BRSM_project

Padakanti Srijith

06/05/2023

Contents

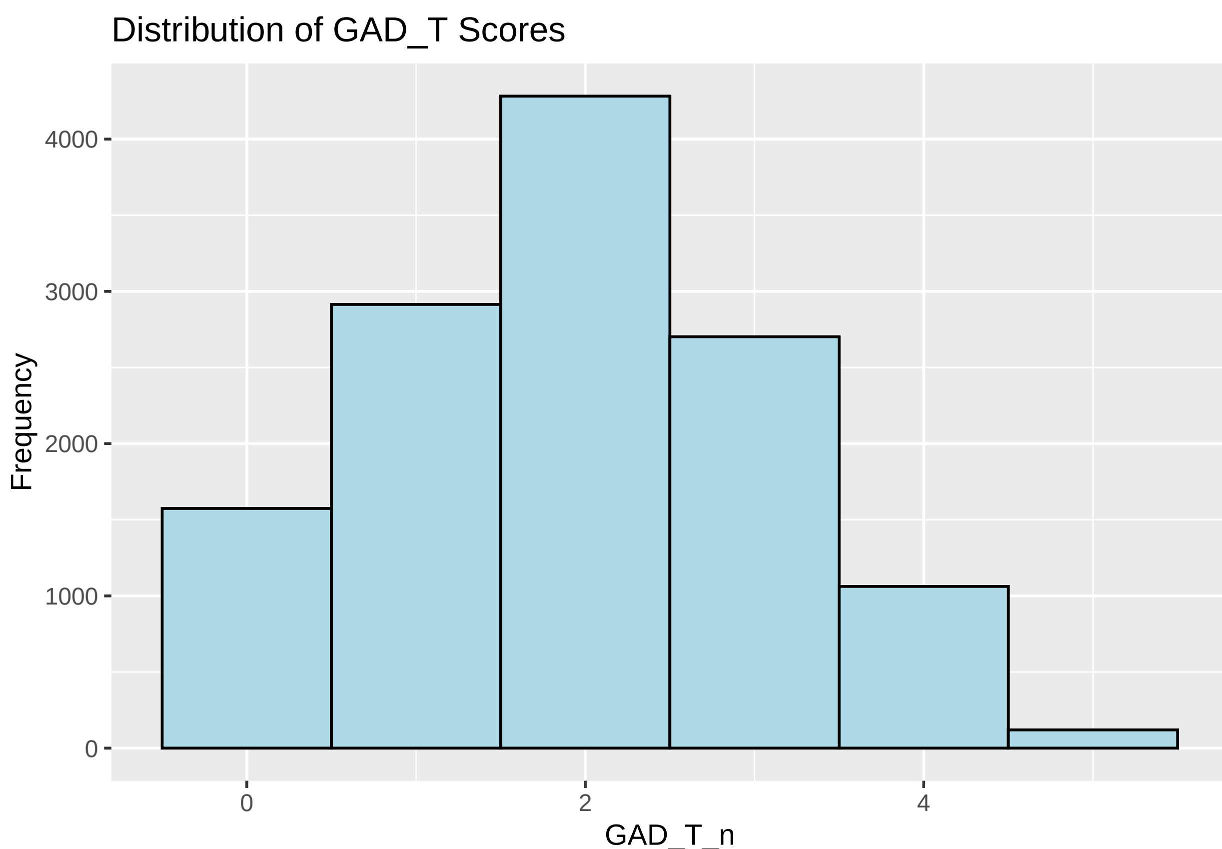
```r
data <- read.csv('cleanedDataset.csv')
data["Gender"][data["Gender"] == "Male"] <- 0
data["Gender"][data["Gender"] == "Female"] <- 1

data$Gender <- as.integer(data$Gender)
```

```
## Warning: NAs introduced by coercion
```

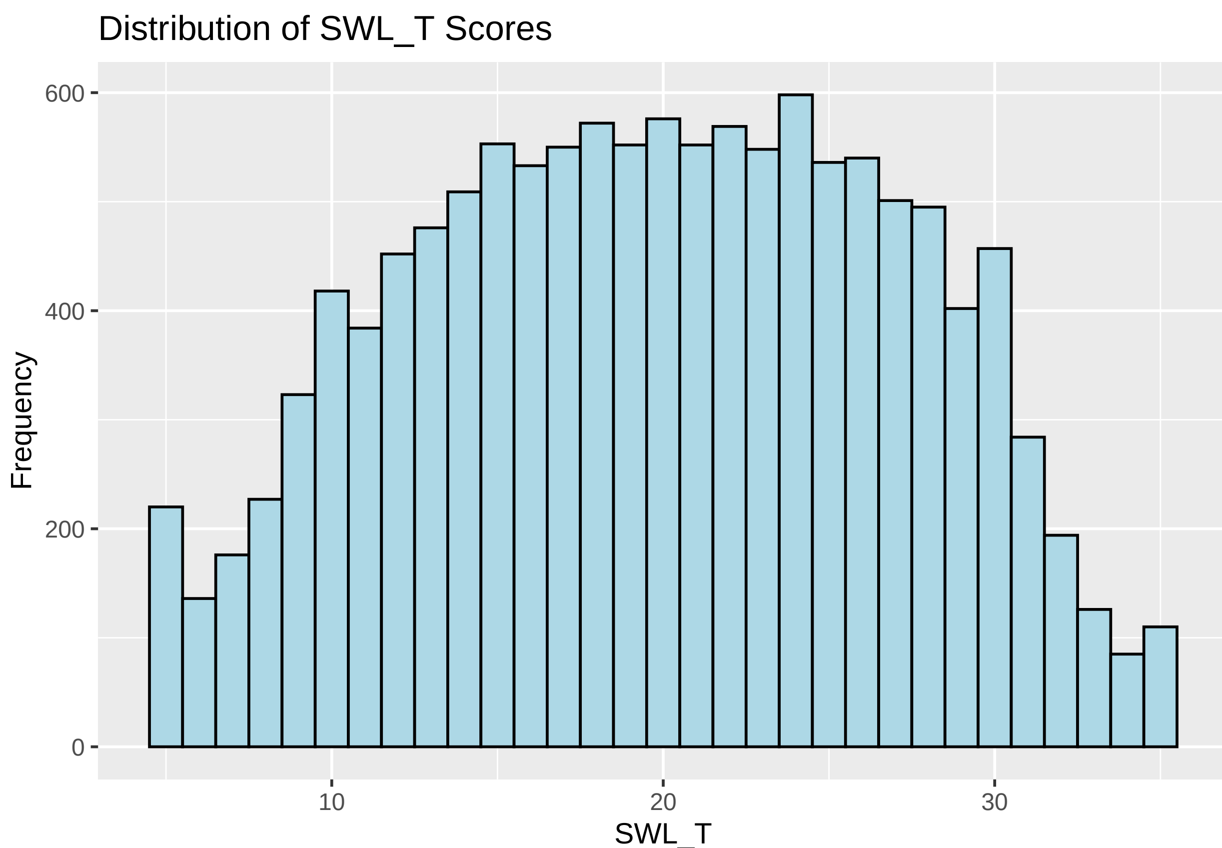
```r
data["Gender"][is.na(data["Gender"])] <- 2
#normalizing GAD and SPIN
data["GAD_T_n"] <- sqrt(data["GAD_T"])
data["SPIN_T_n"] <- sqrt(data["SPIN_T"])


data["birth_res"][data["birth_res"] == "Same"] <- 0
data["birth_res"][data["birth_res"] == "Different"] <- 1
data$birth_res <- as.integer(data$birth_res)

data["Work"][data["Work"] == "Student"] <- 0
data["Work"][data["Work"] == "Unemployed"] <- 1
data["Work"][data["Work"] == "Employed"] <- 2
data$Work <- as.integer(data$Work)
data["Work"][is.na(data["Work"])] <- 1

data["Degree"][data["Degree"] == "Degree"] <- 1
data["Degree"][data["Degree"] == "None"] <- 0
data$Degree <- as.integer(data$Degree)
data["Degree"][is.na(data["Degree"])] <- 0

data["Platform"][data["Platform"] == "PC"] <- 0
data["Platform"][data["Platform"] == "Console (PS, Xbox, ...)"] <- 1
data["Platform"][data["Platform"] == "Smartphone / Tablet"] <- 2
data$Platform <- as.integer(data$Platform)
data["Platform"][is.na(data["Platform"])] <- 1

data["does_stream"] = data["Game"]
data["does_stream"][data["streams"] == 0] <- "No"
data["does_stream"][data["streams"] != 0] <- "Yes"
```

```r
data["does_stream"][is.na(data["streams"])] <- "No"
data["mean_hours"][is.na(data["mean_hours"])] <- "Below"
data["Hours.streams"][is.na(data["Hours.streams"])] <- 10
```

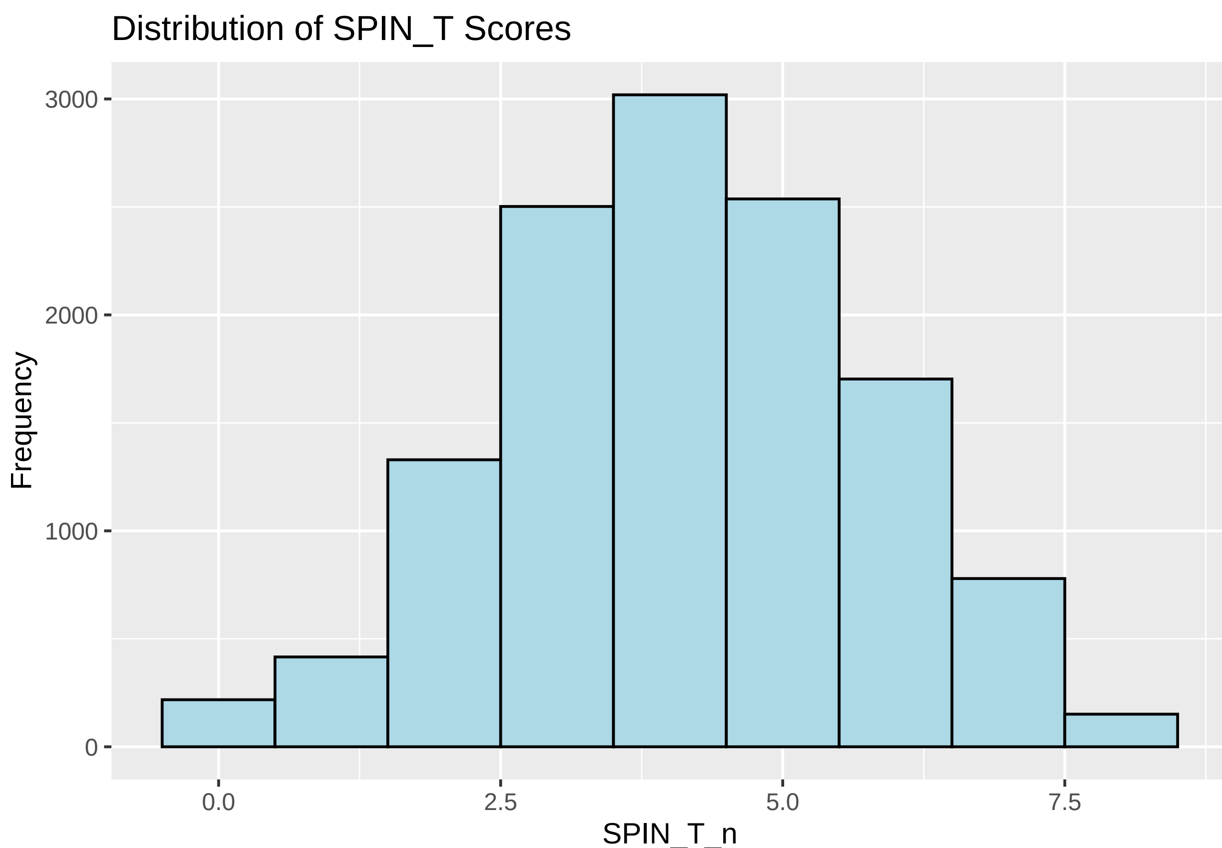
```r
library(ggplot2)
ggplot(data, aes(x = GAD_T_n)) +
  geom_histogram(binwidth = 1, fill = "lightblue", color = "black") +
  labs(x = "GAD_T_n", y = "Frequency") +
  ggtitle("Distribution of GAD_T Scores")
```




```r
ggplot(data, aes(x = SWL_T)) +
  geom_histogram(binwidth = 1, fill = "lightblue", color = "black") +
  labs(x = "SWL_T", y = "Frequency") +
  ggtitle("Distribution of SWL_T Scores")
```




```
ggplot(data, aes(x = SPIN_T_n)) +
  geom_histogram(binwidth = 1, fill = "lightblue", color = "black") +
  labs(x = "SPIN_T_n", y = "Frequency") +
  ggtitle("Distribution of SPIN_T Scores")
```



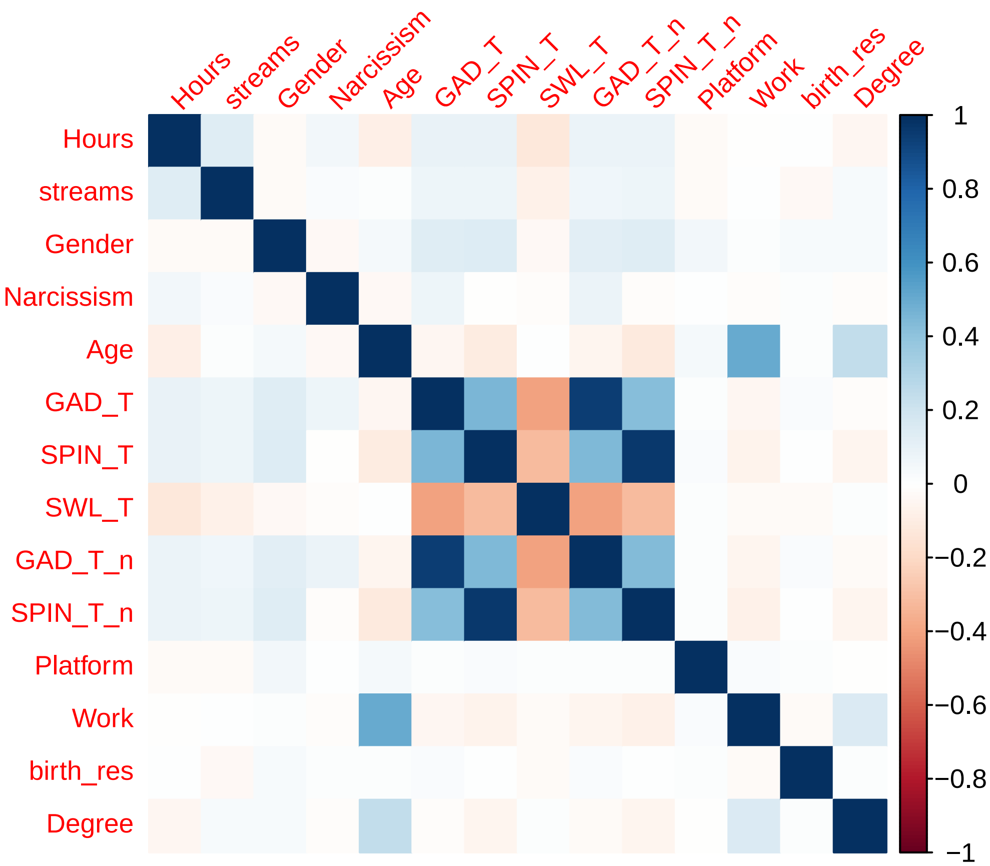

```
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```
data_sub <- subset(data, select = c("Hours", "streams", "Gender", "Narcissism", "Age", "GAD_T", "SPIN_T
corr_matrix <- cor(data_sub)
str(data_sub)
```

```
## 'data.frame':    12654 obs. of  14 variables:
##  $ Hours     : int  15 8 0 20 20 4 30 2 25 14 ...
##  $ streams   : int  0 2 0 5 1 0 8 0 0 0 ...
##  $ Gender    : num  0 0 1 0 0 0 0 1 1 1 ...
##  $ Narcissism: int  1 1 4 2 1 2 2 1 1 1 ...
##  $ Age       : int  25 41 32 28 19 24 29 23 27 21 ...
##  $ GAD_T     : int  1 8 8 0 14 1 0 12 10 19 ...
##  $ SPIN_T    : int  5 33 31 11 13 13 26 55 26 6 ...
##  $ SWL_T     : int  23 16 17 17 14 17 16 12 13 27 ...
##  $ GAD_T_n   : num  1 2.83 2.83 0 3.74 ...
##  $ SPIN_T_n  : num  2.24 5.74 5.57 3.32 3.61 ...
##  $ Platform  : int  1 0 0 0 1 1 0 1 0 0 ...
##  $ Work      : int  1 1 2 2 2 2 2 2 2 0 ...
##  $ birth_res : int  0 0 0 0 1 0 0 0 0 0 ...
##  $ Degree    : int  1 1 1 1 1 1 1 1 1 1 ...
```

```
# Visualize the correlation matrix using a heatmap
corrplot(corr_matrix, method = "color",
         tl.cex = 0.8, tl.srt = 45)
```




```
data$streams_group <- cut(data$streams, breaks=c(-1, 10, 20, 30, Inf), labels=c("0-10", "10-20", "20-30
data["streams_group"][is.na(data["streams_group"])] <- "0-10"

data$age_group <- cut(data$Age, breaks=c(17, 25, 35, 50, Inf), labels=c("18-25", "26-35", "36-50", "50+
data["age_group"][is.na(data["age_group"])] <- "18-25"
```

```
# Gamers those who are unemployed and do not have a degree are less satisfied with life
  # Dependent varaible: SWL, GAD, SPIN (continuous)
  # Independent variable: Works (categorical), Degree (categorical)
# TEST: Factorial ANOVA

# Load necessary library
library(multcomp)
```

```
## Loading required package: mvtnorm

## Loading required package: survival

## Loading required package: TH.data
```

```
## Loading required package: MASS

##
## Attaching package: 'TH.data'

## The following object is masked from 'package:MASS':
##
##     geyser
```

```
# Perform ANOVA
hypo.2.1 <- aov(SWL_T ~ interaction(Work , Degree), data)
summary(hypo.2.1)
```

```
##                                Df Sum Sq Mean Sq F value Pr(>F)
## interaction(Work, Degree)     5  40197    8039   163.9 <2e-16 ***
## Residuals                 12648 620297      49
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Perform pairwise t-tests with Bonferroni correction
pairwise.t.test(data$SWL_T, interaction(data$Degree, data$Work), p.adjust.method = "bonferroni")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  data$SWL_T and interaction(data$Degree, data$Work)
##
##     0.0     1.0     0.1     1.1     0.2
## 1.0 1.000   -       -       -       -
## 0.1 < 2e-16 < 2e-16 -       -       -
## 1.1 < 2e-16 < 2e-16 0.045   -       -
## 0.2 1.000   1.000   1.0e-10 1.3e-08 -
## 1.2 0.042   0.022   < 2e-16 < 2e-16 1.000
##
## P value adjustment method: bonferroni
```

```
hypo.2.2 <- aov(GAD_T_n ~ interaction(Work , Degree), data)
summary(hypo.2.2)
```

```
##                                Df Sum Sq Mean Sq F value Pr(>F)
## interaction(Work, Degree)     5    189   37.83    30.2 <2e-16 ***
## Residuals                 12648  15845    1.25
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Perform pairwise t-tests with Bonferroni correction
pairwise.t.test(data$GAD_T_n, interaction(data$Degree, data$Work), p.adjust.method = "bonferroni")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
```

```
## data:  data$GAD_T_n and interaction(data$Degree, data$Work)
##
##     0.0     1.0     0.1     1.1     0.2
## 1.0 1.00000 -       -       -       -
## 0.1 0.00272 0.00024 -       -       -
## 1.1 0.00151 1.5e-09 0.39646 -       -
## 0.2 0.43452 0.92518 0.00048 0.01058 -
## 1.2 1.5e-09 3.2e-13 2.9e-08 < 2e-16 1.00000
##
## P value adjustment method: bonferroni
```

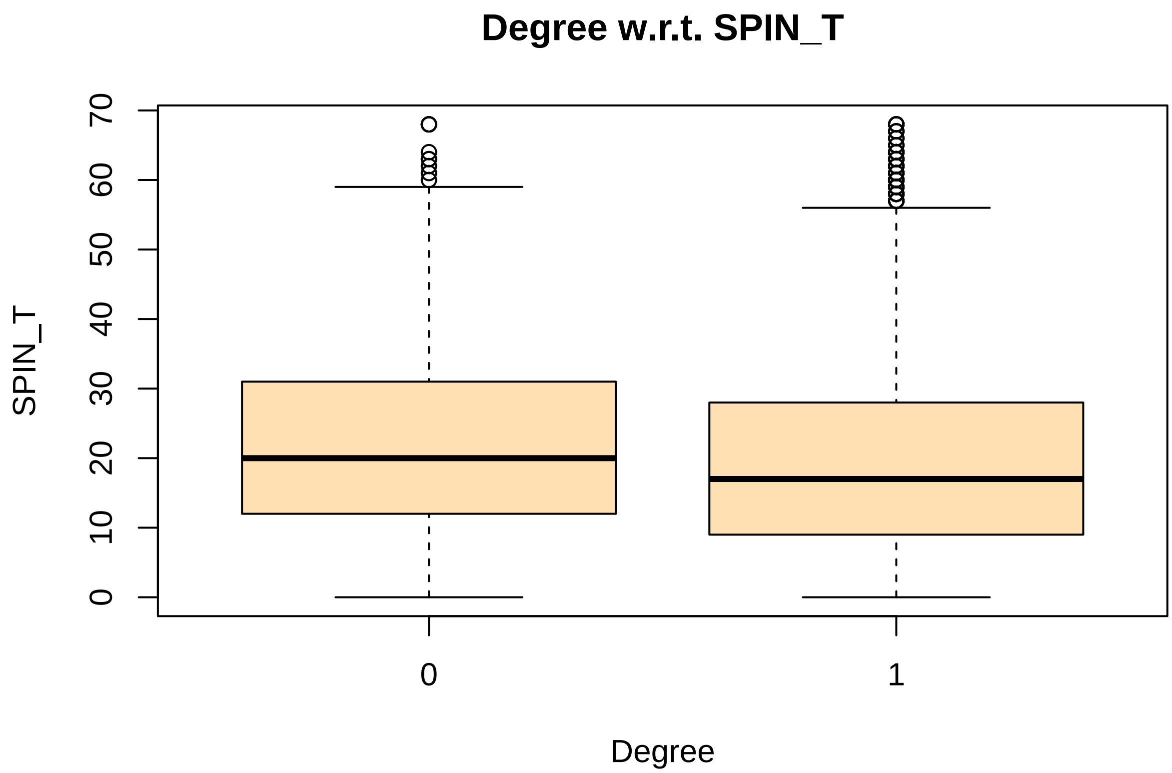
```
hypo.2.3 <- aov(SPIN_T_n ~ interaction(Work , Degree), data)
summary(hypo.2.3)
```

```
##                               Df Sum Sq Mean Sq F value Pr(>F)
## interaction(Work, Degree)     5    529  105.88   42.46 <2e-16 ***
## Residuals                 12648  31541    2.49
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

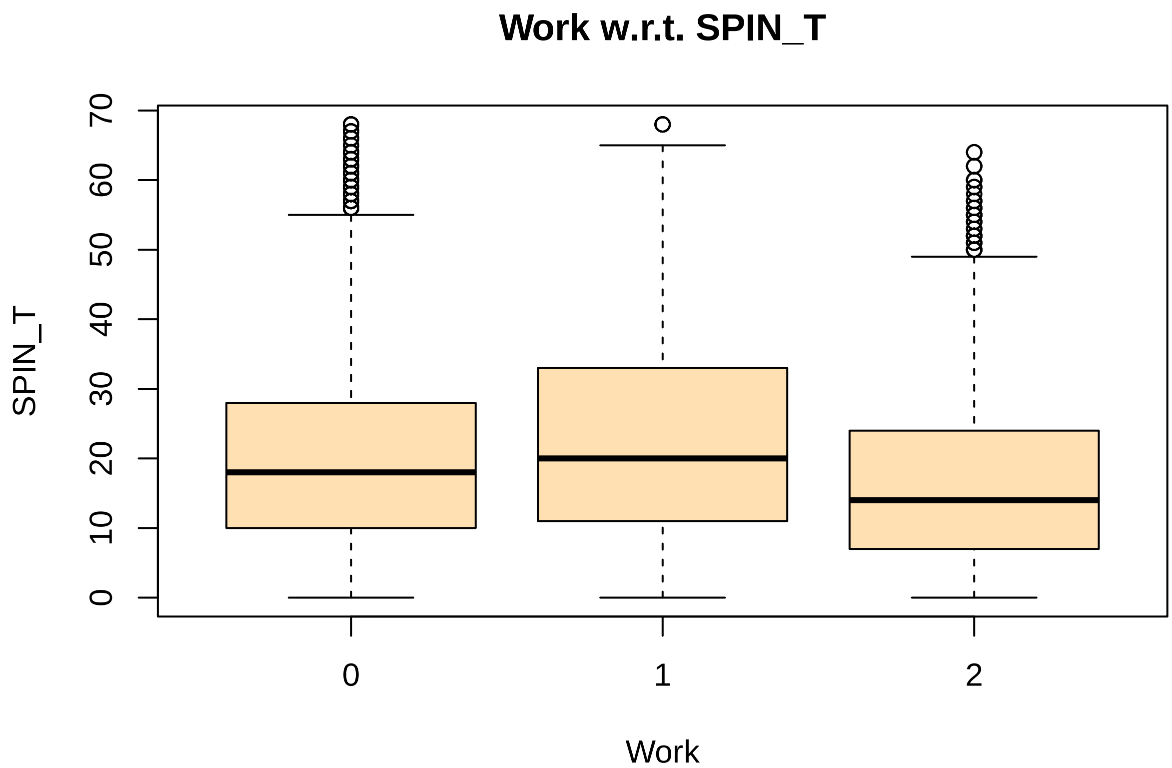
```
# Perform pairwise t-tests with Bonferroni correction
pairwise.t.test(data$SPIN_T_n, interaction(data$Degree, data$Work), p.adjust.method = "bonferroni")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  data$SPIN_T_n and interaction(data$Degree, data$Work)
##
##     0.0     1.0     0.1     1.1     0.2
## 1.0 6.2e-06 -       -       -       -
## 0.1 0.7711  0.0060  -       -       -
## 1.1 1.0000  6.7e-09 1.0000  -       -
## 0.2 0.0387  0.9293  0.0035  0.0130  -
## 1.2 < 2e-16 < 2e-16 2.3e-07 < 2e-16 1.0000
##
## P value adjustment method: bonferroni
```

```
boxplot(SPIN_T ~ Degree, data = data,
        main = "Degree w.r.t. SPIN_T",
        col = c("#FFE0B2"),
        xlab = "Degree",
        ylab = "SPIN_T")
```




```
boxplot(SPIN_T ~ Work, data = data,
        main = "Work w.r.t. SPIN_T",
        col = c("#FFE0B2"),
        xlab = "Work",
        ylab = "SPIN_T")
```




```
# Gamers who are young are more narcissistic than the rest
# Dependent varaible: Narcissism (continuous)
# Independent variable: age_group (categorical)
# TEST: One-way Independent ANOVA
data$age_group <- cut(data$Age, breaks=c(17, 25, 35, 50, Inf), labels=c("18-25", "26-35", "36-50", "50+

model <- aov(Narcissism ~ age_group, data=data)
anova(model)
```

```
## Analysis of Variance Table
##
## Response: Narcissism
##               Df  Sum Sq Mean Sq F value    Pr(>F)
## age_group      3    22.4  7.4579  6.6444 0.0001765 ***
## Residuals 12650 14198.8  1.1224
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(model)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Narcissism ~ age_group, data = data)
##
```

```
## $age_group
##                   diff        lwr         upr     p adj
## 26-35-18-25 -0.14097352 -0.2276733 -0.05427378 0.0001737
## 36-50-18-25 -0.19463414 -0.5729752  0.18370692 0.5489761
## 50+-18-25   -1.04078799 -3.7630271  1.68145115 0.7595433
## 36-50-26-35 -0.05366063 -0.4401483  0.33282701 0.9844493
## 50+-26-35   -0.89981447 -3.6231978  1.82356884 0.8309722
## 50+-36-50   -0.84615385 -3.5943245  1.90201678 0.8585337
```

```
# Narcissism to affect the GAD, SWL, SPIN scores in a positive way.
# Dependent varaible: SWL, GAD, SPIN (continuous)
# Independent variable: Narcissism (continuous)
# TEST: pearson

cor.test(data$SWL_T, data$Narcissism, method="pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  data$SWL_T and data$Narcissism
## t = -1.3191, df = 12652, p-value = 0.1872
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.02914447  0.00569825
## sample estimates:
##         cor
## -0.01172667
```

```
cor.test(data$GAD_T_n, data$Narcissism, method="pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  data$GAD_T_n and data$Narcissism
## t = 9.6373, df = 12652, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.0680442 0.1026378
## sample estimates:
##        cor
## 0.08536674
```

```
cor.test(data$SPIN_T_n, data$Narcissism, method="pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  data$SPIN_T_n and data$Narcissism
## t = -1.1631, df = 12652, p-value = 0.2448
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.02775825  0.00708553
```

```
## sample estimates:
##        cor
## -0.0103395
```

```r
# The GAD and SPIN scores to decrease with age, and the SWL to increase with age.
# Dependent variable: SWL, GAD, SPIN (continuous)
# Independent variable: age_group (categorical)
# TEST: One Way ANOVA
hypo.9 <- aov(SWL_T ~ age_group, data)
summary(hypo.9)
```

```
##                Df Sum Sq Mean Sq F value   Pr(>F)
## age_group       3    884  294.69   5.651 0.000726 ***
## Residuals   12650 659611   52.14
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
TukeyHSD(hypo.9)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = SWL_T ~ age_group, data = data)
##
## $age_group
##                    diff        lwr       upr     p adj
## 26-35-18-25  0.7219085  0.1309785  1.312839 0.0092159
## 36-50-18-25  1.7031773 -0.8755284  4.281883 0.3252339
## 50+-18-25   15.2608696 -3.2934299 33.815169 0.1488502
## 36-50-26-35  0.9812687 -1.6529625  3.615500 0.7737963
## 50+-26-35   14.5389610 -4.0231369 33.101059 0.1833326
## 50+-36-50   13.5576923 -5.1733516 32.288736 0.2456439
```

```r
hypo.9 <- aov(GAD_T_n ~ age_group, data)
summary(hypo.9)
```

```
##                Df Sum Sq Mean Sq F value   Pr(>F)
## age_group       3     44  14.774   11.69 1.21e-07 ***
## Residuals   12650  15989   1.264
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
TukeyHSD(hypo.9)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = GAD_T_n ~ age_group, data = data)
##
## $age_group
##                      diff        lwr        upr     p adj
```

```
## 26-35-18-25 -0.19901228 -0.2910165 -0.1070081 0.0000002
## 36-50-18-25 -0.17871244 -0.5802011  0.2227762 0.6623998
## 50+-18-25   -1.99943410 -4.8882247  0.8893565 0.2838003
## 36-50-26-35  0.02029984 -0.3898338  0.4304335 0.9992657
## 50+-26-35   -1.80042181 -4.6904266  1.0895830 0.3782237
## 50+-36-50   -1.82072166 -4.7370303  1.0955870 0.3762601
```

```
hypo.9 <- aov(SPIN_T_n ~ age_group, data)
summary(hypo.9)
```

```
##                Df Sum Sq Mean Sq F value Pr(>F)
## age_group       3    216   71.86   28.54 <2e-16 ***
## Residuals   12650  31854    2.52
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(hypo.9)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = SPIN_T_n ~ age_group, data = data)
##
## $age_group
##                   diff        lwr        upr     p adj
## 26-35-18-25 -0.4571491 -0.5870097 -0.3272886 0.0000000
## 36-50-18-25 -0.2370790 -0.8037656  0.3296076 0.7048787
## 50+-18-25   -2.7829507 -6.8603734  1.2944720 0.2960367
## 36-50-26-35  0.2200701 -0.3588185  0.7989588 0.7627164
## 50+-26-35   -2.3258016 -6.4049381  1.7533349 0.4588589
## 50+-36-50   -2.5458717 -6.6621351  1.5703917 0.3848303
```