threat ~ sex + (1 | family_id)
##    Data: midus
##
## REML criterion at convergence: 6856.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.5001 -0.3243 -0.1942  0.1409  3.5728
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  family_id (Intercept) 49.61    7.043
##  Residual              21.26    4.611
## Number of obs: 979, groups:  family_id, 844
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  19.9845     0.4053  49.305
## sexFemale     2.3100     0.5307   4.353
##
## Correlation of Fixed Effects:
##           (Intr)
## sexFemale -0.709
```

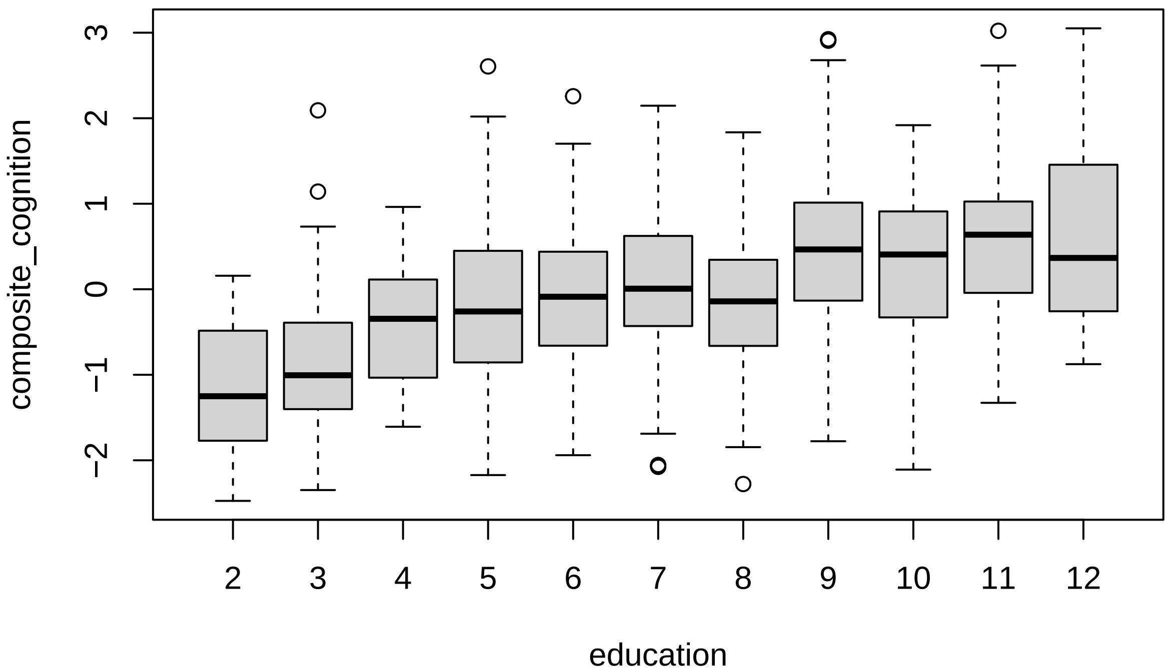
```
knitr::kable(list(c1,c2))
```

Though there is not a significant difference in mean deprivation score across genders, but there is a significant difference in mean threat scores across gender.

We conclude sex to be a confounder of the relationship between childhood trauma and cognition.

Education

```
boxplot(composite_cognition ~ education, data=midus)
```

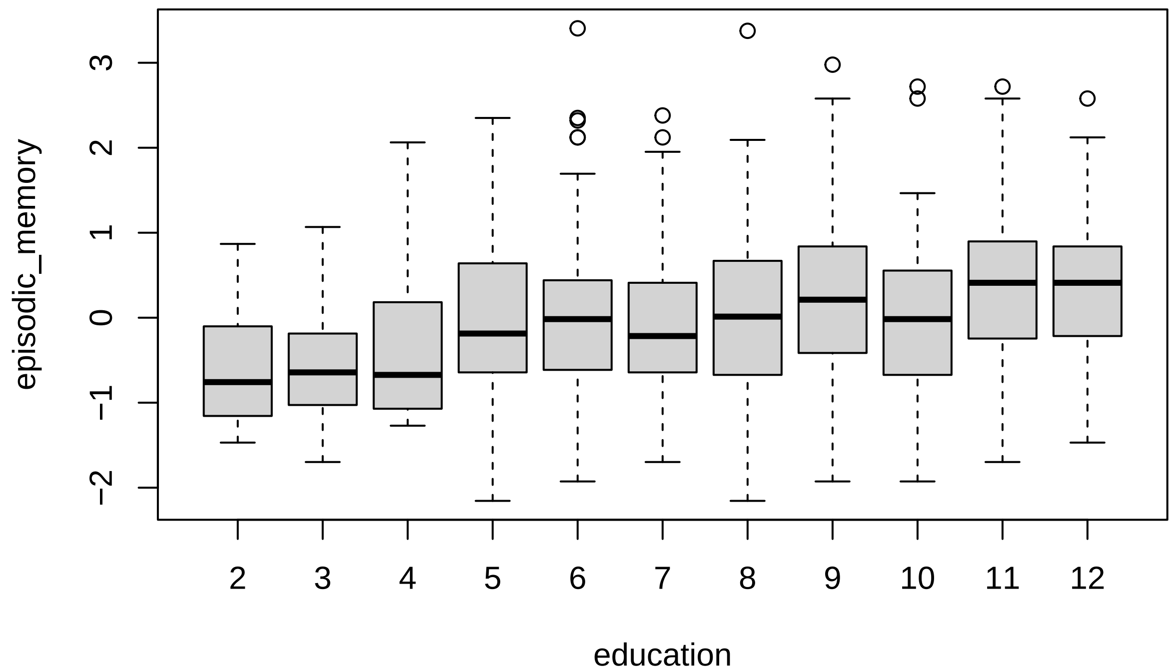



```
education_lm <- lmer(composite_cognition ~ education + (1|family_id), data=midus)
summary(education_lm)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: composite_cognition ~ education + (1 | family_id)
##    Data: midus
##
## REML criterion at convergence: 2405.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.15087 -0.51279 -0.02273  0.46110  2.26212
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  family_id (Intercept) 0.3595   0.5996
##  Residual              0.3463   0.5884
## Number of obs: 976, groups:  family_id, 842
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  -1.0916     0.3335  -3.273
## education3    0.4663     0.3693   1.263
## education4    0.7092     0.4431   1.600
## education5    0.9842     0.3387   2.906
```

```
## education6    1.0302     0.3399   3.031
## education7    1.1971     0.3572   3.351
## education8    1.0894     0.3458   3.150
## education9    1.5728     0.3381   4.651
## education10   1.3470     0.3542   3.803
## education11   1.6696     0.3390   4.924
## education12   1.7157     0.3559   4.820
##
## Correlation of Fixed Effects:
##             (Intr) edctn3 edctn4 edctn5 edctn6 edctn7 edctn8 edctn9 edct10
## education3  -0.903
## education4  -0.752  0.680
## education5  -0.984  0.890  0.743
## education6  -0.981  0.886  0.739  0.967
## education7  -0.934  0.846  0.703  0.920  0.916
## education8  -0.964  0.871  0.738  0.951  0.948  0.902
## education9  -0.986  0.891  0.742  0.971  0.967  0.921  0.951
## education10 -0.941  0.850  0.708  0.927  0.925  0.879  0.908  0.930
## education11 -0.979  0.884  0.737  0.964  0.961  0.914  0.944  0.967  0.923
## education12 -0.937  0.846  0.705  0.922  0.919  0.876  0.903  0.924  0.883
##             edct11
## education3
## education4
## education5
## education6
## education7
## education8
## education9
## education10
## education11
## education12  0.919
```

```
boxplot(episodic_memory ~ education, data=midus)
```

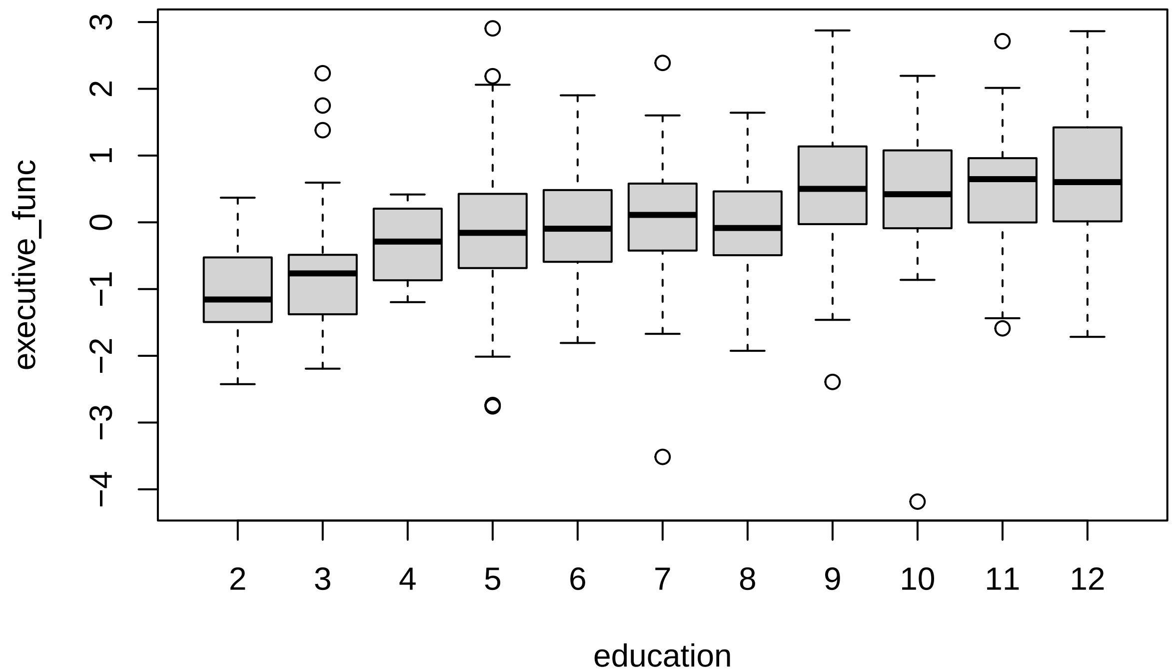



```r
education_lm2 <- lmer(episodic_memory ~ education + (1|family_id), data=midus)
summary(education_lm2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: episodic_memory ~ education + (1 | family_id)
##    Data: midus
##
## REML criterion at convergence: 2535.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1725 -0.5787 -0.0425  0.5561  3.3751
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  family_id (Intercept) 0.2576   0.5075
##  Residual              0.5295   0.7276
## Number of obs: 976, groups:  family_id, 842
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) -0.41759    0.35864  -1.164
## education3  -0.14920    0.39655  -0.376
## education4   0.05993    0.48375   0.124
## education5   0.36704    0.36418   1.008
```

```
## education6   0.42835    0.36552   1.172
## education7   0.38140    0.38405   0.993
## education8   0.46597    0.37207   1.252
## education9   0.65498    0.36360   1.801
## education10  0.41800    0.38117   1.097
## education11  0.77475    0.36536   2.121
## education12  0.70330    0.38276   1.837
##
## Correlation of Fixed Effects:
##             (Intr) edctn3 edctn4 edctn5 edctn6 edctn7 edctn8 edctn9 edct10
## education3  -0.904
## education4  -0.741  0.671
## education5  -0.985  0.891  0.732
## education6  -0.981  0.887  0.728  0.967
## education7  -0.934  0.846  0.692  0.920  0.916
## education8  -0.964  0.872  0.722  0.950  0.947  0.901
## education9  -0.986  0.892  0.731  0.972  0.968  0.921  0.951
## education10 -0.941  0.851  0.698  0.927  0.924  0.879  0.907  0.929
## education11 -0.979  0.885  0.726  0.964  0.961  0.914  0.944  0.967  0.922
## education12 -0.937  0.847  0.695  0.923  0.919  0.876  0.903  0.925  0.882
##             edct11
## education3
## education4
## education5
## education6
## education7
## education8
## education9
## education10
## education11
## education12  0.919
```

```
boxplot(executive_func ~ education, data=midus)
```

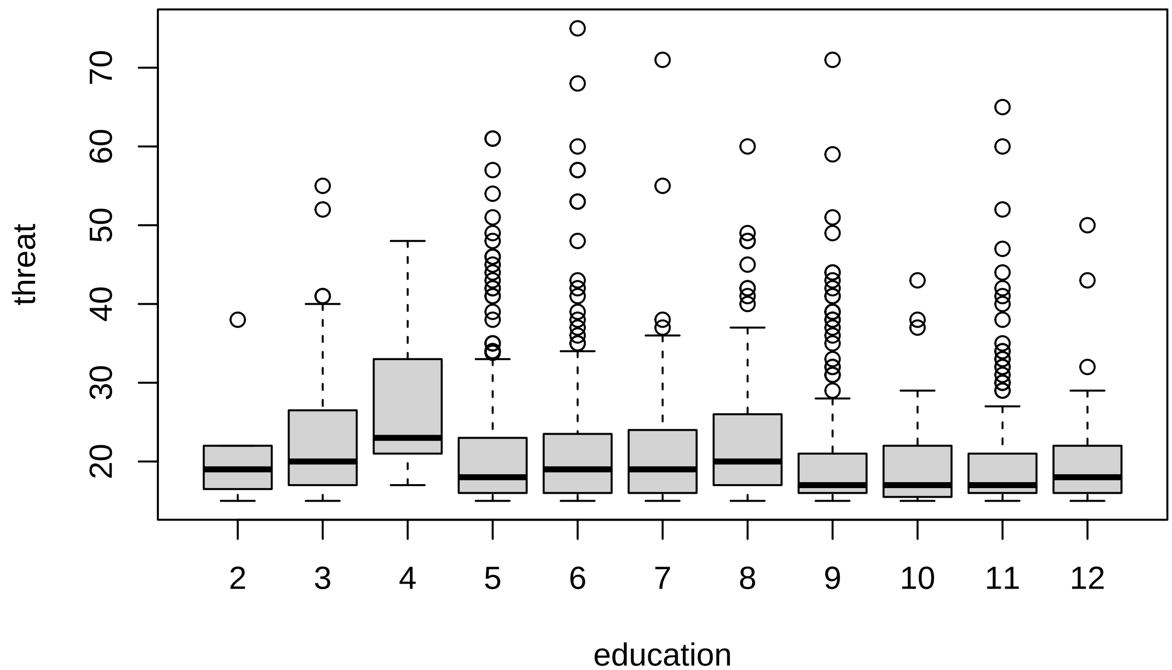



```r
education_lm3 <- lmer(executive_func ~ education + (1|family_id), data=midus)
summary(education_lm3)
```

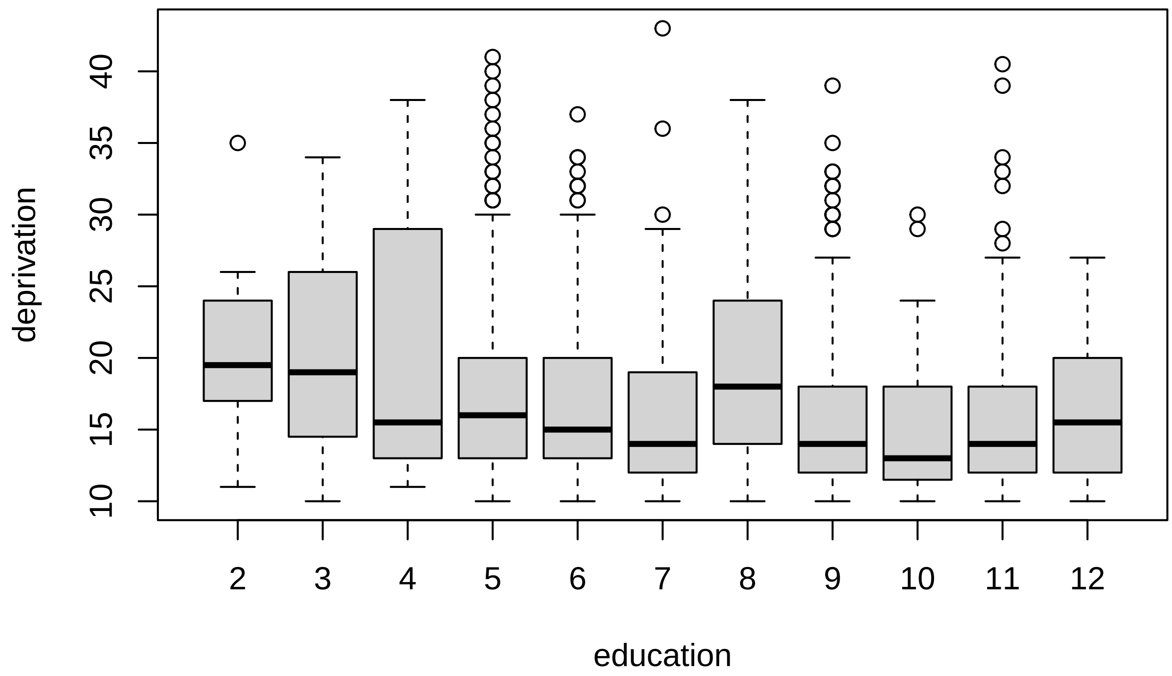
```
## Linear mixed model fit by REML ['lmerMod']
## Formula: executive_func ~ education + (1 | family_id)
##    Data: midus
##
## REML criterion at convergence: 2324.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.6588 -0.4646 -0.0319  0.4353  2.3919
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  family_id (Intercept) 0.3698   0.6081
##  Residual              0.2863   0.5351
## Number of obs: 976, groups:  family_id, 842
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  -0.9334     0.3187  -2.929
## education3    0.4474     0.3531   1.267
## education4    0.5606     0.4208   1.332
## education5    0.9057     0.3237   2.798
```

```
## education6    0.9029     0.3249   2.779
## education7    1.1409     0.3414   3.342
## education8    0.9850     0.3304   2.981
## education9    1.4555     0.3231   4.504
## education10   1.2357     0.3384   3.652
## education11   1.4961     0.3237   4.621
## education12   1.5829     0.3401   4.654
##
## Correlation of Fixed Effects:
##             (Intr) edctn3 edctn4 edctn5 edctn6 edctn7 edctn8 edctn9 edct10
## education3  -0.903
## education4  -0.757  0.684
## education5  -0.984  0.890  0.748
## education6  -0.981  0.885  0.743  0.967
## education7  -0.933  0.846  0.707  0.919  0.916
## education8  -0.965  0.871  0.744  0.952  0.949  0.903
## education9  -0.986  0.891  0.747  0.971  0.967  0.921  0.951
## education10 -0.942  0.850  0.713  0.927  0.925  0.879  0.908  0.930
## education11 -0.979  0.884  0.741  0.964  0.961  0.914  0.944  0.967  0.923
## education12 -0.937  0.845  0.709  0.922  0.919  0.877  0.903  0.924  0.884
##             edct11
## education3
## education4
## education5
## education6
## education7
## education8
## education9
## education10
## education11
## education12  0.920
```

```
boxplot(threat ~ education, data=midus)
```




```
boxplot(deprivation ~ education, data=midus)
```

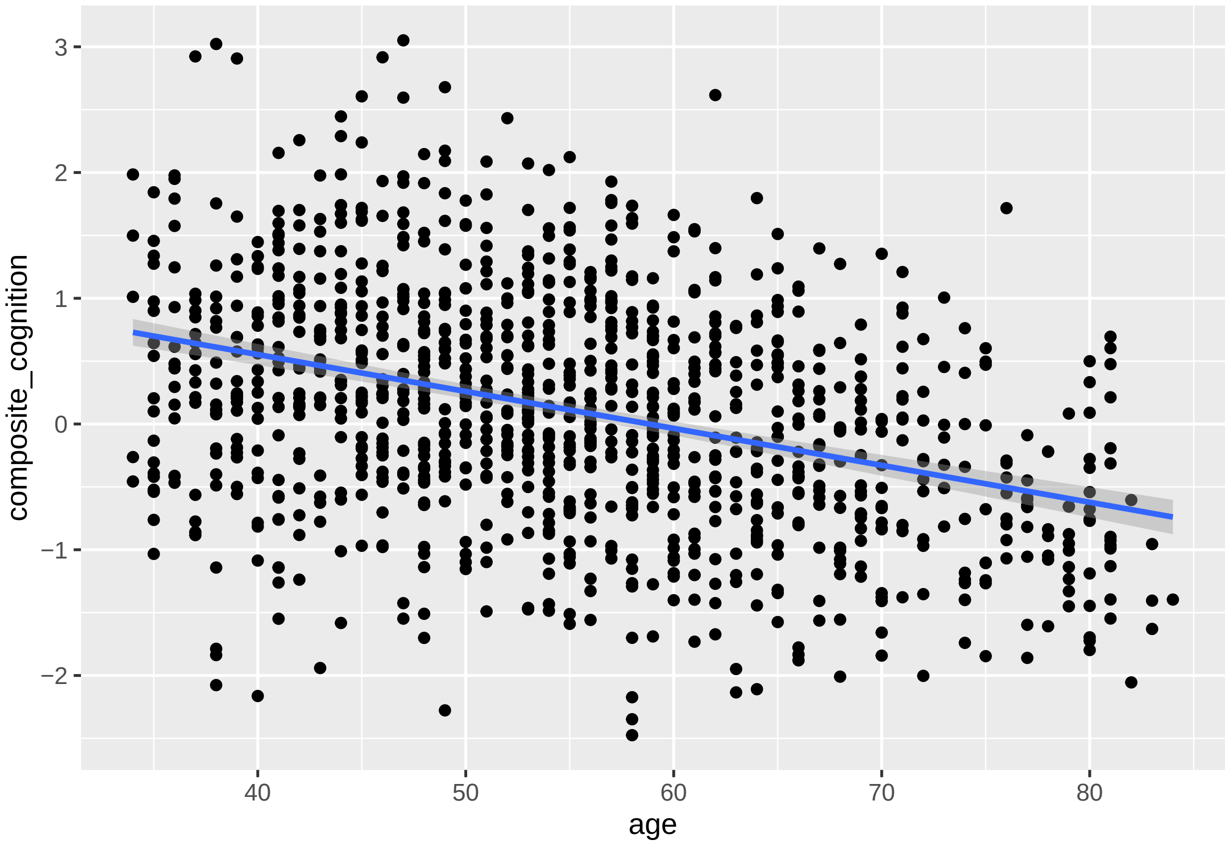


Age

```
par(mfrow=c(3,1))
ggplot(midus,aes(x=age,y=composite_cognition)) + geom_point() + geom_smooth(method=lm)
```

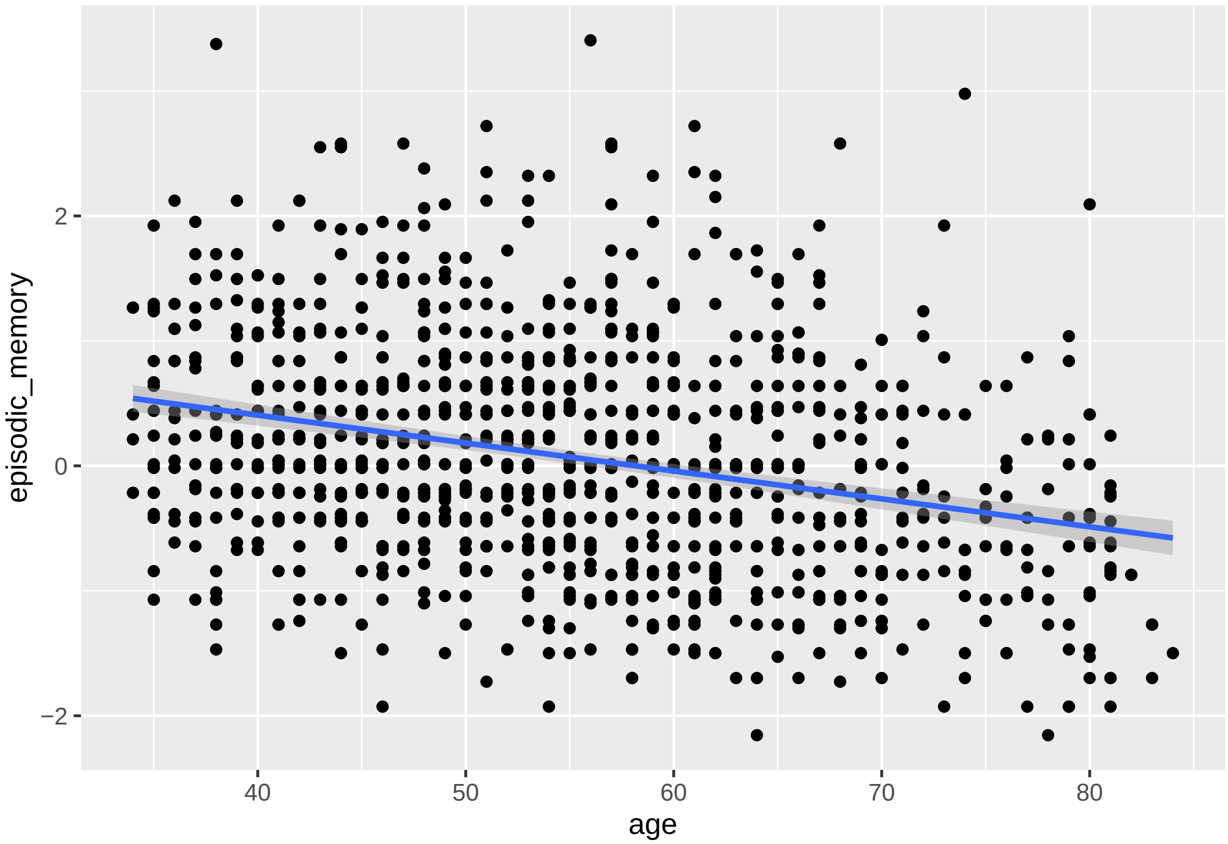
```
## Don't know how to automatically pick scale for object of type haven_labelled/vctrs_vctr/double. Defa
## Don't know how to automatically pick scale for object of type haven_labelled/vctrs_vctr/double. Defa
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
ggplot(midus,aes(x=age,y=episodic_memory)) + geom_point() + geom_smooth(method=lm)
```

```
## Don't know how to automatically pick scale for object of type haven_labelled/vctrs_vctr/double. Defau

## Don't know how to automatically pick scale for object of type haven_labelled/vctrs_vctr/double. Defau

## `geom_smooth()` using formula 'y ~ x'
```

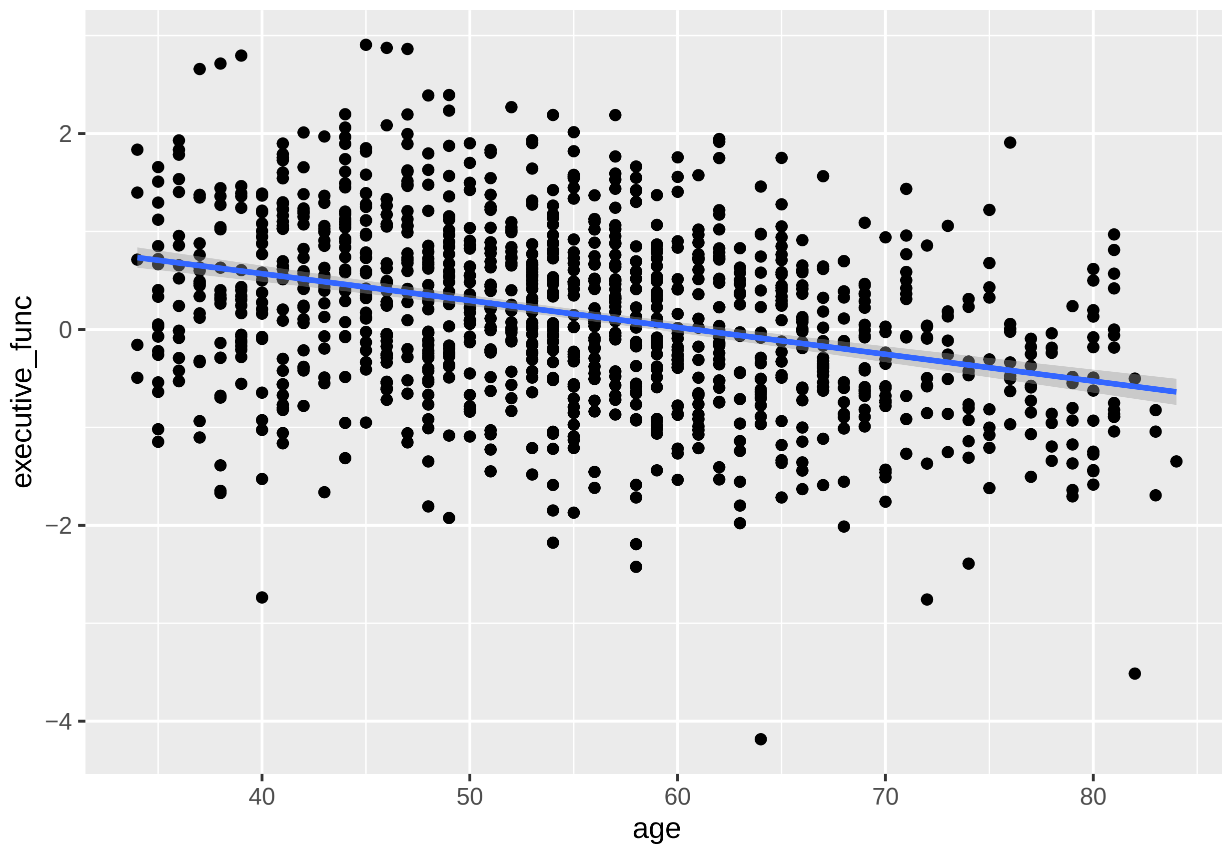



```
ggplot(midus,aes(x=age,y=executive_func)) + geom_point() + geom_smooth(method=lm)
```

```
## Don't know how to automatically pick scale for object of type haven_labelled/vctrs_vctr/double. Defau
```

```
## Don't know how to automatically pick scale for object of type haven_labelled/vctrs_vctr/double. Defau
```

```
## `geom_smooth()` using formula 'y ~ x'
```

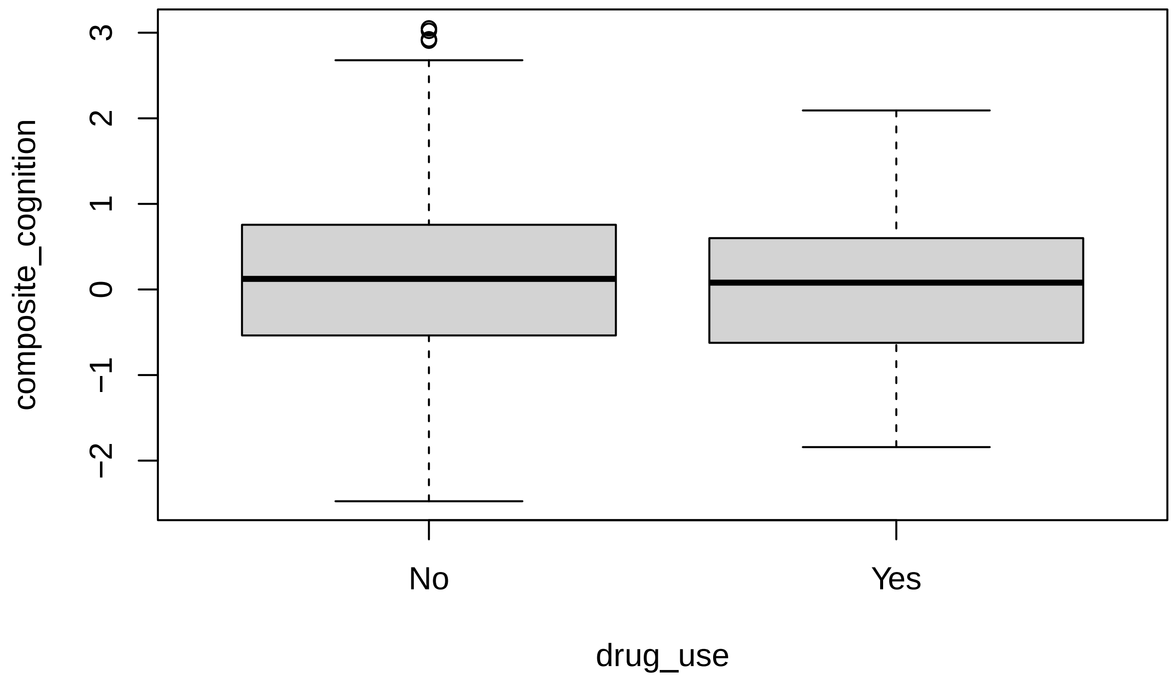


Self Administered Substance Abuse

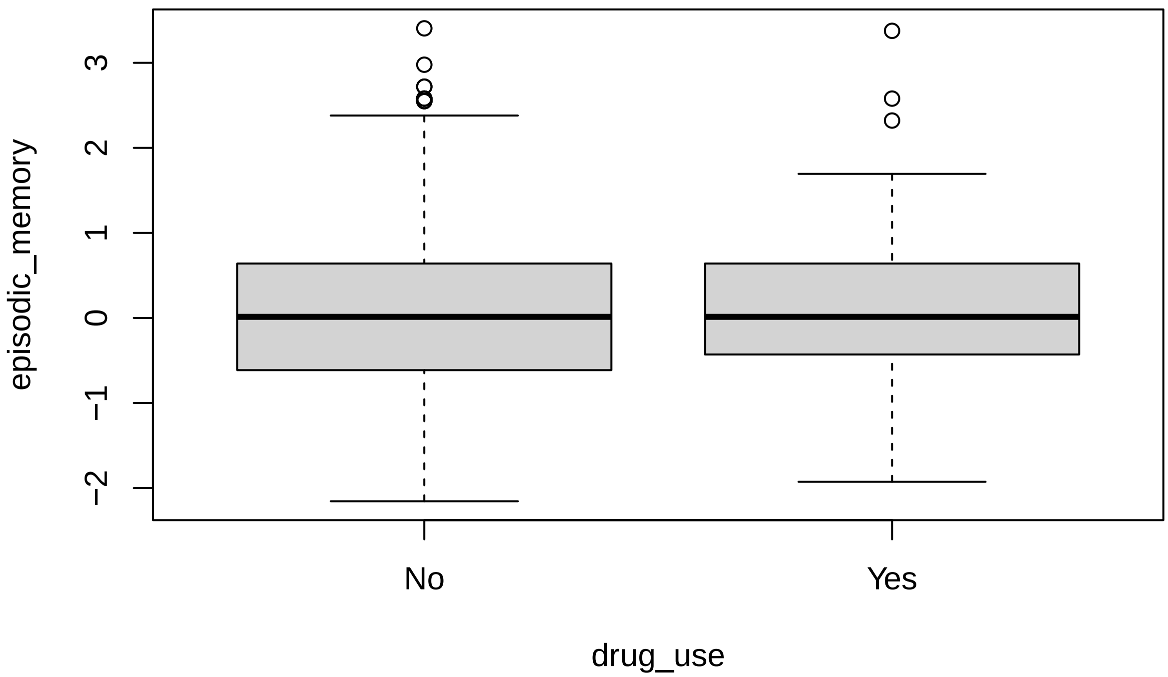
```
midus %>%
  group_by(drug_use) %>%
  summarize(n_obs=n()) %>%
  knitr::kable()
```

drug_use	n_obs
No	997
Yes	104

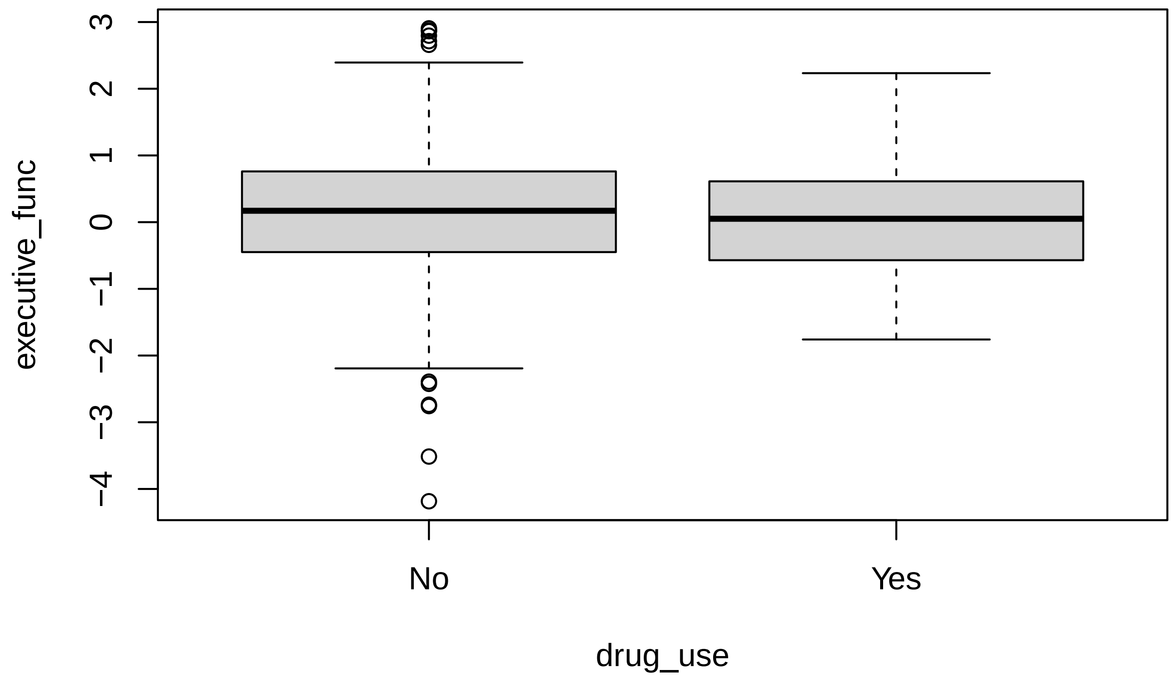
```
boxplot(composite_cognition ~ drug_use, data=midus)
```




```
boxplot(episodic_memory ~ drug_use, data=midus)
```




```
boxplot(executive_func ~ drug_use, data=midus)
```



From the box plots we dont see major differences in average cognition scores between those who did or did not indicate substance use.

```
midus %>%
  group_by(tranquilizer) %>%
  summarize(n_obs=n()) %>%
  knitr::kable()
```

tranquilizer	n_obs
No	1064
Yes	34
NA	3

```
# n=34 for yes

midus %>%
  group_by(stimulants) %>%
  summarize(n_obs=n()) %>%
  knitr::kable()
```

stimulants	n_obs
No	1073
Yes	25

stimulants	n_obs
NA	3

```r
#n = 25

midus %>%
  group_by(inhallants) %>%
  summarize(n_obs=n()) %>%
  knitr::kable()
```

inhallants	n_obs
No	1083
Yes	14
NA	4

```r
# n = 14

midus %>%
  group_by(depressants) %>%
  summarize(n_obs=n()) %>%
  knitr::kable()
```

depressants	n_obs
No	1048
Yes	50
NA	3

```r
#n=49

midus %>%
  group_by(marijuana) %>%
  summarize(n_obs=n()) %>%
  knitr::kable()
```

marijuana	n_obs
No	1094
Yes	1
NA	6

```r
# n = 1

midus %>%
  group_by(cocaine) %>%
  summarize(n_obs=n()) %>%
  knitr::kable()
```

cocaine	n_obs
No	1085
Yes	13
NA	3

```
# n = 13

midus %>%
  group_by(hallucinogens) %>%
  summarize(n_obs=n()) %>%
  knitr::kable()
```

hallucinogens	n_obs
No	1085
Yes	13
NA	3

```
# n = 13

midus %>%
  group_by(heroin) %>%
  summarize(n_obs=n()) %>%
  knitr::kable()
```

heroin	n_obs
Yes	1
No	1096
NA	4

```
# n = 1
```