BestProject

May 10, 2023

[1]:
```
#Cleaning the Dataset

# 1: begin by reading in the WLS dataset
WLS <- read.csv("WLS.csv")

# and have a look at it
summary(WLS)

# 2: delete all missing values.  Look at the WLS Guide to see what indicates a␣
↪missing value
# Look at R Assignment for to see how to set values to NA

WLS[WLS < 0] <- NA

# 3: delete unreported values from the death-year vector; again, look at the␣
↪guide
# this will ONLY apply to the WLS$deatyr vector

WLS$deatyr[WLS$deatyr > 9000] <- NA

# 4: adjust income levels to 2020 and put them in thousands
# this will ONLY apply to the WLS$pi5760 and WLS$yrer74 vectors
WLS$pi5760 <- WLS$pi5760  * (10.16/10)
WLS$yrer74 <- WLS$yrer74  * (7.71/10)

# 5: reverse self-reported health vectors.
# suggestion: make them all go low-to-high

WLS$mx001rer <- 6 - WLS$mx001rer

# 6: some ridiculous outliers were entered into the socializing datasets
# no one can socialize 300 times in four weeks
# come up with a reasonable limit for these

WLS$q1z023rer[WLS$q1z023rer > 100] <- 100
WLS$z023rer[WLS$iz023rer > 100] <- 100
WLS$iz023rer[WLS$iz023rer > 100] <- 100
```

```r
WLS$mz023rer[WLS$mz023rer > 100] <- 100


# 7: OPTIONAL BUT RECOMMENDED
# change the names of the vectors to something REASONABLE
# best to keep it short and meaningful
names(WLS) <- c("ID","Birth","Gender","Death","Income",
                "Population","Income75","Health1","Sad1","Social1",
                "Health2","Sad2","Social2","Cog2","Health3",
                "Sad3","Social3","Cog3","Health4","Sad4",
               "Social4","Cog4","Marriage","Children","Degree")


# check to see if you did it!
summary(WLS)
```

```
     idpub             brdxdy          sexrsp          deatyr
 Min.   :900018   Min.   :-3.00   Min.   :1.000   Min.   :  -2.0
 1st Qu.:908321   1st Qu.:39.00   1st Qu.:1.000   1st Qu.:  -2.0
 Median :916867   Median :39.00   Median :2.000   Median :  -2.0
 Mean   :916879   Mean   :36.47   Mean   :1.516   Mean   : 751.7
 3rd Qu.:925388   3rd Qu.:39.00   3rd Qu.:2.000   3rd Qu.:2002.0
 Max.   :933957   Max.   :40.00   Max.   :2.000   Max.   :9996.0

     pi5760            pop57           yrer74            mx001rer
 Min.   : -3.00   Min.   :2.000   Min.   :  -3.00   Min.   :-3.000
 1st Qu.: 24.00   1st Qu.:3.000   1st Qu.:   0.00   1st Qu.: 4.000
 Median : 49.00   Median :5.000   Median :  50.00   Median : 4.000
 Mean   : 54.18   Mean   :5.037   Mean   :  78.04   Mean   : 4.138
 3rd Qu.: 71.00   3rd Qu.:6.000   3rd Qu.: 130.00   3rd Qu.: 5.000
 Max.   :998.00   Max.   :9.000   Max.   :1650.00   Max.   : 5.000
                                                    NA's   :3442
   ru024rec           mz023rer          gx201re          gu024re
 Min.   :-3.0000   Min.   :-3.000   Min.   :-3.000   Min.   :-5.0000
 1st Qu.:-2.0000   1st Qu.: 1.000   1st Qu.: 2.000   1st Qu.:-2.0000
 Median : 0.0000   Median : 3.000   Median : 2.000   Median : 0.0000
 Mean   :-0.0022   Mean   : 3.906   Mean   : 2.259   Mean   :-0.1027
 3rd Qu.: 0.0000   3rd Qu.: 5.000   3rd Qu.: 3.000   3rd Qu.: 0.0000
 Max.   : 7.0000   Max.   :44.000   Max.   : 5.000   Max.   : 7.0000
 NA's   :1824      NA's   :3442     NA's   :2585     NA's   :3052
   iz023rer          gi503re          hx201re           hu024re
 Min.   :-3.000   Min.   :-5.000   Min.   :-30.000   Min.   :-30.000
 1st Qu.: 1.000   1st Qu.:-2.000   1st Qu.:  2.000   1st Qu.: -2.000
 Median : 2.000   Median : 6.000   Median :  2.000   Median : -2.000
 Mean   : 3.363   Mean   : 4.667   Mean   :  1.409   Mean   : -2.297
 3rd Qu.: 4.000   3rd Qu.: 9.000   3rd Qu.:  3.000   3rd Qu.: -2.000
 Max.   :28.000   Max.   :12.000   Max.   :  5.000   Max.   :  7.000
```

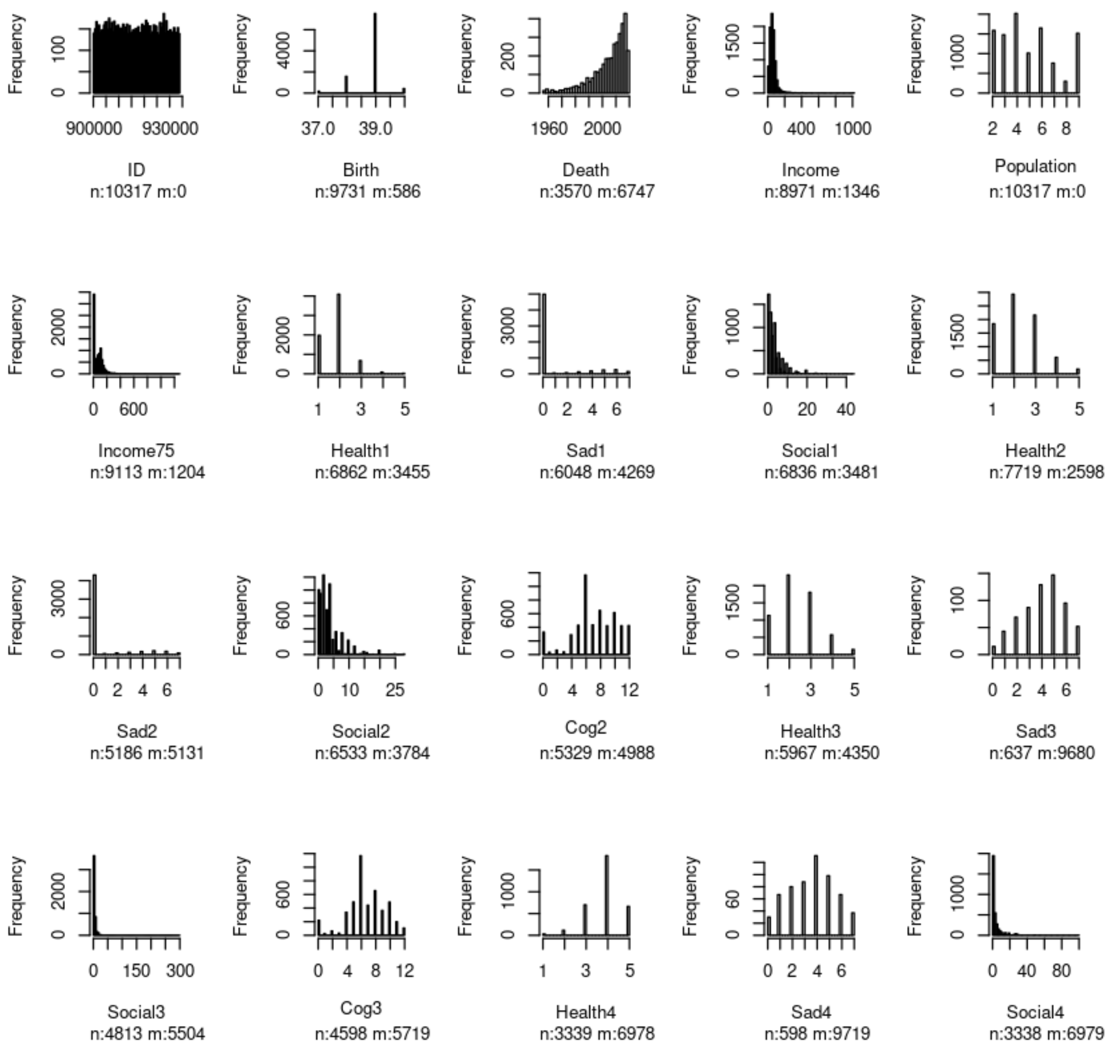
```
 NA's   :3472     NA's   :3052     NA's   :4165      NA's   :4165
   jz023rer         hi503rea         q1x001rer         q1u024re
 Min.   :-29.000   Min.   :-30.000   Min.   :-3.000   Min.   :-2.000
 1st Qu.:  1.000   1st Qu.: -2.000   1st Qu.: 3.000   1st Qu.:-2.000
 Median :  2.000   Median :  6.000   Median : 4.000   Median :-2.000
 Mean   :  0.794   Mean   :  3.616   Mean   : 3.876   Mean   :-0.996
 3rd Qu.:  5.000   3rd Qu.:  8.000   3rd Qu.: 4.000   3rd Qu.:-2.000
 Max.   :300.000   Max.   : 12.000   Max.   : 5.000   Max.   : 7.000
 NA's   :4926      NA's   :4165      NA's   :6969     NA's   :6969
   q1z023rer         q1i503rea         q1c003re          hd201kd
 Min.   : -3.000   Min.   :-5.000   Min.   :-5.000   Min.   : 0.00
 1st Qu.:  0.000   1st Qu.: 4.000   1st Qu.: 1.000   1st Qu.: 2.00
 Median :  2.000   Median : 6.000   Median : 1.000   Median : 3.00
 Mean   :  4.135   Mean   : 5.381   Mean   : 1.187   Mean   : 3.24
 3rd Qu.:  5.000   3rd Qu.: 8.000   3rd Qu.: 1.000   3rd Qu.: 4.00
 Max.   :150.000   Max.   :12.000   Max.   : 5.000   Max.   :10.00
 NA's   :6969      NA's   :6969     NA's   :6969     NA's   :4165
     dglev
 Min.   :-3.000
 1st Qu.:-2.000
 Median :-2.000
 Mean   :-1.078
 3rd Qu.:-2.000
 Max.   : 4.000


       ID             Birth            Gender           Death
 Min.   :900018   Min.   :37.00   Min.   :1.000   Min.   :1957
 1st Qu.:908321   1st Qu.:39.00   1st Qu.:1.000   1st Qu.:1999
 Median :916867   Median :39.00   Median :2.000   Median :2009
 Mean   :916879   Mean   :38.84   Mean   :1.516   Mean   :2005
 3rd Qu.:925388   3rd Qu.:39.00   3rd Qu.:2.000   3rd Qu.:2015
 Max.   :933957   Max.   :40.00   Max.   :2.000   Max.   :2019
                  NA's   :586                     NA's   :6747
     Income          Population        Income75          Health1
 Min.   :   0.00   Min.   :2.000   Min.   :   0.00   Min.   :1.000
 1st Qu.:  34.54   1st Qu.:3.000   1st Qu.:   0.00   1st Qu.:1.000
 Median :  54.86   Median :5.000   Median :  55.51   Median :2.000
 Mean   :  63.76   Mean   :5.037   Mean   :  68.42   Mean   :1.849
 3rd Qu.:  75.18   3rd Qu.:6.000   3rd Qu.: 107.94   3rd Qu.:2.000
 Max.   :1013.97   Max.   :9.000   Max.   :1272.15   Max.   :5.000
 NA's   :1346                      NA's   :1204      NA's   :3455
      Sad1           Social1          Health2           Sad2
 Min.   :0.000   Min.   : 0.000   Min.   :1.000   Min.   :0.000
 1st Qu.:0.000   1st Qu.: 1.000   1st Qu.:2.000   1st Qu.:0.000
 Median :0.000   Median : 3.000   Median :2.000   Median :0.000
 Mean   :0.831   Mean   : 3.945   Mean   :2.268   Mean   :0.752
 3rd Qu.:0.000   3rd Qu.: 5.000   3rd Qu.:3.000   3rd Qu.:0.000
```

```
 Max.   :7.000   Max.   :44.000   Max.   :5.000   Max.   :7.000
 NA's   :4269    NA's   :3481     NA's   :2598    NA's   :5131
    Social2            Cog2           Health3          Sad3
 Min.   : 0.000   Min.   : 0.000   Min.   :1.000   Min.   :0.000
 1st Qu.: 1.000   1st Qu.: 6.000   1st Qu.:2.000   1st Qu.:3.000
 Median : 3.000   Median : 7.000   Median :2.000   Median :4.000
 Mean   : 3.665   Mean   : 7.235   Mean   :2.378   Mean   :4.124
 3rd Qu.: 4.000   3rd Qu.:10.000   3rd Qu.:3.000   3rd Qu.:5.000
 Max.   :28.000   Max.   :12.000   Max.   :5.000   Max.   :7.000
 NA's   :3784     NA's   :4988     NA's   :4350    NA's   :9680
    Social3             Cog3           Health4          Sad4
 Min.   :  0.000   Min.   : 0.000   Min.   :1.000   Min.   :0.000
 1st Qu.:  1.000   1st Qu.: 5.000   1st Qu.:3.000   1st Qu.:2.000
 Median :  3.000   Median : 6.000   Median :4.000   Median :4.000
 Mean   :  4.332   Mean   : 6.754   Mean   :3.891   Mean   :3.622
 3rd Qu.:  5.000   3rd Qu.: 9.000   3rd Qu.:4.000   3rd Qu.:5.000
 Max.   :300.000   Max.   :12.000   Max.   :5.000   Max.   :7.000
 NA's   :5504      NA's   :5719     NA's   :6978    NA's   :9719
    Social4             Cog4           Marriage        Children
 Min.   :  0.000   Min.   : 0.000   Min.   :0.000   Min.   : 0.00
 1st Qu.:  0.000   1st Qu.: 4.000   1st Qu.:1.000   1st Qu.: 2.00
 Median :  2.000   Median : 6.000   Median :1.000   Median : 3.00
 Mean   :  4.119   Mean   : 5.458   Mean   :1.221   Mean   : 3.24
 3rd Qu.:  5.000   3rd Qu.: 8.000   3rd Qu.:1.000   3rd Qu.: 4.00
 Max.   :100.000   Max.   :12.000   Max.   :5.000   Max.   :10.00
 NA's   :6979      NA's   :7002     NA's   :6988    NA's   :4165
     Degree              NA
 Min.   :1.000   Min.   : NA
 1st Qu.:2.000   1st Qu.: NA
 Median :2.000   Median : NA
 Mean   :2.345   Mean   :NaN
 3rd Qu.:3.000   3rd Qu.: NA
 Max.   :4.000   Max.   : NA
 NA's   :7965    NA's   :10317
```

[6]:
```r
install.packages("Hmisc")
library(Hmisc)

hist.data.frame(WLS)
```

```
Installing package into ‘/opt/r’
(as ‘lib’ is unspecified)

also installing the dependencies ‘data.table’, ‘htmlTable’, ‘viridis’, ‘Formula’



Attaching package: ‘Hmisc’
```

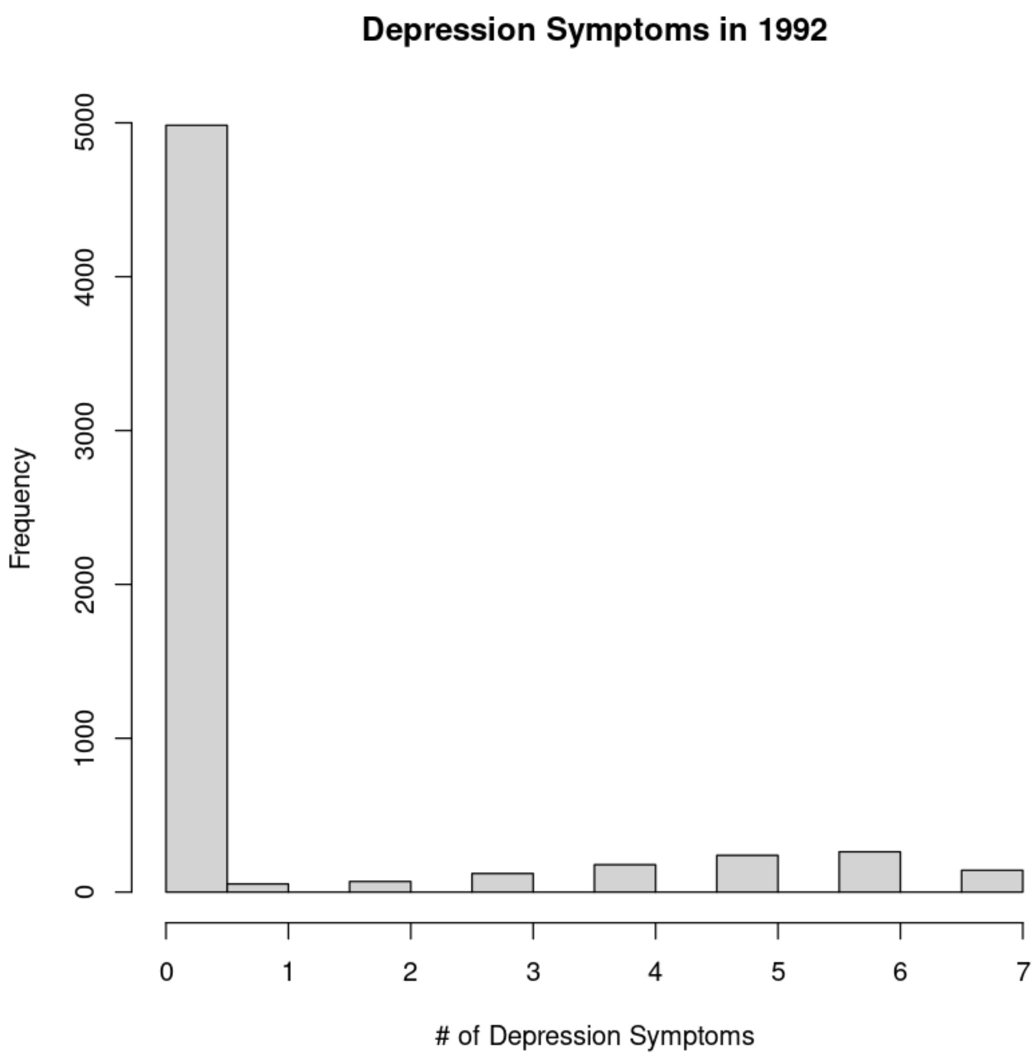
```
The following objects are masked from 'package:base':

    format.pval, units
```



[ ]:
```
#Interesting Comparison #1

Depression Symptoms with year 2020 (year of Covid). Also subseted between male␣
↪and female and create a hypothesis test if
    there is a signifincant amount of difference to say 2020 had a impact on␣
↪the number of depression symptoms.
```

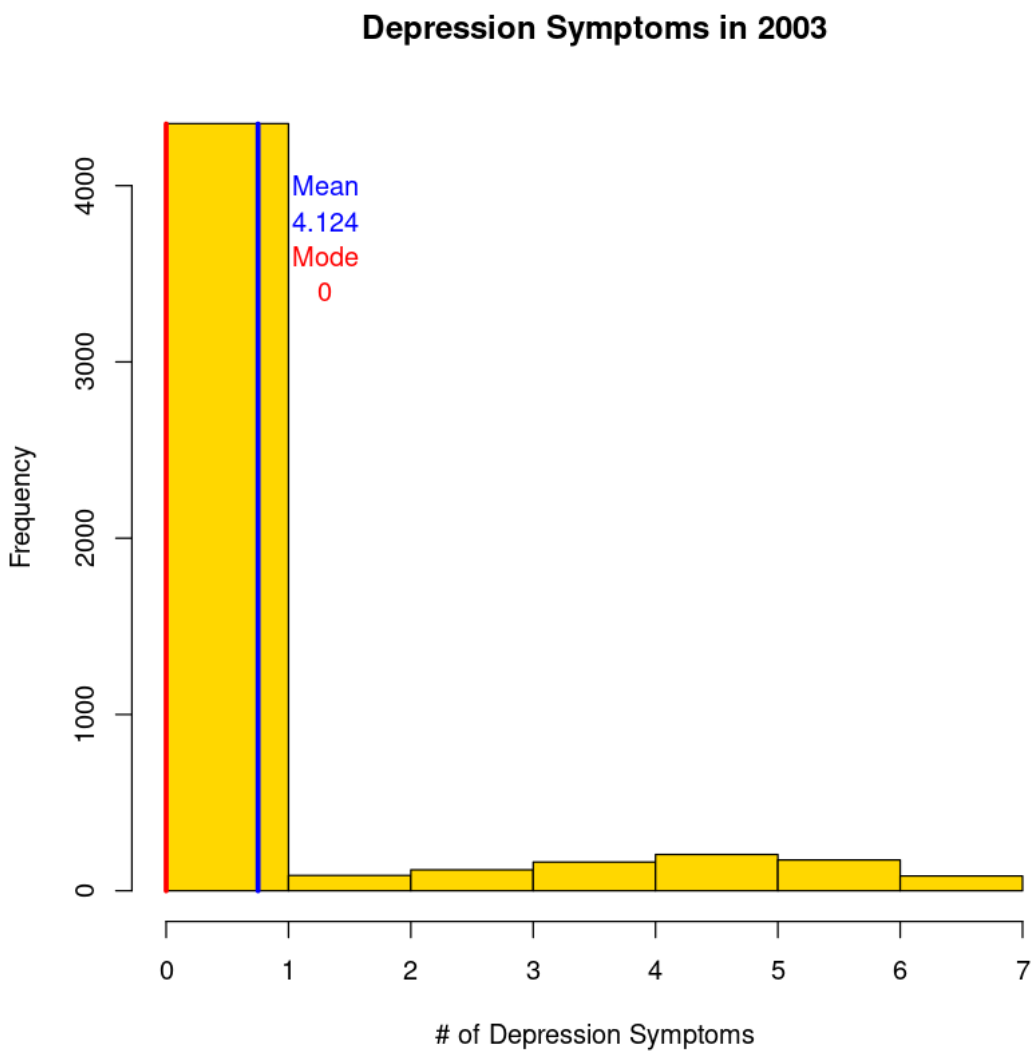
[61]:
```
#Histogram Sad1 Dataset

hist(WLS$Sad1, # dataframe$vector
    breaks = 25, # breaks sets the approximate number of bars
    main = "Depression Symptoms in 1992", # graph title
    xlab = "# of Depression Symptoms", # x-axis title
```

```
    ylab = "Frequency" # y-axis title
    )
```



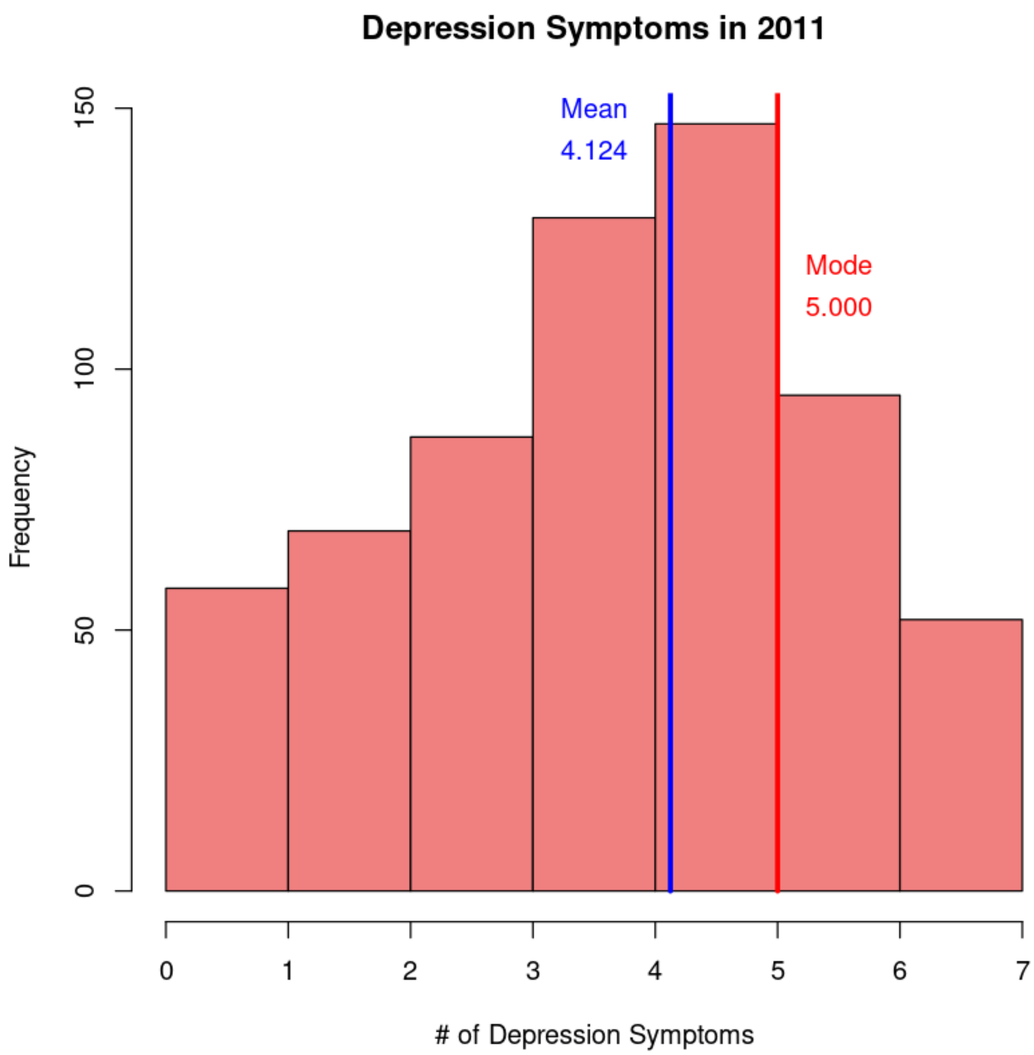
[11]:
```
#Histogram Sad2 Dataset

hist(WLS$Sad2, # dataframe$vector
    breaks = 7, # breaks sets the approximate number of bars
    main = "Depression Symptoms in 2003", # graph title
    xlab = "# of Depression Symptoms", # x-axis title
    ylab = "Frequency",
     col = "gold" # y-axis title
    )
```

```
segments(0.751,0, # x and y coordinates for one end of the line segment
        0.751,4350, # x and y coordinates for the other end.  Note that the x␣
↪value stays the same--a vertical line
        lwd = 3,  # line width--make the line three times as thick for clarity
        col = "blue" # a favorite color of mine. YOU CAN'T USE IT
        )

text(1.3,4000, # x and y coordinates of where the text goes
     labels = "Mean", # what to print there
     cex = 1, # cex means Character Expansion, and can make the characters␣
↪larger or smaller.  1 is the default
     col = "blue"
     )
text(1.3,3800, # same x but y slightly lower
     labels = "4.124",
     cex = 1,
     col = "blue"
     )


segments(0,0, # x and y coordinates for one end of the line segment
        0,4350, # x and y coordinates for the other end.  Note that the x␣
↪value stays the same--a vertical line
        lwd = 3,  # line width--make the line three times as thick for clarity
        col = "red" # a favorite color of mine. YOU CAN'T USE IT
        )

text(1.3,3600, # x and y coordinates of where the text goes
     labels = "Mode", # what to print there
     cex = 1, # cex means Character Expansion, and can make the characters␣
↪larger or smaller.  1 is the default
     col = "red"
     )
text(1.3,3400, # same x but y slightly lower
     labels = "0",
     cex = 1,
     col = "red"
     )
```



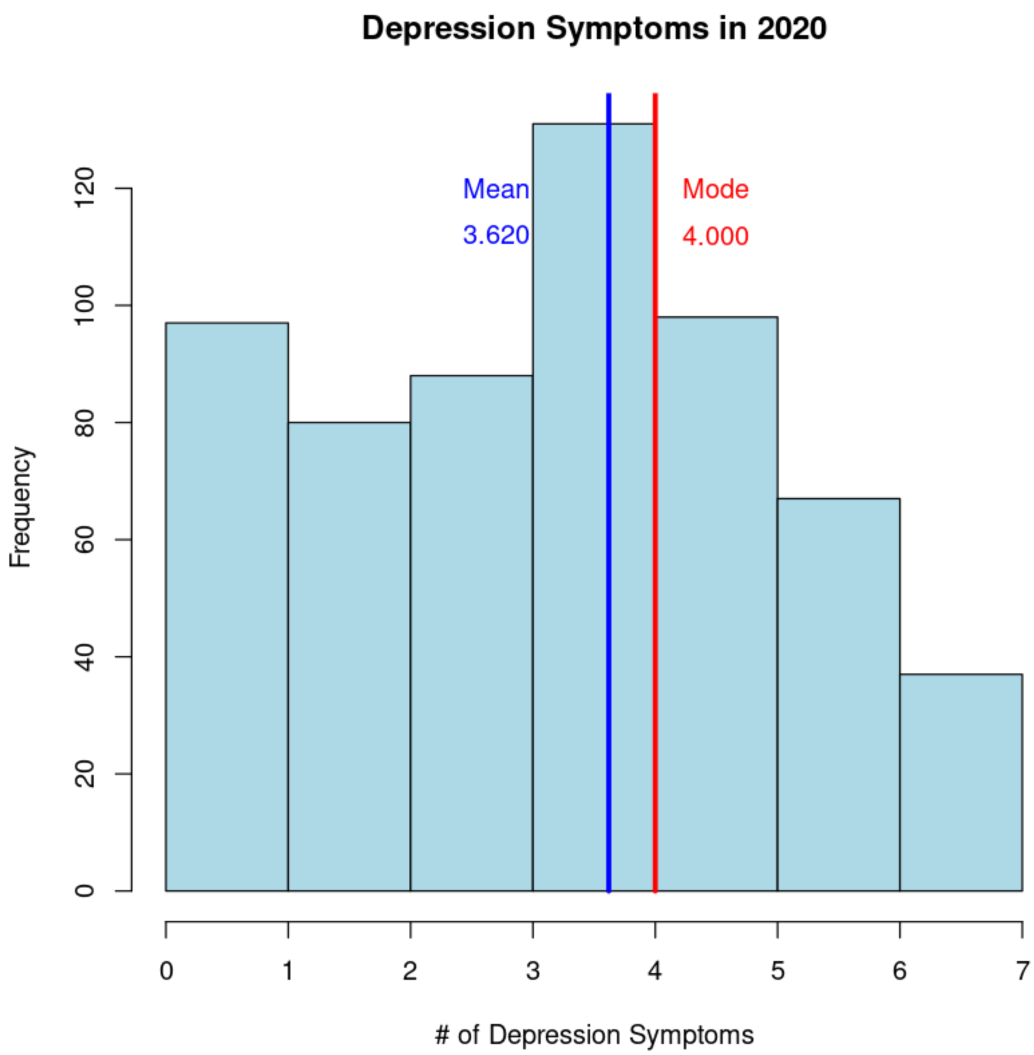
[17]:
```
#Histograms Sad3 Dataset

hist(WLS$Sad3, # dataframe$vector
    breaks = 7, # breaks sets the approximate number of bars
    main = "Depression Symptoms in 2011", # graph title
    xlab = "# of Depression Symptoms", # x-axis title
    ylab = "Frequency",
     col = "lightcoral" # y-axis title
    )

segments(4.124,0, # x and y coordinates for one end of the line segment
         4.124,155, # x and y coordinates for the other end.  Note that the x␣
 ↪value stays the same--a vertical line
```

```r
        lwd = 3,  # line width--make the line three times as thick for clarity
        col = "blue" # a favorite color of mine. YOU CAN'T USE IT
        )

text(3.5,150, # x and y coordinates of where the text goes
     labels = "Mean", # what to print there
     cex = 1, # cex means Character Expansion, and can make the characters␣
 ↪larger or smaller.  1 is the default
     col = "blue"
     )
text(3.5,142, # same x but y slightly lower
     labels = "4.124",
     cex = 1,
     col = "blue"
     )


segments(5,0, # x and y coordinates for one end of the line segment
        5,155, # x and y coordinates for the other end.  Note that the x value␣
 ↪stays the same--a vertical line
        lwd = 3,  # line width--make the line three times as thick for clarity
        col = "red" # a favorite color of mine. YOU CAN'T USE IT
        )

text(5.5, 120, # x and y coordinates of where the text goes
     labels = "Mode", # what to print there
     cex = 1, # cex means Character Expansion, and can make the characters␣
 ↪larger or smaller.  1 is the default
     col = "red"
     )
text(5.5,112, # same x but y slightly lower
     labels = "5.000",
     cex = 1,
     col = "red"
     )
```



[4]:
```
#Histogram for Depression Symptoms in 2020


hist(WLS$Sad4, # dataframe$vector
    breaks = 7, # breaks sets the approximate number of bars
    main = "Depression Symptoms in 2020", # graph title
    xlab = "# of Depression Symptoms", # x-axis title
    ylab = "Frequency",
     col = "lightblue"
    )

segments(3.620,0, # x and y coordinates for one end of the line segment
```

```
        3.620,155, # x and y coordinates for the other end.  Note that the x␣
 ↪value stays the same--a vertical line
        lwd = 3,  # line width--make the line three times as thick for clarity
        col = "blue" # a favorite color of mine. YOU CAN'T USE IT
        )

text(2.7,120, # x and y coordinates of where the text goes
     labels = "Mean", # what to print there
     cex = 1, # cex means Character Expansion, and can make the characters␣
 ↪larger or smaller.  1 is the default
     col = "blue"
     )
text(2.7,112, # same x but y slightly lower
     labels = "3.620",
     cex = 1,
     col = "blue"
     )


segments(4,0, # x and y coordinates for one end of the line segment
        4,155, # x and y coordinates for the other end.  Note that the x value␣
 ↪stays the same--a vertical line
        lwd = 3,  # line width--make the line three times as thick for clarity
        col = "red" # a favorite color of mine. YOU CAN'T USE IT
        )

text(4.5, 120, # x and y coordinates of where the text goes
     labels = "Mode", # what to print there
     cex = 1, # cex means Character Expansion, and can make the characters␣
 ↪larger or smaller.  1 is the default
     col = "red"
     )
text(4.5,112, # same x but y slightly lower
     labels = "4.000",
     cex = 1,
     col = "red"
     )
```



[15]:
```
#Signifcant difference in means between 2003, 2011, and 2020 bulk samples

var.test(WLS$Sad3, WLS$Sad4)

t.test(WLS$Sad3, WLS$Sad4,
       conf.level = 0.99,
       alternative = "g",
       var.equal = TRUE)

prop.test(x = c(4304, 15),
          n = c(5186, 637),
          conf.level = 0.99,
          alternative = "t"
```

```
          )

prop.test(x = c(4304, 15),
          n = c(5186, 637),
          conf.level = 0.99,
          alternative = "g"
          )

zerosad2 <- subset(WLS, Sad2 < 1)
zerosad3 <- subset(WLS, Sad3 < 1)

zerosad22 <- subset(WLS, Sad2 < 8)
zerosad33 <- subset(WLS, Sad3 < 8)

nrow(zerosad2)
nrow(zerosad3)

nrow(zerosad22)
nrow(zerosad33)



zerosad5 <- subset(WLS, Sad3 < 1)
zerosad6 <- subset(WLS, Sad4 < 1)

zerosad55 <- subset(WLS, Sad3 < 8)
zerosad66 <- subset(WLS, Sad4 < 8)

nrow(zerosad5)
nrow(zerosad6)

nrow(zerosad55)
nrow(zerosad66)


#Results: There is not a significant different in their variances, but there is
 ↪a significant different in their means. (Very significant)
```

```
        F test to compare two variances

data:  WLS$Sad3 and WLS$Sad4
F = 0.88097, num df = 636, denom df = 597, p-value = 0.1157
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.7519912 1.0316598
sample estimates:
```

```
ratio of variances
          0.880971


        Two Sample t-test

data:  WLS$Sad3 and WLS$Sad4
t = 4.8462, df = 1233, p-value = 7.091e-07
alternative hypothesis: true difference in means is greater than 0
99 percent confidence interval:
 0.2606826       Inf
sample estimates:
mean of x mean of y
 4.124019  3.622074



        2-sample test for equality of proportions with continuity correction

data:  c(4304, 15) out of c(5186, 637)
X-squared = 1921.4, df = 1, p-value < 2.2e-16
alternative hypothesis: two.sided
99 percent confidence interval:
 0.7850017 0.8277560
sample estimates:
    prop 1     prop 2
0.82992673 0.02354788



        2-sample test for equality of proportions with continuity correction

data:  c(4304, 15) out of c(5186, 637)
X-squared = 1921.4, df = 1, p-value < 2.2e-16
alternative hypothesis: greater
99 percent confidence interval:
 0.7869868 1.0000000
sample estimates:
    prop 1     prop 2
0.82992673 0.02354788
```

4304

15

5186

637

15

30

637

598

[34]:
```r
prop.test(x = c(15, 30),
          n = c(637, 598),
          conf.level = 0.99,
          alternative = "t"
          )

prop.test(x = c(15, 30),
          n = c(637, 598),
          conf.level = 0.99,
          alternative = "l"
          )

zerosad5 <- subset(WLS, Sad3 < 1)
zerosad6 <- subset(WLS, Sad4 < 1)

zerosad55 <- subset(WLS, Sad3 < 8)
zerosad66 <- subset(WLS, Sad4 < 8)

nrow(zerosad5)
nrow(zerosad6)

nrow(zerosad55)
nrow(zerosad66)
```

```
	2-sample test for equality of proportions with continuity correction

data:  c(15, 30) out of c(637, 598)
X-squared = 5.49, df = 1, p-value = 0.01913
alternative hypothesis: two.sided
99 percent confidence interval:
 -0.055956534  0.002717848
sample estimates:
    prop 1     prop 2
0.02354788 0.05016722



	2-sample test for equality of proportions with continuity correction

data:  c(15, 30) out of c(637, 598)
X-squared = 5.49, df = 1, p-value = 0.009563
alternative hypothesis: less
```

```
99 percent confidence interval:
 -1.000000e+00  3.340637e-05
sample estimates:
    prop 1     prop 2
0.02354788 0.05016722
```

15

30

637

598

[8]:
```
#Significant difference in means between 2011 and 2020 in two samples (male and
 ↪female)

male <- subset(WLS, Gender == 1)
female <- subset(WLS, Gender == 2)

var.test(male$Sad4,female$Sad4
         )

t.test(male$Sad4,female$Sad4,
       conf.level = 0.99,
       var.equal = TRUE)
```

```
        F test to compare two variances

data:  male$Sad4 and female$Sad4
F = 1.0931, num df = 200, denom df = 396, p-value = 0.4583
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.8631481 1.3977221
sample estimates:
ratio of variances
          1.093082



        Two Sample t-test

data:  male$Sad4 and female$Sad4
t = 0.45906, df = 596, p-value = 0.6464
alternative hypothesis: true difference in means is not equal to 0
99 percent confidence interval:
 -0.3456273  0.4949563
sample estimates:
```

```
mean of x mean of y
 3.671642  3.596977
```

[41]:
```
#It seems that depression symptoms increased from the 1900's to around 2011,␣
↪then decreased significantly during Covid

mean(WLS$Sad1, na.rm = TRUE)
mean(WLS$Sad2, na.rm = TRUE)
mean(WLS$Sad3, na.rm = TRUE)
mean(WLS$Sad4, na.rm = TRUE)

#Medians

median(WLS$Sad1, na.rm = TRUE)
median(WLS$Sad2, na.rm = TRUE)
median(WLS$Sad3, na.rm = TRUE)
median(WLS$Sad4, na.rm = TRUE)

#Modes

library(gtools)

stat_mode(WLS$Sad1)
stat_mode(WLS$Sad2)
stat_mode(WLS$Sad3)
stat_mode(WLS$Sad4)
```

0.831183862433862

0.751831854994215

4.12401883830455

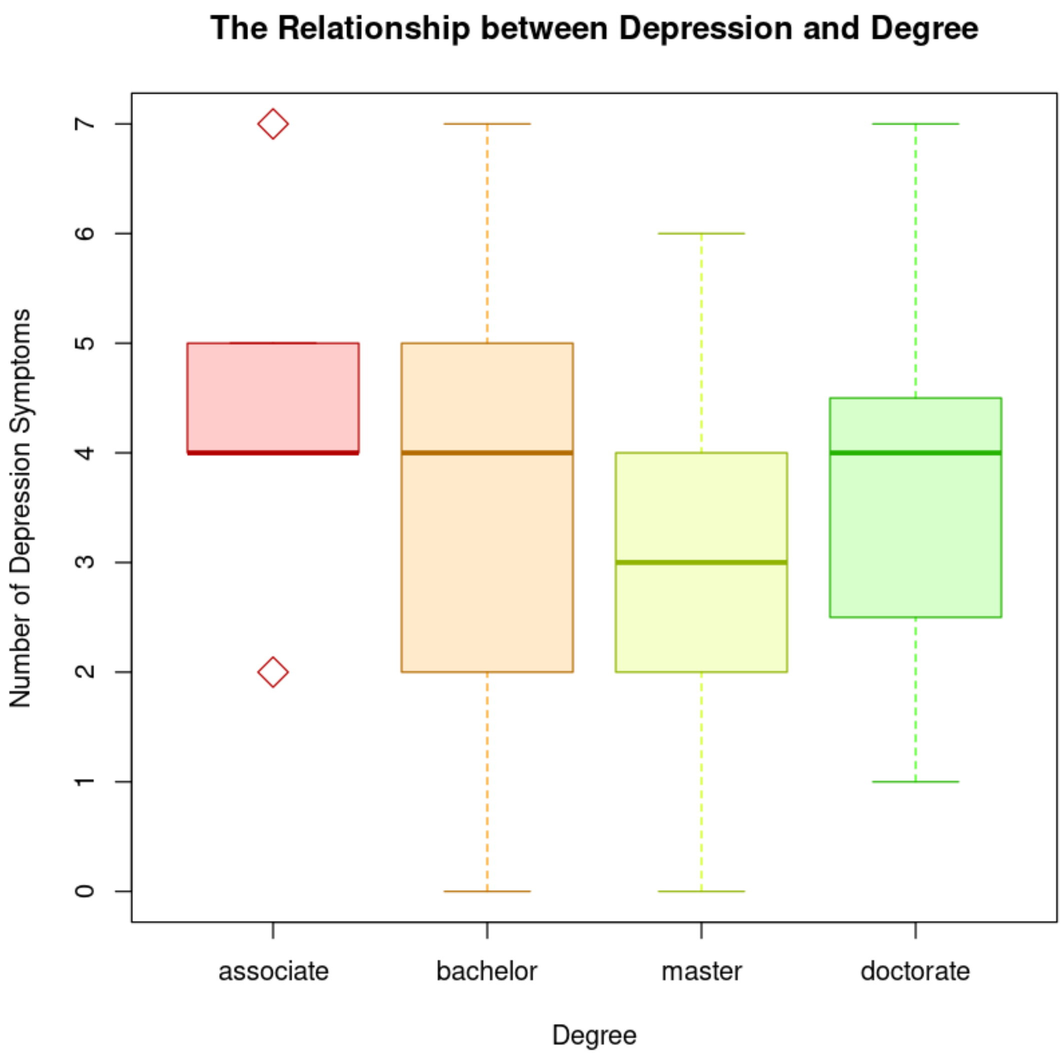
3.62207357859532

[3]:
```
#Does your education level have an effect on the number of depression symptoms?

summary(aov(                    # summarize ANOVA
        Sad4 ~ factor(Degree),   # the formula. factor() means treat this as a␣
↪factor, not numeric
        data = WLS          # the dataframe
        ))

#This is not concrete evidence that there is no relationship with the␣
↪relatively high p-value. There are four factors within the Degree Dataframe.
```

```
                Df Sum Sq Mean Sq F value Pr(>F)
factor(Degree)   3   17.1   5.705   1.661  0.179
```

```
Residuals      119  408.7   3.435
10194 observations deleted due to missingness
```

[4]:
```
c1 <- rainbow(10)
c2 <- rainbow(10, alpha=0.2)
c3 <- rainbow(10, v=0.7)

boxplot(Sad4~Degree,
        data = WLS,
        names = c("associate", "bachelor", "master", "doctorate"),
        xlab = "Degree",
        ylab = "Number of Depression Symptoms",
        main = "The Relationship between Depression and Degree",
       col = c2,
       medcol = c3,
       whiskcol = c1,
       staplecol= c3,
       boxcol = c3,
       outcol = c3,
       pch = 23,
       cex = 2)
```



[2]:
```
#It seems that people with a master's degree have significantly less depression␣
↪symptoms than the rest of the degrees

associate <- subset(WLS, Degree == 1)
bachelor <- subset(WLS, Degree == 2)
master <- subset(WLS, Degree == 3)
doctor <- subset(WLS, Degree == 4)
```

[33]:
```
sum(!is.na(associate))
sum(!is.na(bachelor))
sum(!is.na(master))
sum(!is.na(doctor))
```

```
jesus <- 2476 + 23825 + 9180 + 3584

jesus
```

2476

23825

9180

3584

39065

[6]:
```
#Conduct a t.test to see if the different between the master mean is␣
↪significantly different from any other mean since the rest seem to have a␣
↪similar mean.

#Perform t.test master with associate
t.test(master$Sad4,associate$Sad4,
       conf.level = 0.95,
       var.equal = TRUE)

#Perform t.test master with bachelor
t.test(master$Sad4,bachelor$Sad4,
       conf.level = 0.95,
       alternative = "g",
       var.equal = TRUE)

#Perform t.test master with doctorate
t.test(master$Sad4,doctor$Sad4,
       conf.level = 0.95,
       var.equal = TRUE)

t.test(master$Sad4,associate$Sad4,
       conf.level = 0.99,
       alternative = "l",
       var.equal = TRUE)



#The only signficant difference was between masters and associates. People with␣
↪a master's degree have a significantly lower amount of depression symptoms␣
↪than people with only associate's degrees.
```

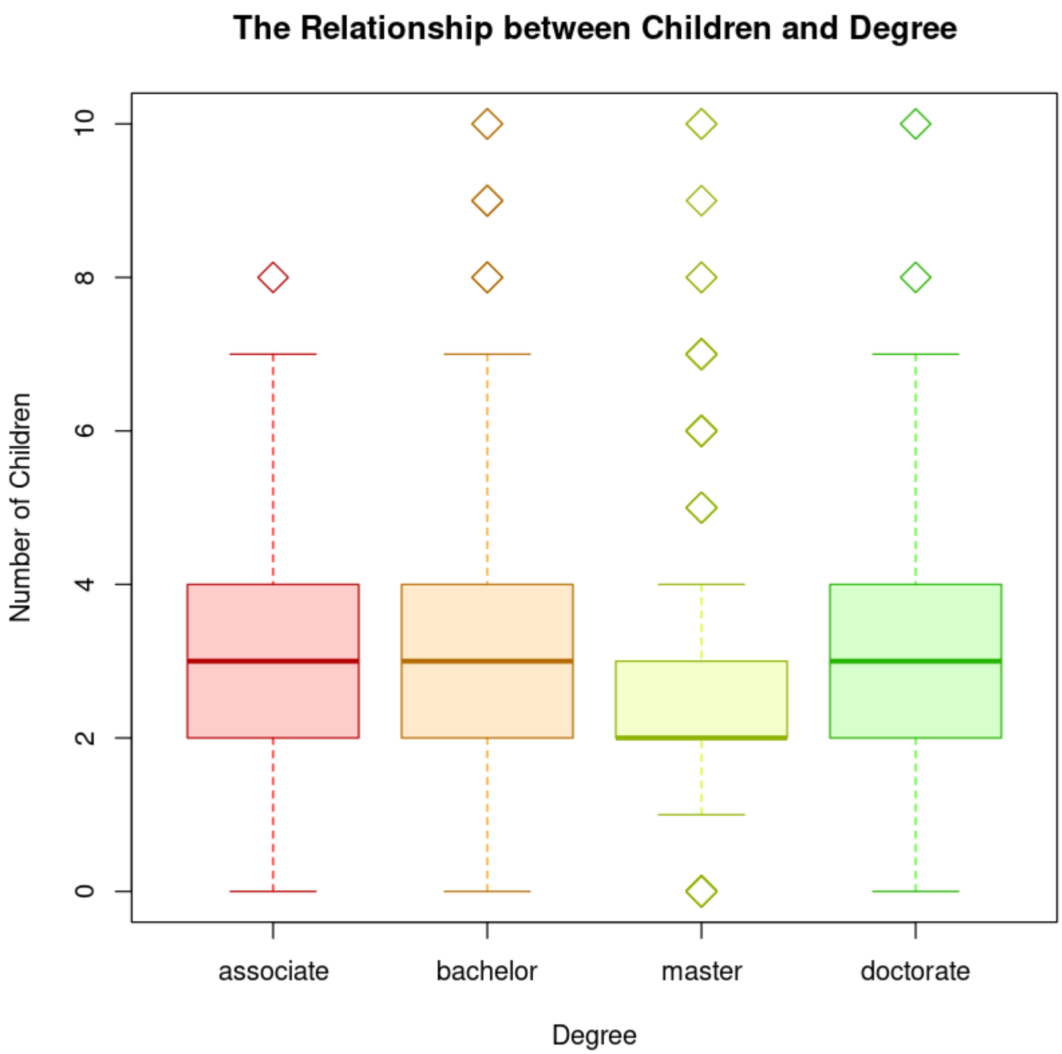
```
        Two Sample t-test

data:  master$Sad4 and associate$Sad4
t = -2.1773, df = 34, p-value = 0.03649
```

```
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.8108392 -0.0968531
sample estimates:
mean of x mean of y
 2.846154  4.300000


        Two Sample t-test

data:  master$Sad4 and bachelor$Sad4
t = -1.3953, df = 99, p-value = 0.917
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -1.329738       Inf
sample estimates:
mean of x mean of y
 2.846154  3.453333


        Two Sample t-test

data:  master$Sad4 and doctor$Sad4
t = -1.2466, df = 36, p-value = 0.2206
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.1553961  0.5143705
sample estimates:
mean of x mean of y
 2.846154  3.666667


        Two Sample t-test

data:  master$Sad4 and associate$Sad4
t = -2.1773, df = 34, p-value = 0.01825
alternative hypothesis: true difference in means is less than 0
99 percent confidence interval:
      -Inf 0.1761856
sample estimates:
mean of x mean of y
 2.846154  4.300000
```

[16]:

```
#Lets dive deep into the sample of people who have masters and people who have
 ↪associates and see what other
```

```
#significantly different factors there are between them. FIRST, LET's SEE IF␣
↪THE NUMBER OF CHILDREN IS CAUSING
#THIS DIFFERENCE!

c1 <- rainbow(10)
c2 <- rainbow(10, alpha=0.2)
c3 <- rainbow(10, v=0.7)

boxplot(Children~Degree,
        data = WLS,
        names = c("associate", "bachelor", "master", "doctorate"),
        xlab = "Degree",
        ylab = "Number of Children",
        main = "The Relationship between Children and Degree",
       col = c2,
       medcol = c3,
       whiskcol = c1,
       staplecol= c3,
       boxcol = c3,
       outcol = c3,
       pch = 23,
       cex = 2)
```



[8]:
```
#Wow, looking at the boxplot only, it seems like people with master's degrees␣
↪have on average less children!

master1 <- subset(WLS, Degree == 3)
associate1 <- subset(WLS, Degree == 1)

#Perform t.test master with associate
t.test(master1$Children,associate1$Children,
       conf.level = 0.99,
       alternative = "l",
       var.equal = TRUE)
```

```
        Two Sample t-test

data:  master1$Children and associate1$Children
t = -3.5965, df = 514, p-value = 0.0001768
alternative hypothesis: true difference in means is less than 0
99 percent confidence interval:
       -Inf -0.2424976
sample estimates:
mean of x mean of y
 2.523002  3.213592
```

[ ]:
```
#There is an extremely signficant difference in means between the number of␣
↪children with associates and the number of children with masters.
```

[8]:
```
#Scatter plot with Sad4/master + Sad4/associate + Children/master + Children/
↪associate

#Plot the means to emphasize difference in means

mean(master1$Children, na.rm = TRUE)
mean(associate1$Children, na.rm = TRUE)

mean(master$Sad4, na.rm = TRUE)
mean(associate$Sad4, na.rm = TRUE)
```

2.52300242130751

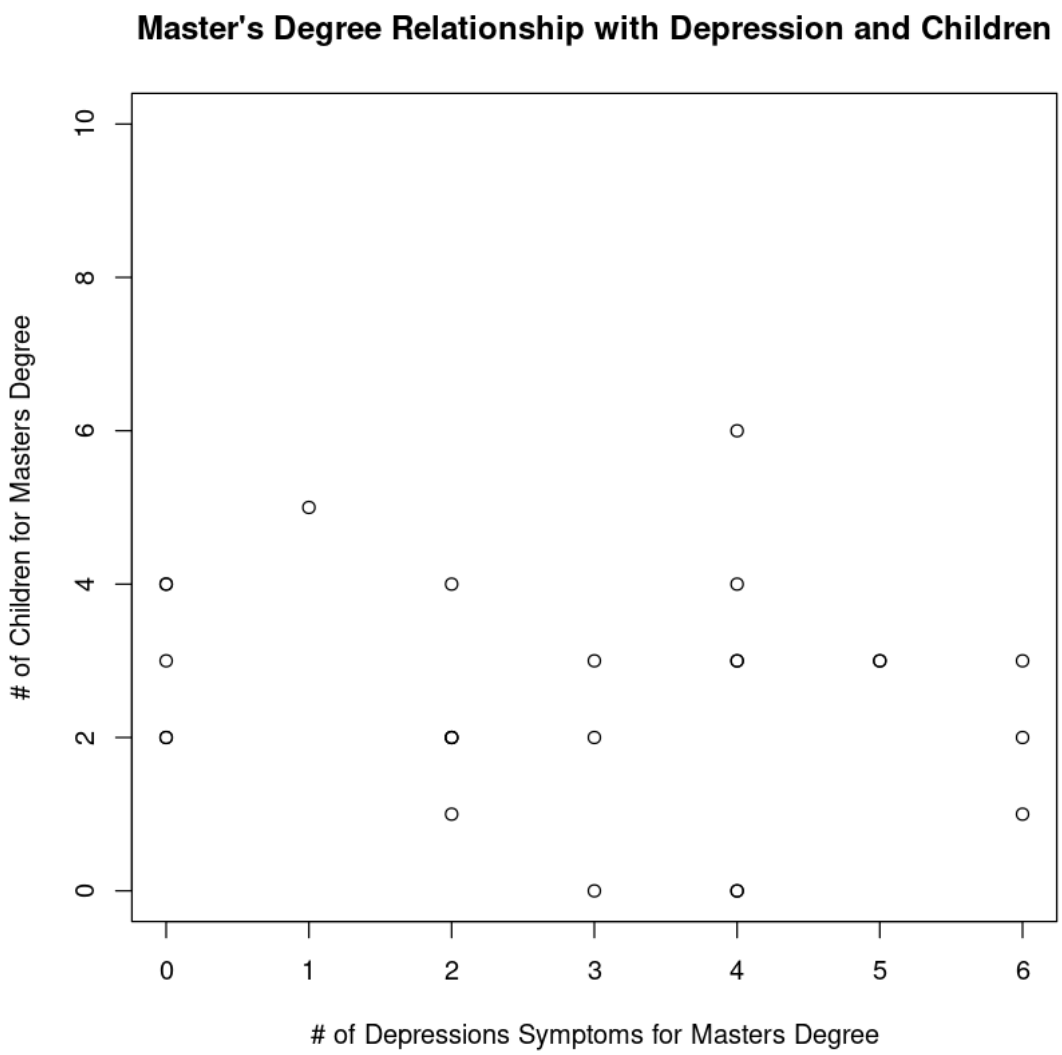
3.21359223300971

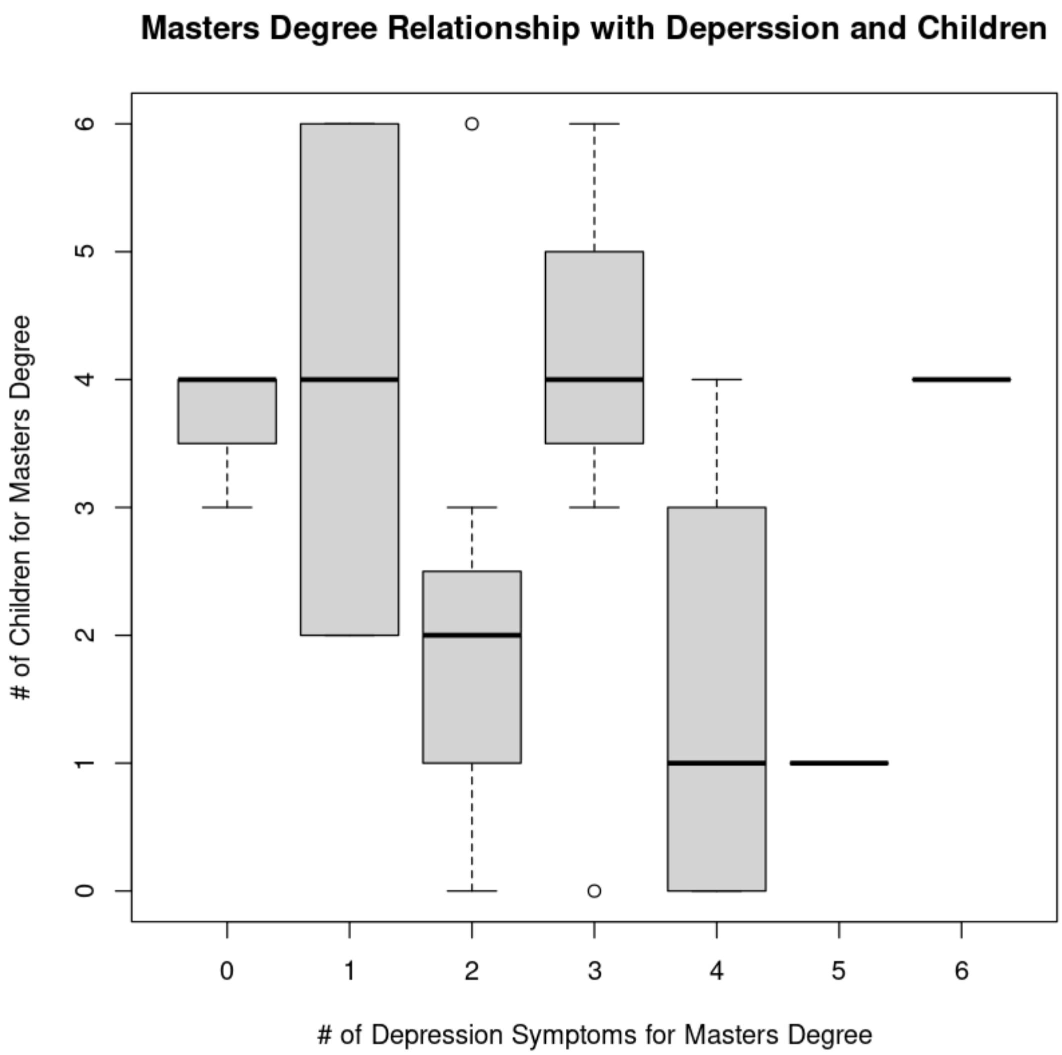
2.84615384615385

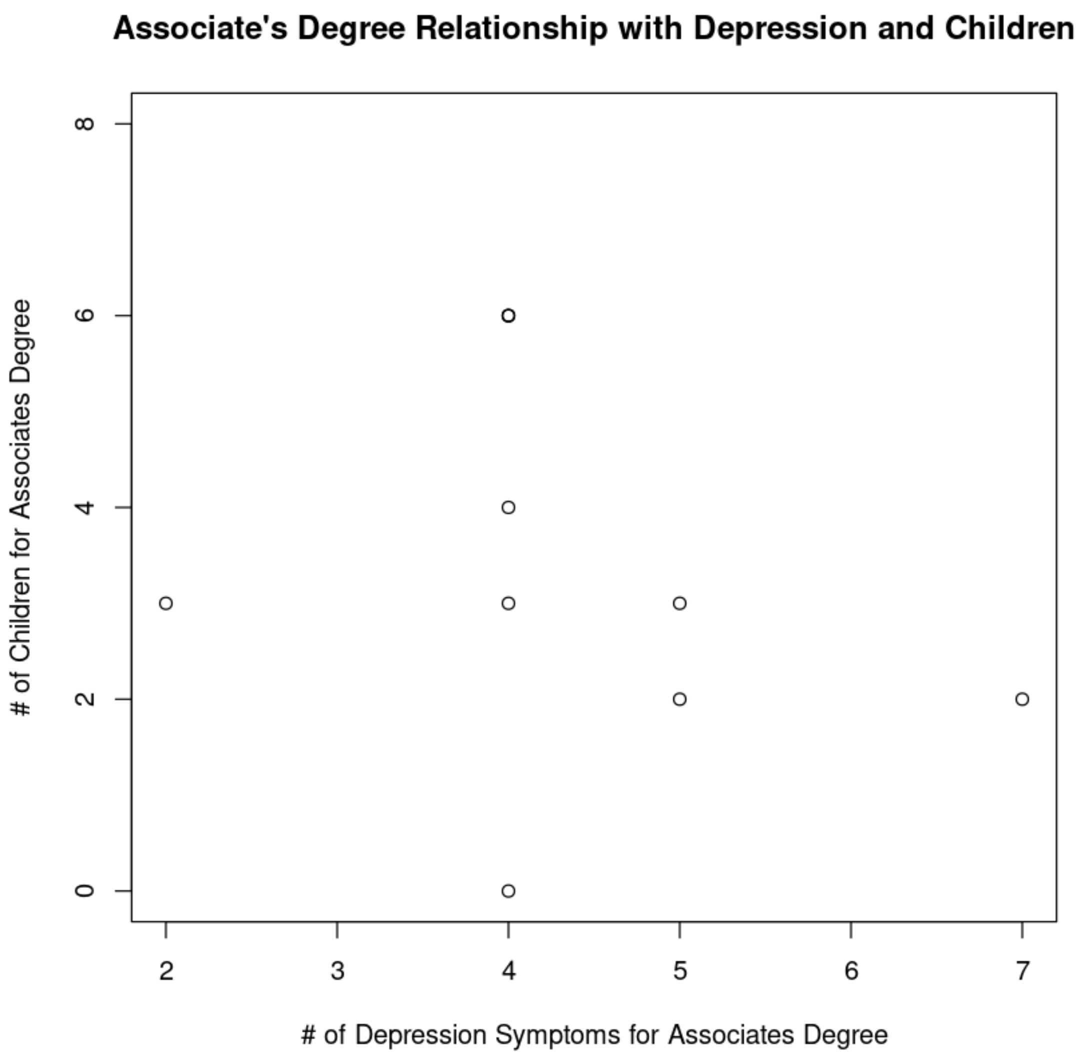
4.3

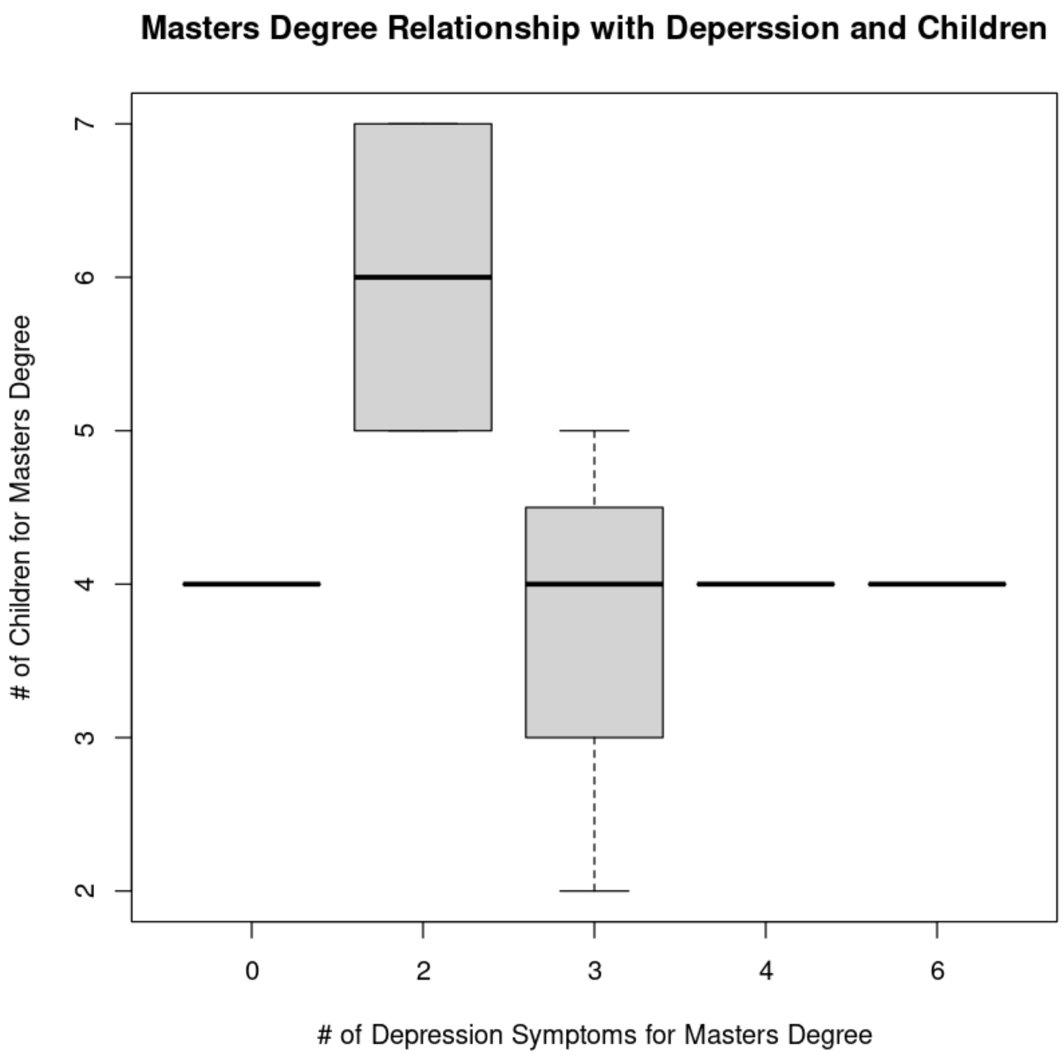
[6]:
```
#With the masters subset, plot the Sad4 and Children Against each other

plot(master$Sad4,master1$Children,  # x vector, y vector
     main = "Master's Degree Relationship with Depression and Children",
     xlab = "# of Depressions Symptoms for Masters Degree",
     ylab = "# of Children for Masters Degree"
    )

boxplot(master$Sad4~master1$Children,
        data = WLS,
        xlab = "# of Depression Symptoms for Masters Degree",
        ylab = "# of Children for Masters Degree",
        main = "Masters Degree Relationship with Deperssion and Children")

plot(associate$Sad4, associate$Children,
```

```
    main = "Associate's Degree Relationship with Depression and Children",
    xlab = "# of Depression Symptoms for Associates Degree",
    ylab = "# of Children for Associates Degree")

boxplot(associate$Sad4~associate1$Children,
        data = WLS,
        xlab = "# of Depression Symptoms for Masters Degree",
        ylab = "# of Children for Masters Degree",
        main = "Masters Degree Relationship with Deperssion and Children")
```



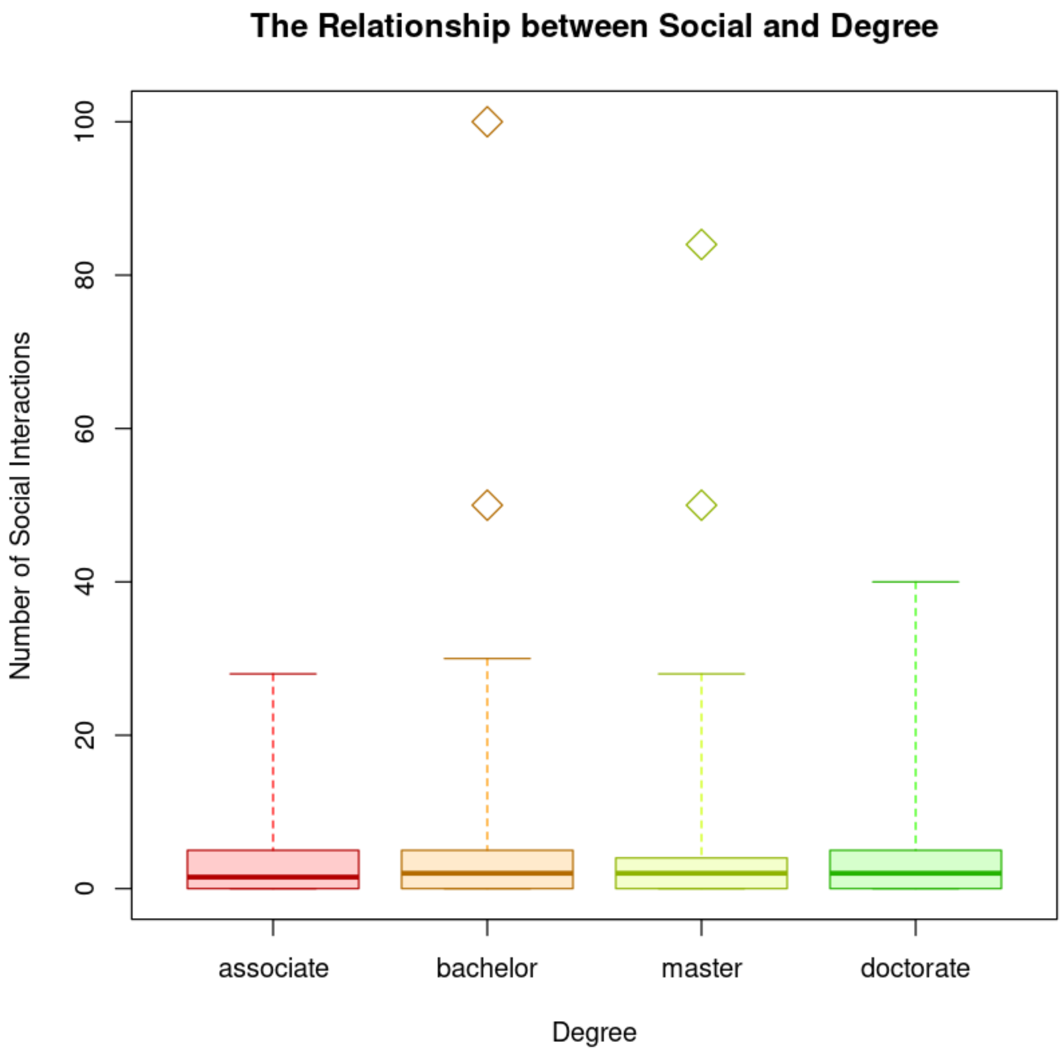

Masters Degree Relationship with Deperssion and Children
# of Children for Masters Degree
0
1
2
3
4
5
6
# of Depression Symptoms for Masters Degree


Associate's Degree Relationship with Depression and Children
# of Children for Associates Degree
# of Depression Symptoms for Associates Degree
0
2
4
6
8
2
3
4
5
6
7



[13]:
```
#Also, you can maybe do that "even though associates had more meeting's with␣
↪friends they had more depression symptoms
#than people with masters.

c1 <- rainbow(10)
c2 <- rainbow(10, alpha=0.2)
c3 <- rainbow(10, v=0.7)

boxplot(Social4~Degree,
        data = WLS,
        names = c("associate", "bachelor", "master", "doctorate"),
        xlab = "Degree",
        ylab = "Number of Social Interactions",
```

```
 main = "The Relationship between Social and Degree",
 range = 8,
col = c2,
medcol = c3,
whiskcol = c1,
staplecol= c3,
boxcol = c3,
outcol = c3,
pch = 23,
cex = 2)
```



[38]:
```
master2 <- subset(WLS, Degree == 3)
associate2 <- subset(WLS, Degree == 1)

t.test(master2$Social4,associate2$Social4,
       conf.level = 0.99,
       var.equal = TRUE)

var.test(associate$Social, master$Social)
```

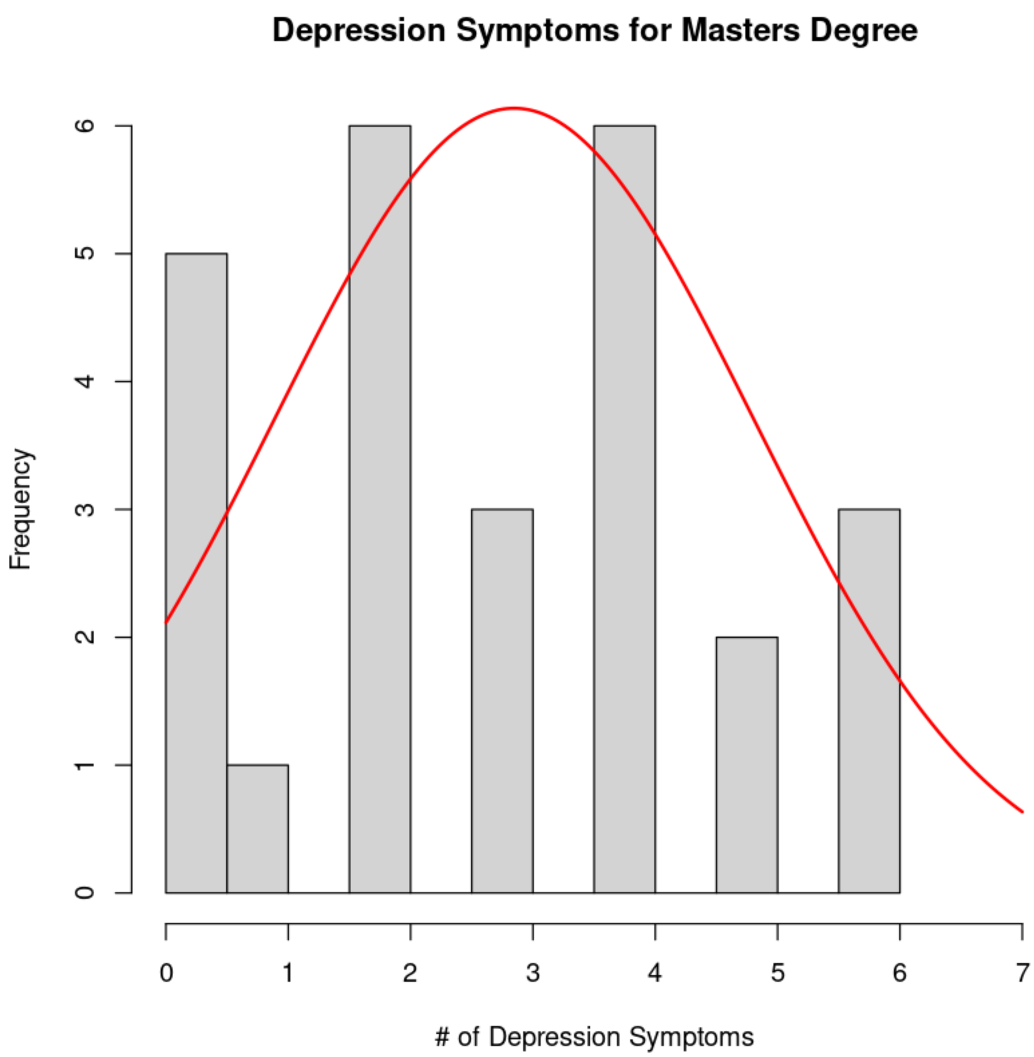
```
        Two Sample t-test

data:  master2$Social4 and associate2$Social4
t = 0.49644, df = 302, p-value = 0.6199
alternative hypothesis: true difference in means is not equal to 0
99 percent confidence interval:
 -2.452257  3.614040
sample estimates:
mean of x mean of y
 4.253968  3.673077
```

```
Error in var.test.default(associate$Social, master$Social): not enough 'x'␣
↪observations
Traceback:

1. var.test(associate$Social, master$Social)
2. var.test.default(associate$Social, master$Social)
3. stop("not enough 'x' observations")
```

[ ]:
```
#There is no signficance between the number of interactions and degree. If␣
↪there is no signifance here, there is no point
# in searching if depression symptoms are related to number of interactions for␣
↪the subsets associate and masters.
```

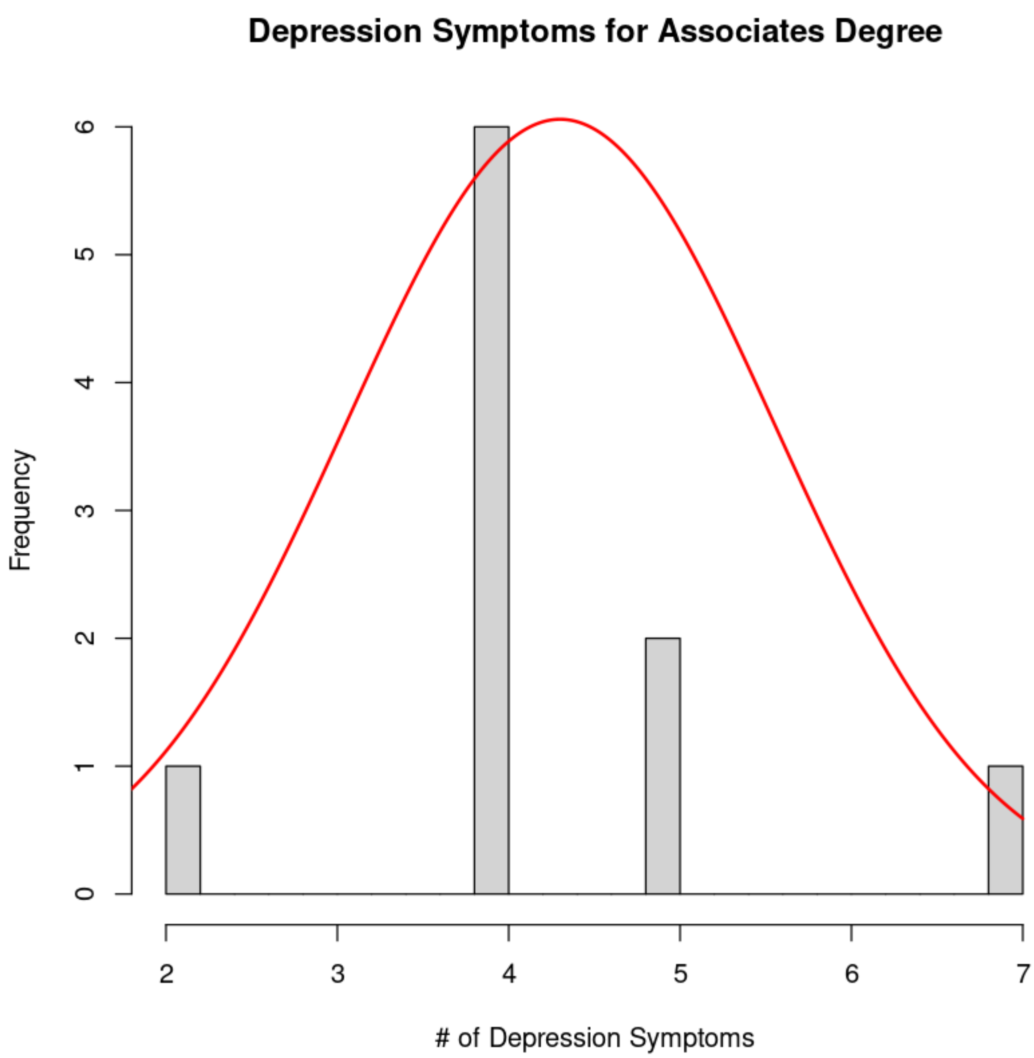
[45]:
```
#Use normal distributions in order to compare the mean values of the subsets␣
↪for depression and children

hist(master$Sad4, # dataframe$vector
    breaks = 20, # breaks sets the approximate number of bars
    main = "Depression Symptoms for Masters Degree", # graph title
    xlab = "# of Depression Symptoms", # x-axis title
    ylab = "Frequency",
     xlim = c(0,7)# y-axis title
    )
```

```
x <- seq(0,7,0.01)


lines(x,                # x coordinate
      30*dnorm(x, mean = 2.846, sd = 1.95),    # first peak
      col = "red",
      lwd = 2
      )
```



[44]:
```
hist(associate$Sad4, # dataframe$vector
    breaks = 20, # breaks sets the approximate number of bars
    main = "Depression Symptoms for Associates Degree", # graph title
    xlab = "# of Depression Symptoms", # x-axis title
```

```r
    ylab = "Frequency" # y-axis title
    )

x <- seq(0,7,0.01)

lines(x,                # x coordinate
      19*dnorm(x, mean = 4.3, sd = 1.251),    # first peak
      col = "red",
      lwd = 2
      )
```



[16]:
```
#Many people who had an associates degree did not report depression symptoms.

length(associate$Sad4)
length(master$Sad4)

#All I am trying to do is convert these histograms into distributions so that I␣
↪can plot them all into one and it will
# have 4 distributions, 2 for children, 2 for depression symptoms, and all for␣
↪the masters and associates degree.
```

154

548

[18]:
```
#Also add another conditions where you only include the

associate$Sad4
```

1. <NA> 2. <NA> 3. <NA> 4. <NA> 5. 5 6. <NA> 7. 5 8. <NA> 9. <NA> 10. <NA> 11. <NA> 12. <NA> 13. <NA> 14. <NA> 15. <NA> 16. <NA> 17. <NA> 18. <NA> 19. <NA> 20. 7 21. <NA> 22. <NA> 23. <NA> 24. <NA> 25. <NA> 26. 2 27. <NA> 28. <NA> 29. <NA> 30. <NA> 31. <NA> 32. <NA> 33. <NA> 34. <NA> 35. <NA> 36. <NA> 37. <NA> 38. <NA> 39. <NA> 40. <NA> 41. <NA> 42. <NA> 43. <NA> 44. <NA> 45. <NA> 46. <NA> 47. <NA> 48. <NA> 49. <NA> 50. <NA> 51. <NA> 52. <NA> 53. <NA> 54. <NA> 55. <NA> 56. <NA> 57. <NA> 58. <NA> 59. <NA> 60. <NA> 61. <NA> 62. <NA> 63. <NA> 64. <NA> 65. <NA> 66. <NA> 67. <NA> 68. <NA> 69. <NA> 70. <NA> 71. <NA> 72. <NA> 73. <NA> 74. <NA> 75. <NA> 76. 4 77. <NA> 78. <NA> 79. <NA> 80. <NA> 81. <NA> 82. <NA> 83. <NA> 84. <NA> 85. <NA> 86. <NA> 87. <NA> 88. <NA> 89. <NA> 90. <NA> 91. <NA> 92. 4 93. <NA> 94. <NA> 95. <NA> 96. <NA> 97. <NA> 98. 4 99. <NA> 100. <NA> 101. <NA> 102. <NA> 103. <NA> 104. <NA> 105. <NA> 106. <NA> 107. <NA> 108. <NA> 109. <NA> 110. <NA> 111. <NA> 112. <NA> 113. 4 114. <NA> 115. <NA> 116. <NA> 117. <NA> 118. <NA> 119. <NA> 120. <NA> 121. <NA> 122. <NA> 123. <NA> 124. <NA> 125. <NA> 126. <NA> 127. <NA> 128. <NA> 129. <NA> 130. <NA> 131. <NA> 132. <NA> 133. <NA> 134. <NA> 135. 4 136. 4 137. <NA> 138. <NA> 139. <NA> 140. <NA> 141. <NA> 142. <NA> 143. <NA> 144. <NA> 145. <NA> 146. <NA> 147. <NA> 148. <NA> 149. <NA> 150. <NA> 151. <NA> 152. <NA> 153. <NA> 154. <NA>

[39]:
```
sd(associate$Sad4, na.rm = TRUE)
```

1.25166555703457

[ ]:
```
hist(master1$Children, # dataframe$vector
    breaks = 20, # breaks sets the approximate number of bars
    main = "# of Children for Masters Degree", # graph title
    xlab = "# of Depression Symptoms", # x-axis title
    ylab = "Frequency" # y-axis title
    )

hist(associate$Sad4, # dataframe$vector
```

```
    breaks = 20, # breaks sets the approximate number of bars
    main = "Depression Symptoms for Associates Degree", # graph title
    xlab = "# of Depression Symptoms", # x-axis title
    ylab = "Frequency" # y-axis title
    )

hist(associate1$Children, # dataframe$vector
    breaks = 20, # breaks sets the approximate number of bars
    main = "# of Children for Associates Degree", # graph title
    xlab = "# of Depression Symptoms", # x-axis title
    ylab = "Frequency" # y-axis title
    )
```

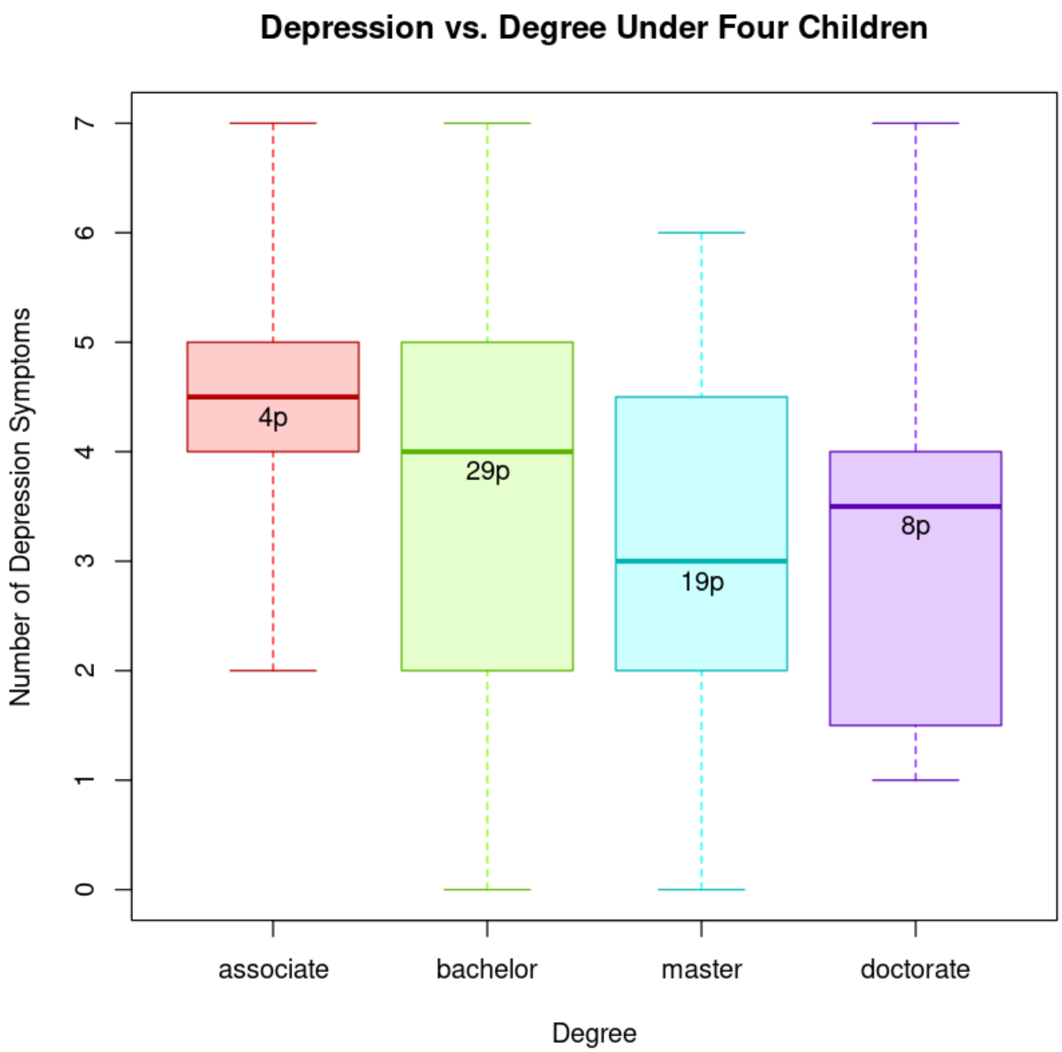
[ ]:
```
#RESET FOR HISTOGRAMS WITH LINES. CREATE A PLOT with 4 means and alternate␣
↪colors to show means.
```

[14]:
```
associate <- subset(WLS, Degree == 1)
bachelor <- subset(WLS, Degree == 2)
master <- subset(WLS, Degree == 3)
doctor <- subset(WLS, Degree == 4)


associate_less <- subset(associate, Children < 3)
bachelor_less <- subset(bachelor, Children < 3)
master_less <- subset(master, Children < 3)
doctorate_less <- subset(doctor, Children < 3)






t.test(associate_less$Children,master_less$Children,
       conf.level = 0.99,
       var.equal = TRUE)
```

```
	Two Sample t-test

data:  associate_less$Children and master_less$Children
t = 2.2558, df = 265, p-value = 0.0249
alternative hypothesis: true difference in means is not equal to 0
99 percent confidence interval:
 -0.05014971  0.71816590
sample estimates:
mean of x mean of y
 1.641026  1.307018
```

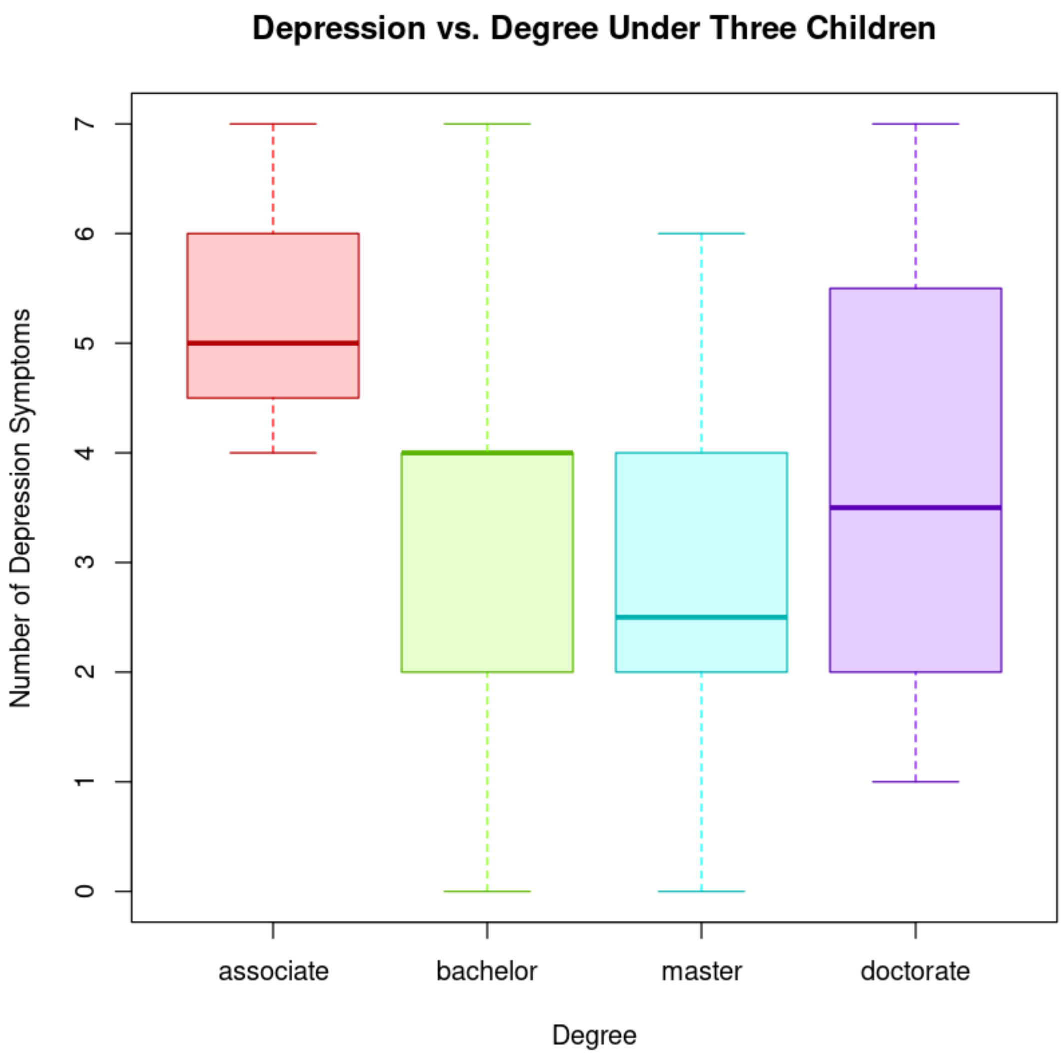
[6]:
```r
WLS99 <- subset(WLS, Children < 4)



associate99 <- subset(WLS99, Degree == 1)
bachelor99 <- subset(WLS99, Degree == 2)
master99 <- subset(WLS99, Degree == 3)
doctor99 <- subset(WLS99, Degree == 4)

c1 <- rainbow(4)
c2 <- rainbow(4, alpha=0.2)
c3 <- rainbow(4, v=0.7)

boxplot(Sad4~Degree,
        data = WLS99,
        names = c("associate", "bachelor", "master", "doctorate"),
        xlab = "Degree",
        ylab = "Number of Depression Symptoms",
        main = "Depression vs. Degree Under Four Children",
        range = 8,
       col = c2,
       medcol = c3,
       whiskcol = c1,
       staplecol= c3,
       boxcol = c3,
       outcol = c3,
       pch = 23,
       cex = 2)

text(1,4.3, # same x but y slightly lower
     labels = "4p",
     cex = 1,
     col = "black"
     )

text(2,3.8, # same x but y slightly lower
     labels = "29p",
     cex = 1,
     col = "black"
     )

text(3,2.8, # same x but y slightly lower
     labels = "19p",
     cex = 1,
     col = "black"
     )
```

```
text(4,3.3, # same x but y slightly lower
     labels = "8p",
     cex = 1,
     col = "black"
     )
```



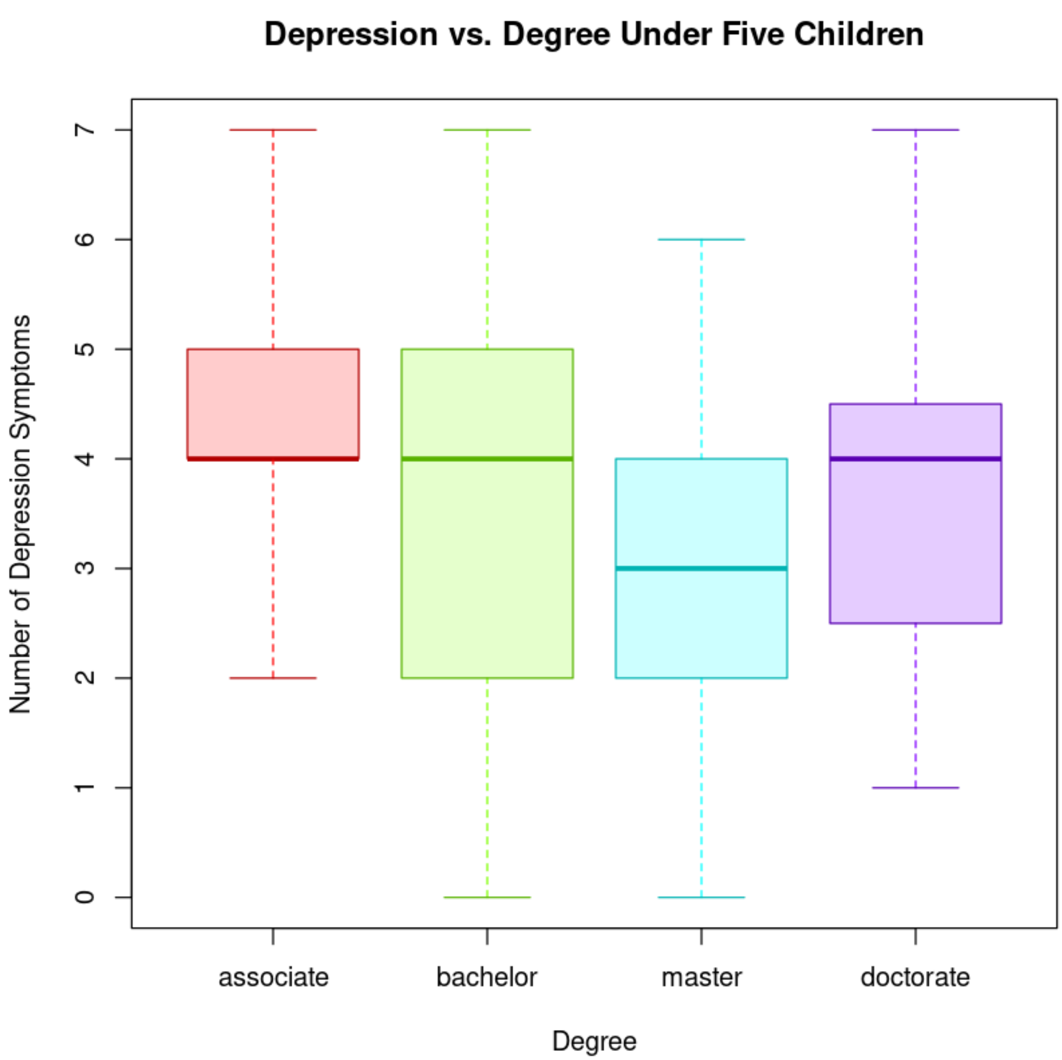
[10]:
```
WLSA <- subset(WLS, Children < 3)



associate99A <- subset(WLSA, Degree == 1)
bachelor99A <- subset(WLSA, Degree == 2)
master99A <- subset(WLSA, Degree == 3)
```

```r
doctor99A <- subset(WLSA, Degree == 4)

c1 <- rainbow(4)
c2 <- rainbow(4, alpha=0.2)
c3 <- rainbow(4, v=0.7)

boxplot(Sad4~Degree,
        data = WLSA,
        names = c("associate", "bachelor", "master", "doctorate"),
        xlab = "Degree",
        ylab = "Number of Depression Symptoms",
        main = "Depression vs. Degree Under Three Children",
        range = 8,
       col = c2,
       medcol = c3,
       whiskcol = c1,
       staplecol= c3,
       boxcol = c3,
       outcol = c3,
       pch = 23,
       cex = 2)
```

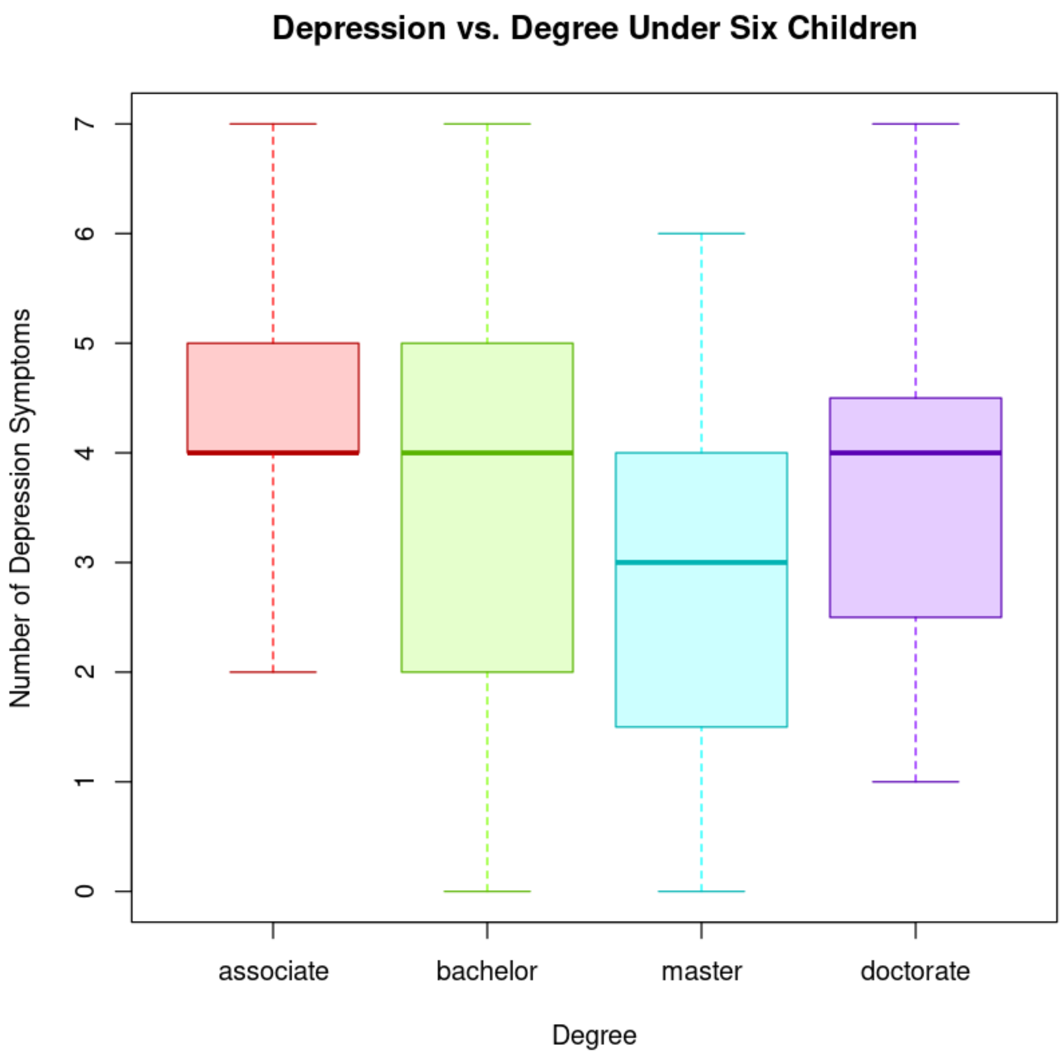


[9]:
```
WLSB <- subset(WLS, Children < 5)



associate99B <- subset(WLSB, Degree == 1)
bachelor99B <- subset(WLSB, Degree == 2)
master99B <- subset(WLSB, Degree == 3)
doctor99B <- subset(WLSB, Degree == 4)

c1 <- rainbow(4)
c2 <- rainbow(4, alpha=0.2)
c3 <- rainbow(4, v=0.7)
```

```r
boxplot(Sad4~Degree,
        data = WLSB,
        names = c("associate", "bachelor", "master", "doctorate"),
        xlab = "Degree",
        ylab = "Number of Depression Symptoms",
        main = "Depression vs. Degree Under Five Children",
        range = 8,
       col = c2,
       medcol = c3,
       whiskcol = c1,
       staplecol= c3,
       boxcol = c3,
       outcol = c3,
       pch = 23,
       cex = 2)
```

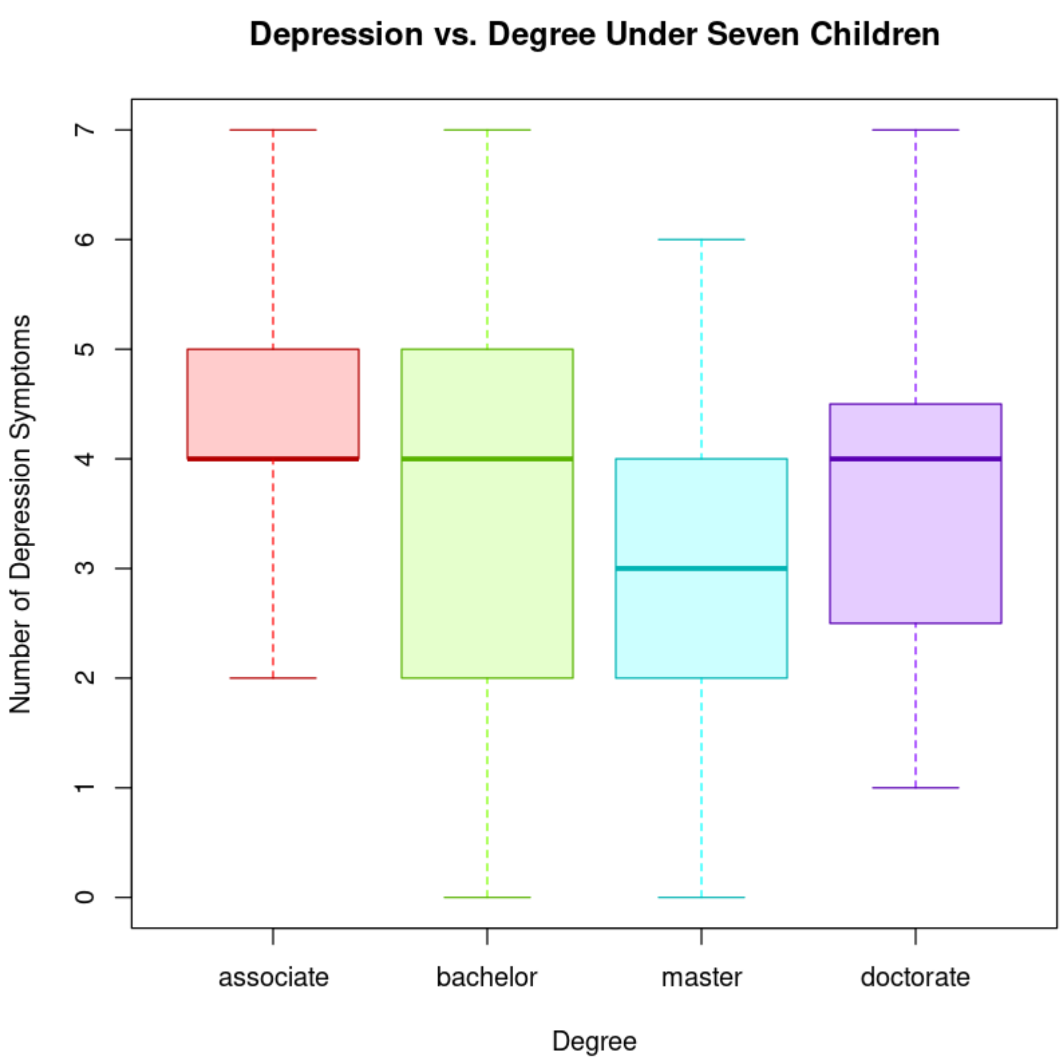


[11]:
```
WLSC <- subset(WLS, Children < 6)



associate99C <- subset(WLSC, Degree == 1)
bachelor99C <- subset(WLSC, Degree == 2)
master99C <- subset(WLSC, Degree == 3)
doctor99C <- subset(WLSC, Degree == 4)

c1 <- rainbow(4)
c2 <- rainbow(4, alpha=0.2)
c3 <- rainbow(4, v=0.7)

boxplot(Sad4~Degree,
        data = WLSC,
        names = c("associate", "bachelor", "master", "doctorate"),
        xlab = "Degree",
        ylab = "Number of Depression Symptoms",
        main = "Depression vs. Degree Under Six Children",
        range = 8,
       col = c2,
       medcol = c3,
       whiskcol = c1,
       staplecol= c3,
       boxcol = c3,
       outcol = c3,
       pch = 23,
       cex = 2)
```

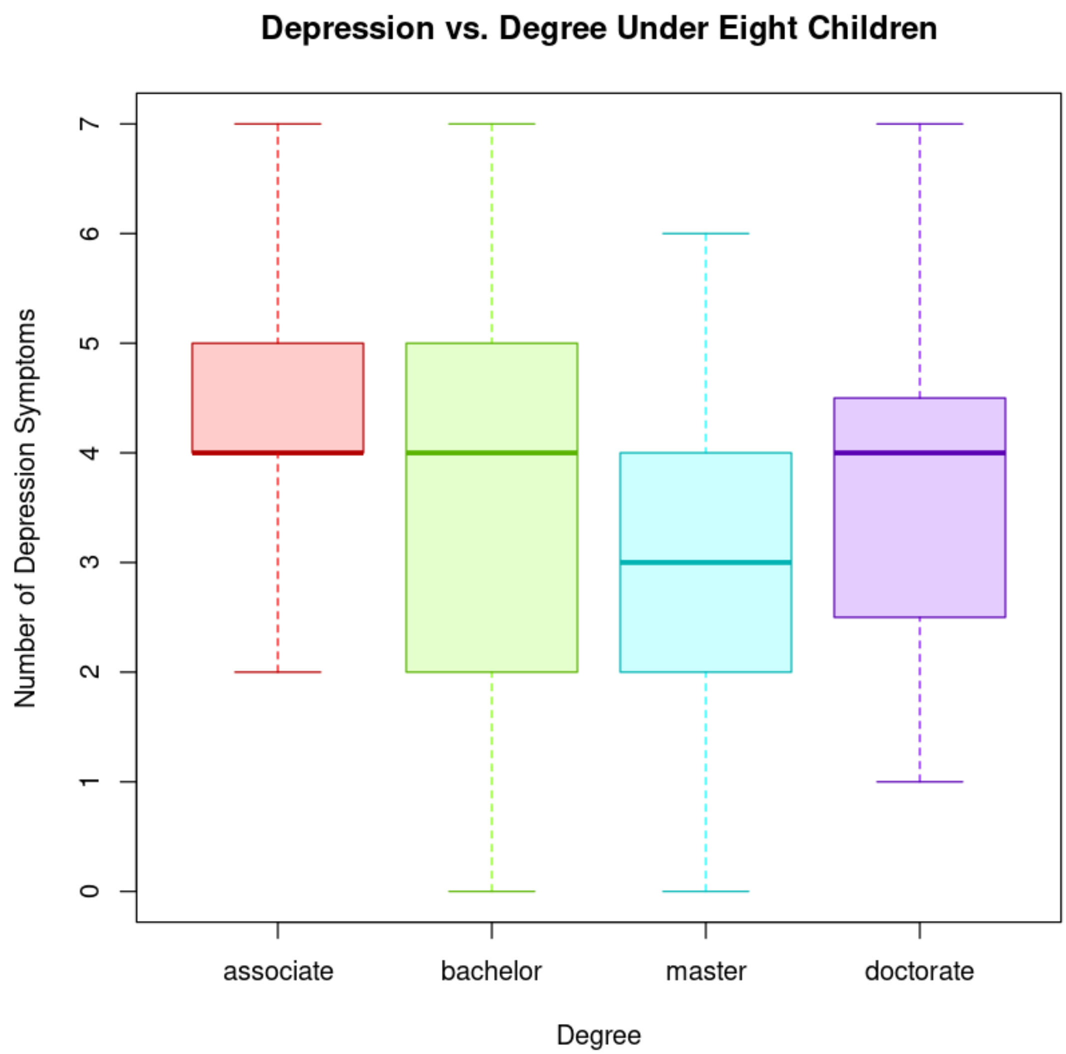


[12]:
```
WLSD <- subset(WLS, Children < 7)



associate99D <- subset(WLSD, Degree == 1)
bachelor99D <- subset(WLSD, Degree == 2)
master99D <- subset(WLSD, Degree == 3)
doctor99D <- subset(WLSD, Degree == 4)

c1 <- rainbow(4)
c2 <- rainbow(4, alpha=0.2)
c3 <- rainbow(4, v=0.7)
```

```r
boxplot(Sad4~Degree,
        data = WLSD,
        names = c("associate", "bachelor", "master", "doctorate"),
        xlab = "Degree",
        ylab = "Number of Depression Symptoms",
        main = "Depression vs. Degree Under Seven Children",
        range = 8,
       col = c2,
       medcol = c3,
       whiskcol = c1,
       staplecol= c3,
       boxcol = c3,
       outcol = c3,
       pch = 23,
       cex = 2)
```



[13]:
```
WLSE <- subset(WLS, Children < 8)



associate99E <- subset(WLSE, Degree == 1)
bachelor99E <- subset(WLSE, Degree == 2)
master99E <- subset(WLSE, Degree == 3)
doctor99E <- subset(WLSE, Degree == 4)

c1 <- rainbow(4)
c2 <- rainbow(4, alpha=0.2)
c3 <- rainbow(4, v=0.7)

boxplot(Sad4~Degree,
        data = WLSE,
        names = c("associate", "bachelor", "master", "doctorate"),
        xlab = "Degree",
        ylab = "Number of Depression Symptoms",
        main = "Depression vs. Degree Under Eight Children",
        range = 8,
       col = c2,
       medcol = c3,
       whiskcol = c1,
       staplecol= c3,
       boxcol = c3,
       outcol = c3,
       pch = 23,
       cex = 2)
```



[14]:
```
#Under 3 Children
t.test(associate99$Children,master99$Children,
       conf.level = 0.99,
       alternative = "g",
       var.equal = TRUE)

#Under 4 Children
t.test(associate99A$Children,master99A$Children,
       conf.level = 0.99,
       alternative = "g",
       var.equal = TRUE)

#Under 5 Children
```

```r
t.test(associate99B$Children,master99B$Children,
       conf.level = 0.99,
       alternative = "g",
       var.equal = TRUE)

#Under 6 Children
t.test(associate99C$Children,master99C$Children,
       conf.level = 0.99,
       alternative = "g",
       var.equal = TRUE)

#Under 7 Children
t.test(associate99D$Children,master99D$Children,
       conf.level = 0.99,
       alternative = "g",
       var.equal = TRUE)

#Under 8 Children
t.test(associate99E$Children,master99E$Children,
       conf.level = 0.99,
       alternative = "g",
       var.equal = TRUE)
```

```
        Two Sample t-test

data:  associate99$Children and master99$Children
t = 2.6658, df = 377, p-value = 0.004005
alternative hypothesis: true difference in means is greater than 0
99 percent confidence interval:
 0.0470054       Inf
sample estimates:
mean of x mean of y 
 2.158730  1.778481 
```

```
        Two Sample t-test

data:  associate99A$Children and master99A$Children
t = 2.2558, df = 265, p-value = 0.01245
alternative hypothesis: true difference in means is greater than 0
99 percent confidence interval:
 -0.01253996         Inf
sample estimates:
mean of x mean of y 
 1.641026  1.307018 
```

```
        Two Sample t-test

data:  associate99B$Children and master99B$Children
t = 3.6382, df = 452, p-value = 0.0001531
alternative hypothesis: true difference in means is greater than 0
99 percent confidence interval:
 0.1926661       Inf
sample estimates:
mean of x mean of y
 2.635294  2.097561



        Two Sample t-test

data:  associate99C$Children and master99C$Children
t = 3.6776, df = 478, p-value = 0.0001311
alternative hypothesis: true difference in means is greater than 0
99 percent confidence interval:
 0.210242      Inf
sample estimates:
mean of x mean of y
 2.815217  2.239691



        Two Sample t-test

data:  associate99D$Children and master99D$Children
t = 3.9322, df = 496, p-value = 4.808e-05
alternative hypothesis: true difference in means is greater than 0
99 percent confidence interval:
 0.2673313       Inf
sample estimates:
mean of x mean of y
 3.010204  2.352500



        Two Sample t-test

data:  associate99E$Children and master99E$Children
t = 4.0474, df = 508, p-value = 2.993e-05
alternative hypothesis: true difference in means is greater than 0
99 percent confidence interval:
 0.3061423       Inf
sample estimates:
mean of x mean of y
```

```
3.166667  2.443627
```

[15]:
```r
#Under 4 Children
t.test(bachelor99$Children,master99$Children,
       conf.level = 0.99,
       alternative = "g",
       var.equal = TRUE)

#Under 3 Children
t.test(bachelor99A$Children,master99A$Children,
       conf.level = 0.99,
       alternative = "g",
       var.equal = TRUE)

#Under 5 Children
t.test(bachelor99B$Children,master99B$Children,
       conf.level = 0.99,
       alternative = "g",
       var.equal = TRUE)

#Under 6 Children
t.test(bachelor99C$Children,master99C$Children,
       conf.level = 0.99,
       alternative = "g",
       var.equal = TRUE)

#Under 7 Children
t.test(bachelor99D$Children,master99D$Children,
       conf.level = 0.99,
       alternative = "g",
       var.equal = TRUE)

#Under 8 Children
t.test(bachelor99E$Children,master99E$Children,
       conf.level = 0.99,
       alternative = "g",
       var.equal = TRUE)
```

```
        Two Sample t-test

data:  bachelor99$Children and master99$Children
t = 3.7873, df = 1073, p-value = 8.038e-05
alternative hypothesis: true difference in means is greater than 0
99 percent confidence interval:
 0.09792243        Inf
sample estimates:
```

```
mean of x mean of y
 2.032938  1.778481


        Two Sample t-test

data:  bachelor99A$Children and master99A$Children
t = 2.3652, df = 704, p-value = 0.009146
alternative hypothesis: true difference in means is greater than 0
99 percent confidence interval:
 0.002231035         Inf
sample estimates:
mean of x mean of y
 1.464435  1.307018


        Two Sample t-test

data:  bachelor99B$Children and master99B$Children
t = 3.2646, df = 1264, p-value = 0.0005629
alternative hypothesis: true difference in means is greater than 0
99 percent confidence interval:
 0.06818482        Inf
sample estimates:
mean of x mean of y
 2.335563  2.097561


        Two Sample t-test

data:  bachelor99C$Children and master99C$Children
t = 4.0043, df = 1366, p-value = 3.277e-05
alternative hypothesis: true difference in means is greater than 0
99 percent confidence interval:
 0.1345163       Inf
sample estimates:
mean of x mean of y
 2.561224  2.239691


        Two Sample t-test

data:  bachelor99D$Children and master99D$Children
t = 3.8627, df = 1415, p-value = 5.861e-05
alternative hypothesis: true difference in means is greater than 0
99 percent confidence interval:
```

```
 0.132549      Inf
sample estimates:
mean of x mean of y
 2.686332  2.352500



        Two Sample t-test

data:  bachelor99E$Children and master99E$Children
t = 3.3576, df = 1438, p-value = 0.0004033
alternative hypothesis: true difference in means is greater than 0
99 percent confidence interval:
 0.09356767        Inf
sample estimates:
mean of x mean of y
 2.749031  2.443627
```

[16]:

```
#Under 4 Children
t.test(doctor99$Children,master99$Children,
       conf.level = 0.99,
       alternative = "g",
       var.equal = TRUE)

#Under 3 Children
t.test(doctor99A$Children,master99A$Children,
       conf.level = 0.99,
       alternative = "g",
       var.equal = TRUE)

#Under 5 Children
t.test(doctor99B$Children,master99B$Children,
       conf.level = 0.99,
       alternative = "g",
       var.equal = TRUE)

#Under 6 Children
t.test(doctor99C$Children,master99C$Children,
       conf.level = 0.99,
       alternative = "g",
       var.equal = TRUE)

#Under 7 Children
t.test(doctor99D$Children,master99D$Children,
       conf.level = 0.99,
       alternative = "g",
       var.equal = TRUE)
```

```r
#Under 8 Children
t.test(doctor99E$Children,master99E$Children,
       conf.level = 0.99,
       alternative = "g",
       var.equal = TRUE)
```

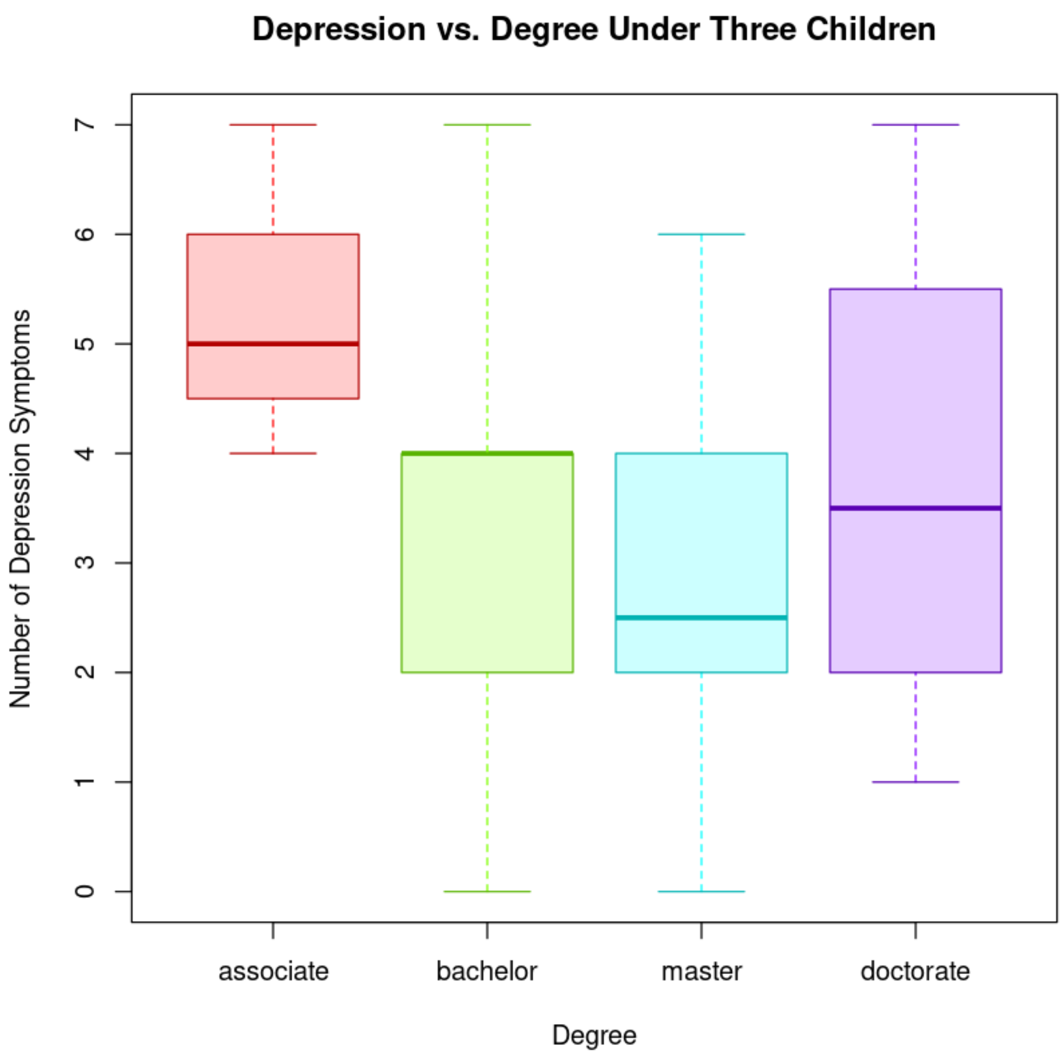
```
        Two Sample t-test

data:  doctor99$Children and master99$Children
t = 2.6402, df = 423, p-value = 0.004295
alternative hypothesis: true difference in means is greater than 0
99 percent confidence interval:
 0.03513418        Inf
sample estimates:
mean of x mean of y
 2.082569  1.778481


        Two Sample t-test

data:  doctor99A$Children and master99A$Children
t = 1.4896, df = 292, p-value = 0.0687
alternative hypothesis: true difference in means is greater than 0
99 percent confidence interval:
 -0.1014176        Inf
sample estimates:
mean of x mean of y
 1.484848  1.307018


        Two Sample t-test

data:  doctor99B$Children and master99B$Children
t = 3.0665, df = 504, p-value = 0.001141
alternative hypothesis: true difference in means is greater than 0
99 percent confidence interval:
 0.09005374        Inf
sample estimates:
mean of x mean of y
 2.474453  2.097561


        Two Sample t-test
```

```
data:  doctor99C$Children and master99C$Children
t = 2.5997, df = 529, p-value = 0.004796
alternative hypothesis: true difference in means is greater than 0
99 percent confidence interval:
 0.03489662        Inf
sample estimates:
mean of x mean of y
 2.580420  2.239691



        Two Sample t-test

data:  doctor99D$Children and master99D$Children
t = 3.0414, df = 550, p-value = 0.001233
alternative hypothesis: true difference in means is greater than 0
99 percent confidence interval:
 0.1002285       Inf
sample estimates:
mean of x mean of y
 2.782895  2.352500



        Two Sample t-test

data:  doctor99E$Children and master99E$Children
t = 2.7892, df = 561, p-value = 0.002731
alternative hypothesis: true difference in means is greater than 0
99 percent confidence interval:
 0.06884338        Inf
sample estimates:
mean of x mean of y
 2.864516  2.443627
```

[84]:

```
WLS3 <- subset(WLS, Children < 3)



associate <- subset(WLS3, Degree == 1)
bachelor <- subset(WLS3, Degree == 2)
master <- subset(WLS3, Degree == 3)
doctor <- subset(WLS3, Degree == 4)

c1 <- rainbow(4)
c2 <- rainbow(4, alpha=0.2)
c3 <- rainbow(4, v=0.7)
```

```r
boxplot(Sad4~Degree,
        data = WLS3,
        names = c("associate", "bachelor", "master", "doctorate"),
        xlab = "Degree",
        ylab = "Number of Depression Symptoms",
        main = "Depression vs. Degree Under Three Children",
        range = 8,
       col = c2,
       medcol = c3,
       whiskcol = c1,
       staplecol= c3,
       boxcol = c3,
       outcol = c3,
       pch = 23,
       cex = 2)
```



[87]:
```
associate$Sad4 #3
bachelor$Sad4 #28
master$Sad4 #12
doctor$Sad4 #4
```

1. <NA> 2. <NA> 3. 5 4. <NA> 5. <NA> 6. <NA> 7. 7 8. <NA> 9. <NA> 10. <NA> 11. <NA> 12. <NA> 13. <NA> 14. <NA> 15. <NA> 16. <NA> 17. <NA> 18. <NA> 19. <NA> 20. <NA> 21. <NA> 22. <NA> 23. <NA> 24. <NA> 25. <NA> 26. <NA> 27. <NA> 28. <NA> 29. <NA> 30. 4 31. <NA> 32. <NA> 33. <NA> 34. <NA> 35. <NA> 36. <NA> 37. <NA> 38. <NA> 39. <NA>

1. <NA> 2. <NA> 3. <NA> 4. <NA> 5. <NA> 6. <NA> 7. <NA> 8. <NA> 9. <NA> 10. <NA> 11. <NA> 12. <NA> 13. <NA> 14. <NA> 15. <NA> 16. <NA> 17. <NA> 18. 3 19. <NA> 20. 4 21. <NA> 22. <NA> 23. <NA> 24. <NA> 25. <NA> 26. <NA> 27. <NA> 28. <NA> 29. <NA> 30. <NA> 31. <NA> 32. <NA> 33. <NA> 34. <NA> 35. <NA> 36. <NA> 37. <NA> 38. <NA> 39. 1 40. <NA> 41. <NA> 42. <NA> 43. <NA> 44. <NA> 45. <NA> 46. <NA> 47. <NA> 48. <NA> 49. <NA> 50. <NA> 51. <NA> 52. <NA> 53. <NA> 54. <NA> 55. <NA> 56. <NA> 57. 5 58. <NA> 59. <NA> 60. 4 61. <NA> 62. <NA> 63. <NA> 64. <NA> 65. <NA> 66. <NA> 67. <NA> 68. <NA> 69. <NA> 70. <NA> 71. <NA> 72. <NA> 73. <NA> 74. <NA> 75. <NA> 76. <NA> 77. <NA> 78. <NA> 79. <NA> 80. <NA> 81. <NA> 82. <NA> 83. <NA> 84. <NA> 85. <NA> 86. <NA> 87. <NA> 88. <NA> 89. <NA> 90. <NA> 91. <NA> 92. <NA> 93. <NA> 94. 2 95. <NA> 96. <NA> 97. <NA> 98. <NA> 99. <NA> 100. <NA> 101. <NA> 102. <NA> 103. <NA> 104. 3 105. <NA> 106. <NA> 107. <NA> 108. 6 109. <NA> 110. 5 111. <NA> 112. <NA> 113. <NA> 114. <NA> 115. <NA> 116. <NA> 117. <NA> 118. <NA> 119. <NA> 120. <NA> 121. <NA> 122. <NA> 123. <NA> 124. <NA> 125. <NA> 126. <NA> 127. <NA> 128. <NA> 129. <NA> 130. <NA> 131. <NA> 132. <NA> 133. 5 134. <NA> 135. <NA> 136. <NA> 137. <NA> 138. <NA> 139. <NA> 140. <NA> 141. <NA> 142. <NA> 143. <NA> 144. <NA> 145. <NA> 146. <NA> 147. <NA> 148. <NA> 149. <NA> 150. <NA> 151. <NA> 152. <NA> 153. <NA> 154. <NA> 155. <NA> 156. <NA> 157. <NA> 158. <NA> 159. <NA> 160. 4 161. <NA> 162. 3 163. <NA> 164. <NA> 165. 1 166. <NA> 167. <NA> 168. <NA> 169. <NA> 170. <NA> 171. <NA> 172. 4 173. <NA> 174. <NA> 175. <NA> 176. <NA> 177. 4 178. <NA> 179. <NA> 180. <NA> 181. <NA> 182. <NA> 183. <NA> 184. <NA> 185. 7 186. <NA> 187. <NA> 188. <NA> 189. <NA> 190. <NA> 191. <NA> 192. <NA> 193. <NA> 194. <NA> 195. <NA> 196. <NA> 197. <NA> 198. <NA> 199. <NA> 200. <NA> 201. 202. <NA> 203. <NA> 204. <NA> 205. <NA> 206. <NA> 207. <NA> 208. <NA> 209. <NA> 210. <NA> 211. <NA> 212. <NA> 213. <NA> 214. <NA> 215. <NA> 216. <NA> 217. <NA> 218. <NA> 219. <NA> 220. <NA> 221. <NA> 222. <NA> 223. <NA> 224. <NA> 225. <NA> 226. <NA> 227. <NA> 228. <NA> 229. <NA> 230. <NA> 231. <NA> 232. <NA> 233. <NA> 234. <NA> 235. <NA> 236. <NA> 237. <NA> 238. <NA> 239. <NA> 240. <NA> 241. 2 242. <NA> 243. <NA> 244. <NA> 245. <NA> 246. <NA> 247. <NA> 248. <NA> 249. <NA> 250. 3 251. <NA> 252. <NA> 253. <NA> 254. <NA> 255. <NA> 256. <NA> 257. 1 258. <NA> 259. <NA> 260. <NA> 261. <NA> 262. <NA> 263. 0 264. <NA> 265. <NA> 266. <NA> 267. <NA> 268. <NA> 269. <NA> 270. <NA> 271. <NA> 272. 2 273. <NA> 274. <NA> 275. <NA> 276. <NA> 277. <NA> 278. <NA> 279. <NA> 280. <NA> 281. <NA> 282. <NA> 283. <NA> 284. <NA> 285. <NA> 286. <NA>

287. <NA> 288. <NA> 289. <NA> 290. <NA> 291. <NA> 292. <NA> 293. <NA> 294. <NA> 295. 4 296. <NA> 297. <NA> 298. <NA> 299. <NA> 300. <NA> 301. <NA> 302. <NA> 303. <NA> 304. <NA> 305. <NA> 306. <NA> 307. <NA> 308. <NA> 309. <NA> 310. <NA> 311. <NA> 312. <NA> 313. <NA> 314. <NA> 315. <NA> 316. <NA> 317. <NA> 318. <NA> 319. <NA> 320. <NA> 321. <NA> 322. <NA> 323. <NA> 324. <NA> 325. <NA> 326. 4 327. <NA> 328. <NA> 329. <NA> 330. 7 331. <NA> 332. <NA> 333. <NA> 334. <NA> 335. <NA> 336. <NA> 337. <NA> 338. <NA> 339. <NA> 340. <NA> 341. <NA> 342. <NA> 343. <NA> 344. <NA> 345. 3 346. <NA> 347. <NA> 348. <NA> 349. <NA> 350. <NA> 351. <NA> 352. <NA> 353. <NA> 354. <NA> 355. <NA> 356. <NA> 357. <NA> 358. <NA> 359. <NA> 360. <NA> 361. <NA> 362. <NA> 363. 4 364. <NA> 365. <NA> 366. <NA> 367. <NA> 368. <NA> 369. <NA> 370. <NA> 371. <NA> 372. <NA> 373. <NA> 374. <NA> 375. <NA> 376. <NA> 377. 4 378. <NA> 379. <NA> 380. <NA> 381. <NA> 382. 2 383. <NA> 384. <NA> 385. <NA> 386. <NA> 387. <NA> 388. <NA> 389. <NA> 390. <NA> 391. <NA> 392. <NA> 393. <NA> 394. <NA> 395. <NA> 396. <NA> 397. <NA> 398. <NA> 399. <NA> 400. <NA> 401. <NA>

1. <NA> 2. <NA> 3. <NA> 4. <NA> 5. <NA> 6. <NA> 7. <NA> 8. <NA> 9. <NA> 10. <NA> 11. <NA> 12. <NA> 13. <NA> 14. <NA> 15. <NA> 16. <NA> 17. <NA> 18. <NA> 19. <NA> 20. <NA> 21. <NA> 22. <NA> 23. <NA> 24. <NA> 25. <NA> 26. <NA> 27. <NA> 28. <NA> 29. 3 30. <NA> 31. <NA> 32. <NA> 33. 0 34. <NA> 35. <NA> 36. 4 37. <NA> 38. <NA> 39. <NA> 40. <NA> 41. <NA> 42. <NA> 43. <NA> 44. <NA> 45. <NA> 46. <NA> 47. <NA> 48. <NA> 49. <NA> 50. <NA> 51. <NA> 52. <NA> 53. <NA> 54. <NA> 55. <NA> 56. <NA> 57. <NA> 58. <NA> 59. <NA> 60. <NA> 61. <NA> 62. <NA> 63. <NA> 64. <NA> 65. <NA> 66. <NA> 67. <NA> 68. <NA> 69. <NA> 70. <NA> 71. <NA> 72. <NA> 73. <NA> 74. <NA> 75. <NA> 76. <NA> 77. <NA> 78. <NA> 79. <NA> 80. <NA> 81. 6 82. <NA> 83. <NA> 84. <NA> 85. <NA> 86. <NA> 87. <NA> 88. <NA> 89. <NA> 90. 3 91. <NA> 92. <NA> 93. <NA> 94. <NA> 95. <NA> 96. <NA> 97. <NA> 98. <NA> 99. <NA> 100. <NA> 101. <NA> 102. <NA> 103. <NA> 104. <NA> 105. <NA> 106. <NA> 107. <NA> 108. <NA> 109. <NA> 110. <NA> 111. <NA> 112. <NA> 113. <NA> 114. <NA> 115. <NA> 116. <NA> 117. <NA> 118. <NA> 119. <NA> 120. <NA> 121. <NA> 122. <NA> 123. <NA> 124. <NA> 125. <NA> 126. <NA> 127. <NA> 128. <NA> 129. <NA> 130. <NA> 131. <NA> 132. <NA> 133. <NA> 134. <NA> 135. <NA> 136. <NA> 137. <NA> 138. <NA> 139. <NA> 140. <NA> 141. <NA> 142. <NA> 143. <NA> 144. <NA> 145. <NA> 146. <NA> 147. <NA> 148. <NA> 149. <NA> 150. 2 151. <NA> 152. <NA> 153. <NA> 154. <NA> 155. <NA> 156. <NA> 157. <NA> 158. <NA> 159. <NA> 160. <NA> 161. <NA> 162. <NA> 163. <NA> 164. <NA> 165. <NA> 166. <NA> 167. <NA> 168. <NA> 169. <NA> 170. <NA> 171. <NA> 172. <NA> 173. <NA> 174. 4 175. <NA> 176. <NA> 177. <NA> 178. 2 179. <NA> 180. <NA> 181. <NA> 182. 6 183. <NA> 184. <NA> 185. <NA> 186. <NA> 187. <NA> 188. <NA> 189. <NA> 190. <NA> 191. <NA> 192. 2 193. <NA> 194. <NA> 195. <NA> 196. <NA> 197. <NA> 198. 0 199. <NA> 200. <NA> 201. <NA> 202. <NA> 203. <NA> 204. <NA> 205. <NA> 206. <NA> 207. <NA> 208. <NA> 209. <NA> 210. <NA> 211. <NA> 212. <NA> 213. 2 214. <NA> 215. <NA> 216. <NA> 217. <NA> 218. <NA> 219. <NA> 220. <NA> 221. <NA> 222. <NA> 223. <NA> 224. <NA> 225. <NA> 226. <NA> 227. <NA> 228. <NA>

1. <NA> 2. 3 3. <NA> 4. <NA> 5. 7 6. <NA> 7. <NA> 8. <NA> 9. <NA> 10. <NA> 11. <NA> 12. <NA> 13. <NA> 14. <NA> 15. <NA> 16. <NA> 17. <NA> 18. <NA> 19. <NA> 20. <NA> 21. <NA> 22. <NA> 23. 4 24. <NA> 25. <NA> 26. <NA> 27. <NA> 28. <NA> 29. <NA> 30. <NA> 31. <NA> 32. <NA> 33. <NA> 34. <NA> 35. <NA> 36. <NA> 37. <NA> 38. <NA> 39. <NA> 40. <NA> 41. <NA> 42. <NA> 43. <NA> 44. <NA> 45. <NA> 46. <NA> 47. <NA>

48. <NA> 49. <NA> 50. <NA> 51. <NA> 52. <NA> 53. <NA> 54. <NA> 55. <NA> 56. <NA> 57. <NA> 58. <NA> 59. <NA> 60. <NA> 61. <NA> 62. <NA> 63. <NA> 64. <NA> 65. 1 66. <NA>

[89]:
```
t.test(associate$Sad4,doctor$Sad4,
       conf.level = 0.99,
       var.equal = TRUE)

var.test(associate$Sad4, doctor$Sad4)

prop.test(x = c(12, 18),
          n = c(47, 60),
          conf.level = 0.99,
          alternative = "t"
          )
```

```
	Two Sample t-test

data:  associate$Sad4 and doctor$Sad4
t = 0.95794, df = 5, p-value = 0.3821
alternative hypothesis: true difference in means is not equal to 0
99 percent confidence interval:
 -5.081232  8.247899
sample estimates:
mean of x mean of y
 5.333333  3.750000



	F test to compare two variances

data:  associate$Sad4 and doctor$Sad4
F = 0.37333, num df = 2, denom df = 3, p-value = 0.567
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  0.02326919 14.62178464
sample estimates:
ratio of variances
         0.3733333



	2-sample test for equality of proportions with continuity correction

data:  c(12, 18) out of c(47, 60)
X-squared = 0.086337, df = 1, p-value = 0.7689
alternative hypothesis: two.sided
99 percent confidence interval:
```

```
 -0.2873992  0.1980375
sample estimates:
   prop 1    prop 2
0.2553191 0.3000000
```

[3]:

```
WLS3 <- subset(WLS, Social4 > 3.5)
WLS4 <- subset(WLS, Social4 < 3.5)
```

[19]:

```
quantile(WLS$Social4, 0.75, na.rm = TRUE)
```

75\%: 5

[6]:

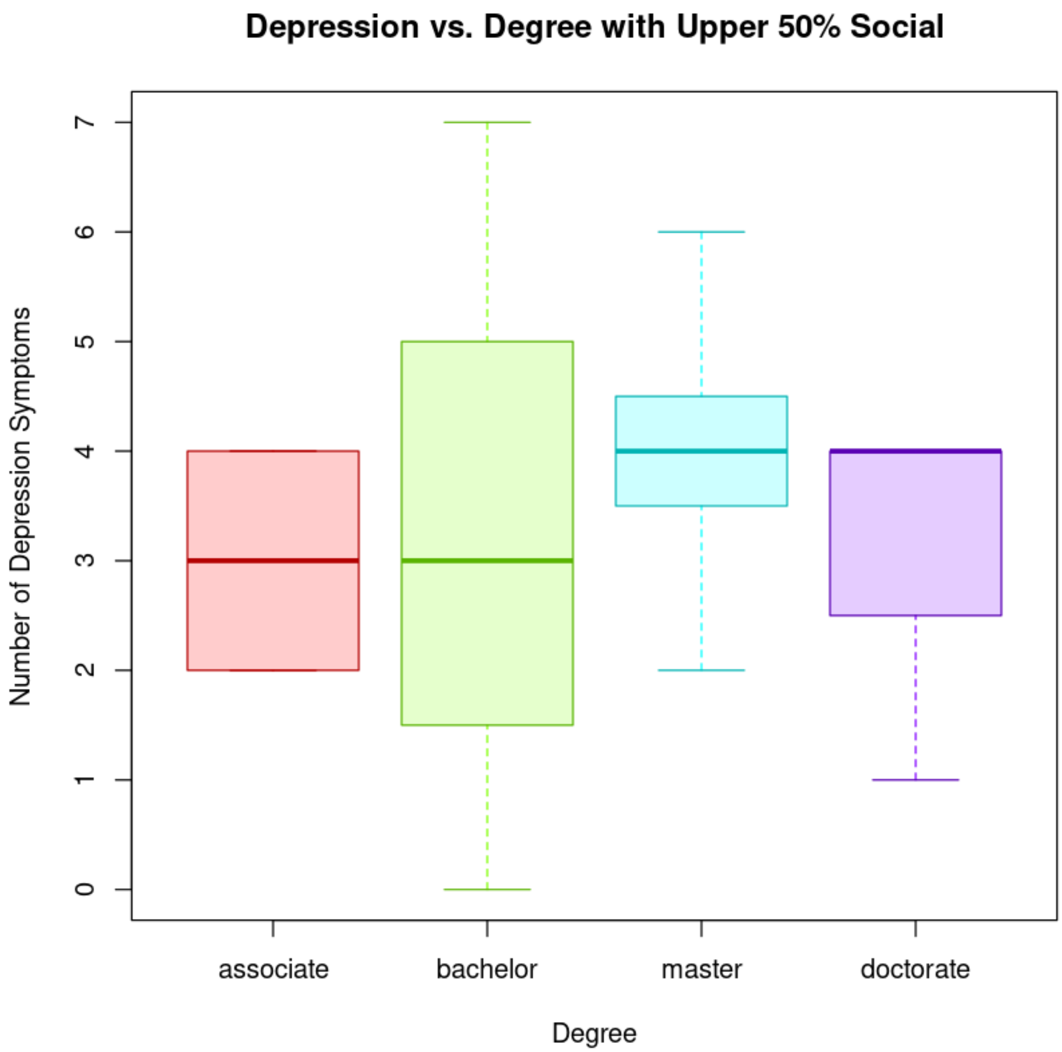
```
associate7 <- subset(WLS3, Degree == 1)
bachelor7 <- subset(WLS3, Degree == 2)
master7 <- subset(WLS3, Degree == 3)
doctor7 <- subset(WLS3, Degree == 4)

c1 <- rainbow(4)
c2 <- rainbow(4, alpha=0.2)
c3 <- rainbow(4, v=0.7)

boxplot(Sad4~Degree,
        data = WLS3,
        names = c("associate", "bachelor", "master", "doctorate"),
        xlab = "Degree",
        ylab = "Number of Depression Symptoms",
        main = "Depression vs. Degree with Upper 50% Social",
        range = 8,
       col = c2,
       medcol = c3,
       whiskcol = c1,
       staplecol= c3,
       boxcol = c3,
       outcol = c3,
       pch = 23,
       cex = 2)

mean(associate7$Children, na.rm = TRUE)
mean(bachelor7$Children, na.rm = TRUE)
mean(master7$Children, na.rm = TRUE)
mean(doctor7$Children, na.rm = TRUE)
```

3.17647058823529

2.78139534883721

2.63333333333333

3.07894736842105



[7]:
```
associate8 <- subset(WLS4, Degree == 1)
bachelor8 <- subset(WLS4, Degree == 2)
master8 <- subset(WLS4, Degree == 3)
doctor8 <- subset(WLS4, Degree == 4)

c1 <- rainbow(4)
c2 <- rainbow(4, alpha=0.2)
c3 <- rainbow(4, v=0.7)

boxplot(Sad4~Degree,
        data = WLS4,
```

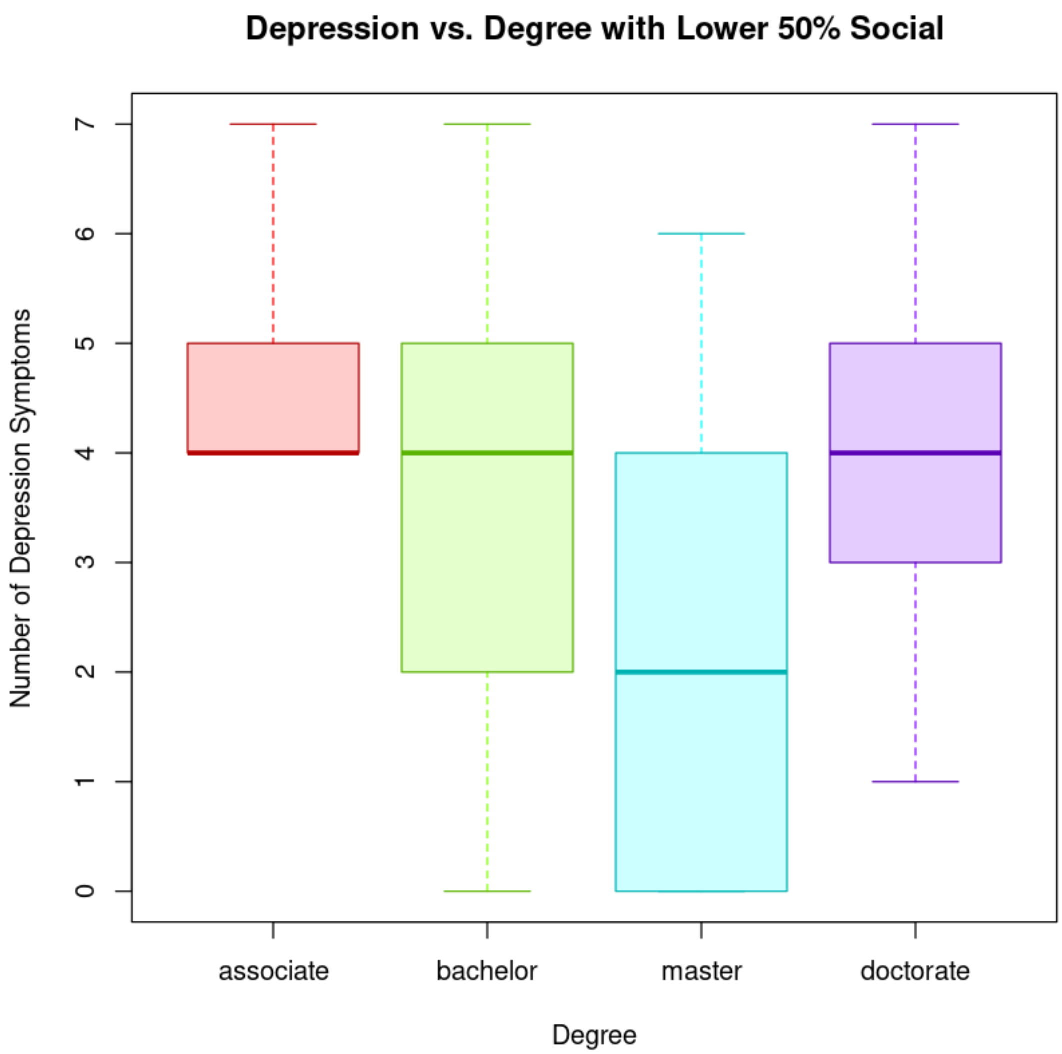
```
        names = c("associate", "bachelor", "master", "doctorate"),
        xlab = "Degree",
        ylab = "Number of Depression Symptoms",
        main = "Depression vs. Degree with Lower 50% Social",
        range = 8,
       col = c2,
       medcol = c3,
       whiskcol = c1,
       staplecol= c3,
       boxcol = c3,
       outcol = c3,
       pch = 23,
       cex = 2)

mean(associate8$Children, na.rm = TRUE)
mean(bachelor8$Children, na.rm = TRUE)
mean(master8$Children, na.rm = TRUE)
mean(doctor8$Children, na.rm = TRUE)
```

3.4

2.80366492146597

2.72258064516129

3.02985074626866



[17]:
```
t.test(associate7$Sad4,associate8$Sad4,
       conf.level = 0.99,
       var.equal = TRUE)

t.test(master7$Sad4,master8$Sad4,
       conf.level = 0.99,
       var.equal = TRUE)

t.test(doctor7$Sad4,doctor8$Sad4,
       conf.level = 0.99,
       var.equal = TRUE)

t.test(associate8$Sad4,master8$Sad4,
```

```r
        conf.level = 0.99,
        alternative = "g",
        var.equal = TRUE)

t.test(associate7$Sad4,master7$Sad4,
        conf.level = 0.99,
        var.equal = TRUE)
```

```
        Two Sample t-test

data:  associate7$Sad4 and associate8$Sad4
t = -1.8501, df = 8, p-value = 0.1015
alternative hypothesis: true difference in means is not equal to 0
99 percent confidence interval:
 -4.572177  1.322177
sample estimates:
mean of x mean of y
    3.000     4.625


        Two Sample t-test

data:  master7$Sad4 and master8$Sad4
t = 2.1483, df = 24, p-value = 0.04199
alternative hypothesis: true difference in means is not equal to 0
99 percent confidence interval:
 -0.5031741  3.8365075
sample estimates:
mean of x mean of y
 4.000000  2.333333


        Two Sample t-test

data:  doctor7$Sad4 and doctor8$Sad4
t = -0.75864, df = 10, p-value = 0.4656
alternative hypothesis: true difference in means is not equal to 0
99 percent confidence interval:
 -4.602270  2.824492
sample estimates:
mean of x mean of y
 3.000000  3.888889


        Two Sample t-test
```

```
data:  associate8$Sad4 and master8$Sad4
t = 2.994, df = 24, p-value = 0.003147
alternative hypothesis: true difference in means is greater than 0
99 percent confidence interval:
 0.3841222       Inf
sample estimates:
mean of x mean of y
 4.625000  2.333333



        Two Sample t-test

data:  associate7$Sad4 and master7$Sad4
t = -1.0328, df = 8, p-value = 0.3319
alternative hypothesis: true difference in means is not equal to 0
99 percent confidence interval:
 -4.24884  2.24884
sample estimates:
mean of x mean of y
        3         4
```

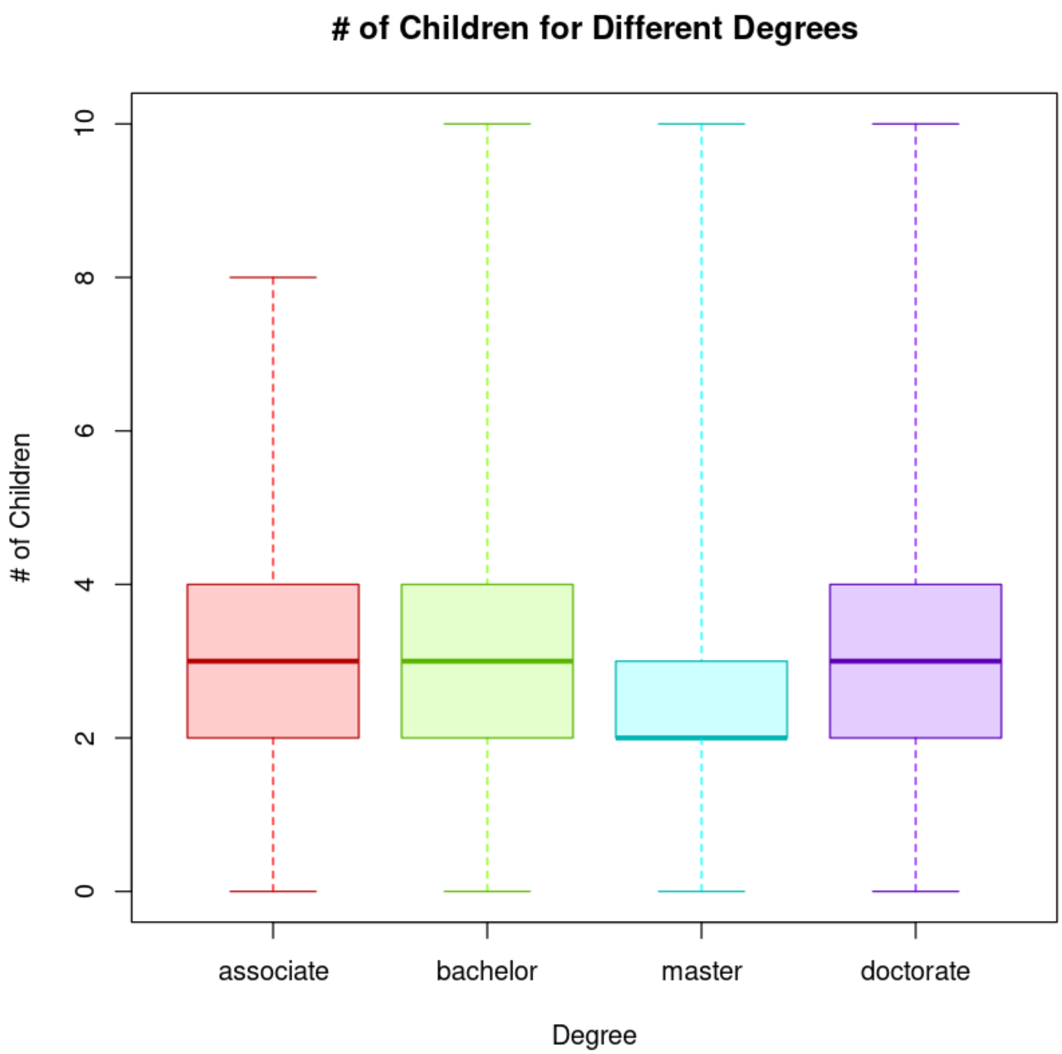
[36]:

```
prop.test(x = c(2, 3584),
          n = c(20, 39065),
          conf.level = 0.99,
          alternative = "g"
          )
```

```
Warning message in prop.test(x = c(2, 3584), n = c(20, 39065), conf.level =
0.99, :
"Chi-squared approximation may be incorrect"


        2-sample test for equality of proportions with continuity correction

data:  c(2, 3584) out of c(20, 39065)
X-squared = 1.8703e-28, df = 1, p-value = 0.5
alternative hypothesis: greater
99 percent confidence interval:
 -0.1560931  1.0000000
sample estimates:
    prop 1     prop 2
0.10000000 0.09174453
```

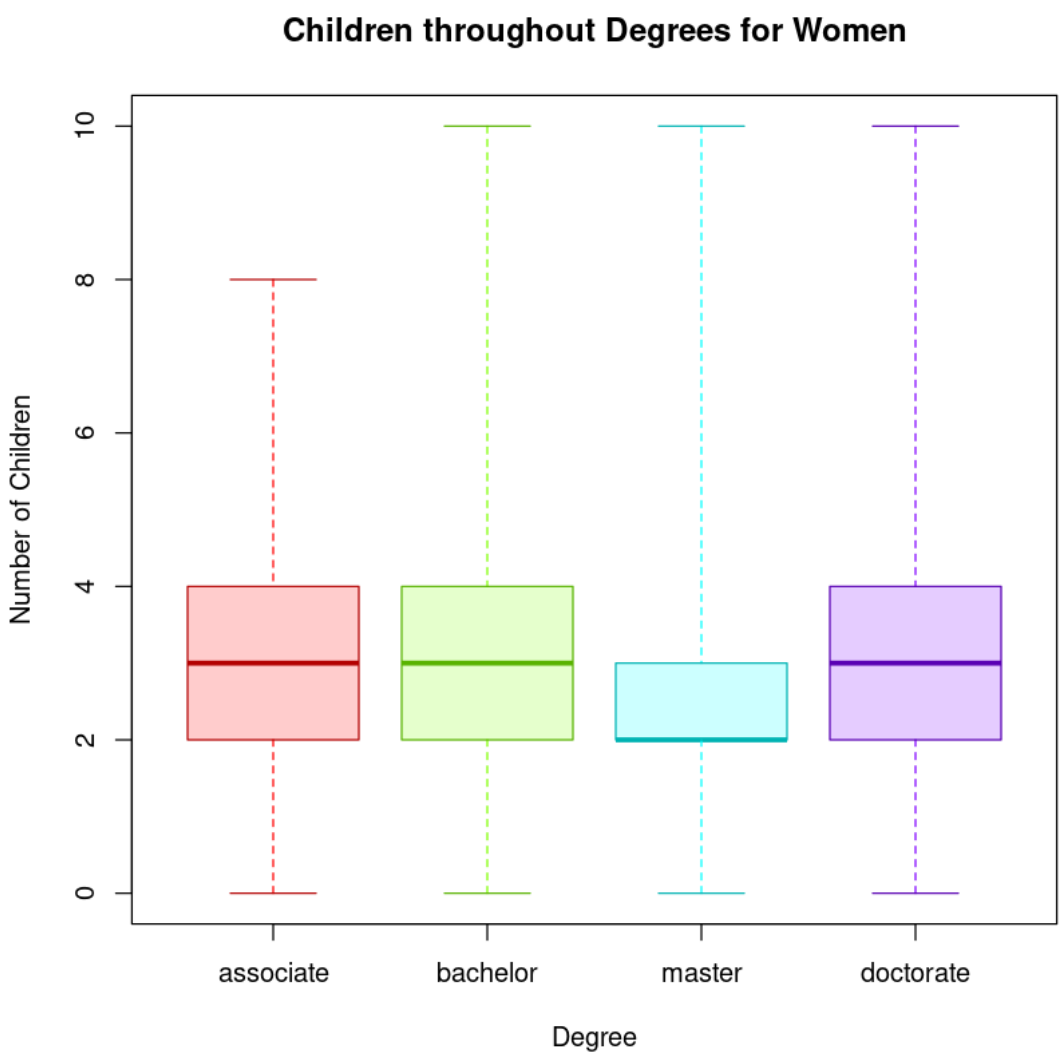
[27]:
```
WL4MALE <- subset(WLS, Gender = 1)
WL4FMALE <- subset(WLS, Gender = 2)
```

[29]:
```
c1 <- rainbow(4)
c2 <- rainbow(4, alpha=0.2)
c3 <- rainbow(4, v=0.7)

boxplot(Children~Degree,
        data = WL4MALE,
        names = c("associate", "bachelor", "master", "doctorate"),
        xlab = "Degree",
        ylab = "# of Children",
        main = "# of Children for Different Degrees",
        range = 8,
       col = c2,
       medcol = c3,
       whiskcol = c1,
       staplecol= c3,
       boxcol = c3,
       outcol = c3,
       pch = 23,
       cex = 2)
```



[31]:
```
c1 <- rainbow(4)
c2 <- rainbow(4, alpha=0.2)
c3 <- rainbow(4, v=0.7)



boxplot(Children~Degree,
        data = WL4FMALE,
        names = c("associate", "bachelor", "master", "doctorate"),
        xlab = "Degree",
        ylab = "Number of Children",
        main = "Children throughout Degrees for Women",
        range = 8,
```

```
        col = c2,
        medcol = c3,
        whiskcol = c1,
        staplecol= c3,
        boxcol = c3,
        outcol = c3,
        pch = 23,
        cex = 2)
```



[15]:
```
mean(female$Children, na.rm = TRUE)
mean(male$Children, na.rm = TRUE)
```

3.23959687906372

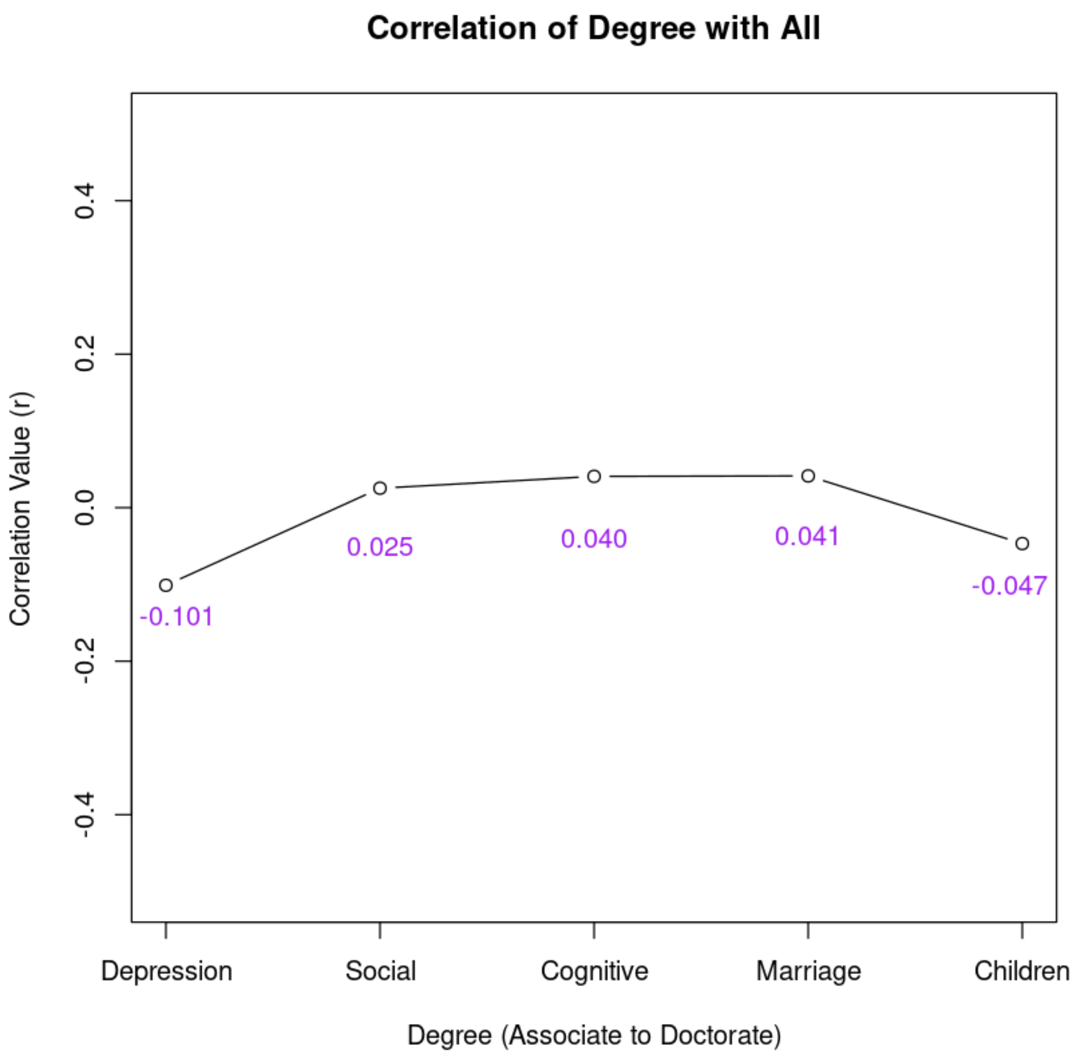
3.23959687906372

[17]:
```
summary(WLS$Gender)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  1.000   1.000   2.000   1.516   2.000   2.000 
```

[74]:
```
b <- cor(WLS$Degree,WLS$Sad4,use="p")
c <- cor(WLS$Degree,WLS$Social4,use="p")
d <- cor(WLS$Degree,WLS$Cog4,use="p")
e <- cor(WLS$Degree,WLS$Marriage,use="p")
f <- cor(WLS$Degree,WLS$Children,use="p")


# Finally, make a plot of these correlation values.  The x-axis simply goes␣
 ↪from 1 to 13,
# one number per month.  Relabel these with dates.

plot(c(1,2,3,4,5),  # x values, made into a vector
     c(b,c,d,e,f), # y values, made into a vector
     type = "b", # this means plot (b)oth lines and points
     ylim = c(-0.5,0.5), # make the y-axis go from OWLS$.5 to 1, you might␣
 ↪adjust this
     xlab = "Degree (Associate to Doctorate)",
     ylab = "Correlation Value (r)",
     main = "Correlation of Degree with All",
     xaxt = "n" # this means (n)o x-axis, because we will add our own with the␣
 ↪names of the months
     )

text(1.05,-0.14, # same x but y slightly lower
     labels = "-0.101",
     cex = 1,
     col = "purple"
     )

text(2,-0.05, # same x but y slightly lower
     labels = "0.025",
     cex = 1,
     col = "purple"
     )

text(3,-0.04, # same x but y slightly lower
     labels = "0.040",
     cex = 1,
     col = "purple"
     )
```

```r
text(4,-0.035, # same x but y slightly lower
     labels = "0.041",
     cex = 1,
     col = "purple"
     )

text(4.94,-0.1, # same x but y slightly lower
     labels = "-0.047",
     cex = 1,
     col = "purple"
     )

axis(1,  # 1 is the bottom axis
     at = c(1,2,3,4,5),  # for space reasons, plot just 5 months at intervals␣
 ↪of three
     label = c("Depression","Social","Cognitive","Marriage","Children")
     )


#With more education, you do not get happier are sadder, you do not make more␣
 ↪children, you do not get smarter,
```

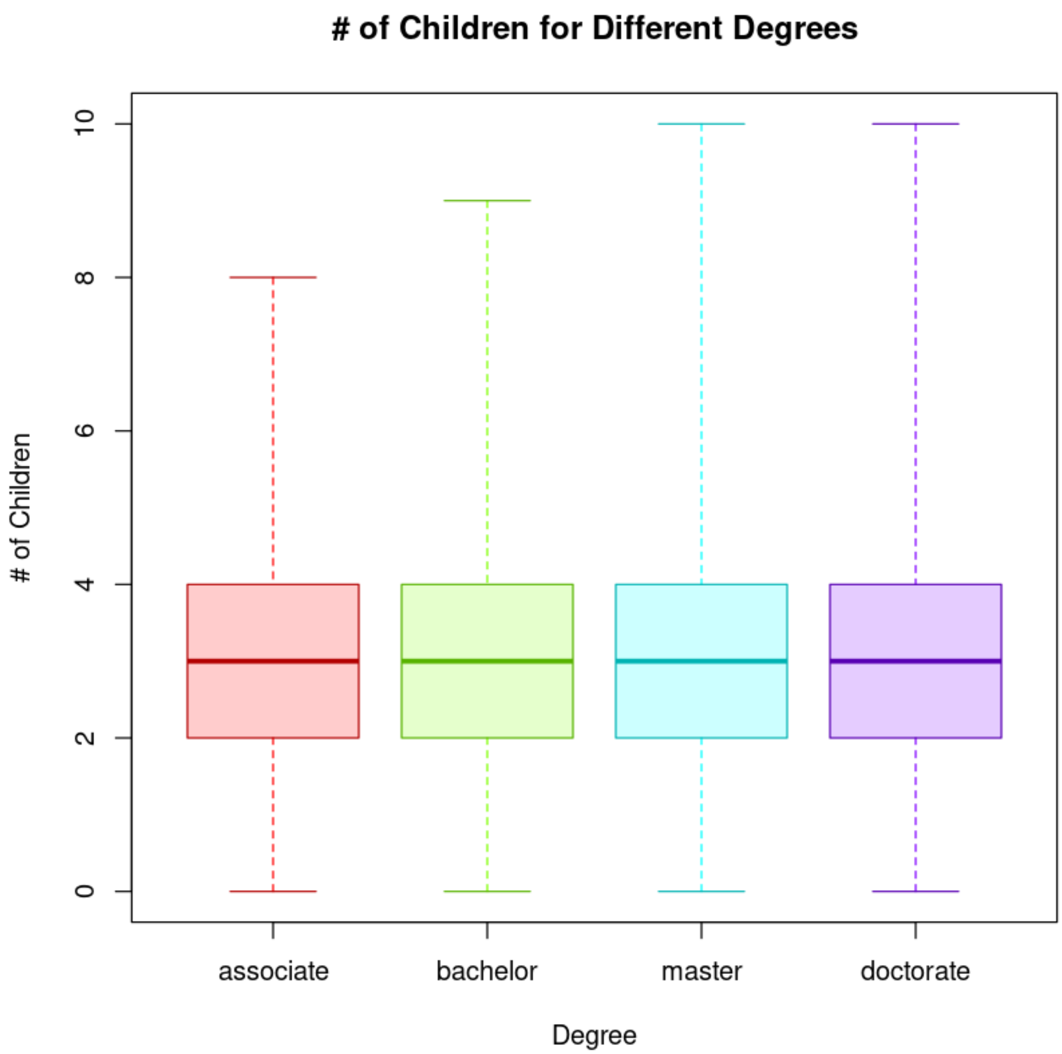


[75]:
```
cor(WLS$Sad4,WLS$Children,use="p")
```

0.00853522569344083

[48]:
```
#Just Male Degree vs. Children

WLS10 <- subset(WLS, Gender == 1)

associate11 <- subset(WLS10, Degree == 1)
bachelor11 <- subset(WLS10, Degree == 2)
master11 <- subset(WLS10, Degree == 3)
doctor11 <- subset(WLS10, Degree == 4)
```

```r
c1 <- rainbow(4)
c2 <- rainbow(4, alpha=0.2)
c3 <- rainbow(4, v=0.7)

boxplot(Children~Degree,
        data = WLS10,
        names = c("associate", "bachelor", "master", "doctorate"),
        xlab = "Degree",
        ylab = "# of Children",
        main = "# of Children for Different Degrees",
        range = 8,
       col = c2,
       medcol = c3,
       whiskcol = c1,
       staplecol= c3,
       boxcol = c3,
       outcol = c3,
       pch = 23,
       cex = 2)
```



[49]:
```
mean(associate11$Children, na.rm = TRUE)
mean(bachelor11$Children, na.rm = TRUE)
mean(master11$Children, na.rm = TRUE)
mean(doctor11$Children, na.rm = TRUE)

median(associate11$Children, na.rm = TRUE)
median(bachelor11$Children, na.rm = TRUE)
median(master11$Children, na.rm = TRUE)
median(doctor11$Children, na.rm = TRUE)
```

3.14

2.8263358778626

2.78798586572438

2.96575342465753

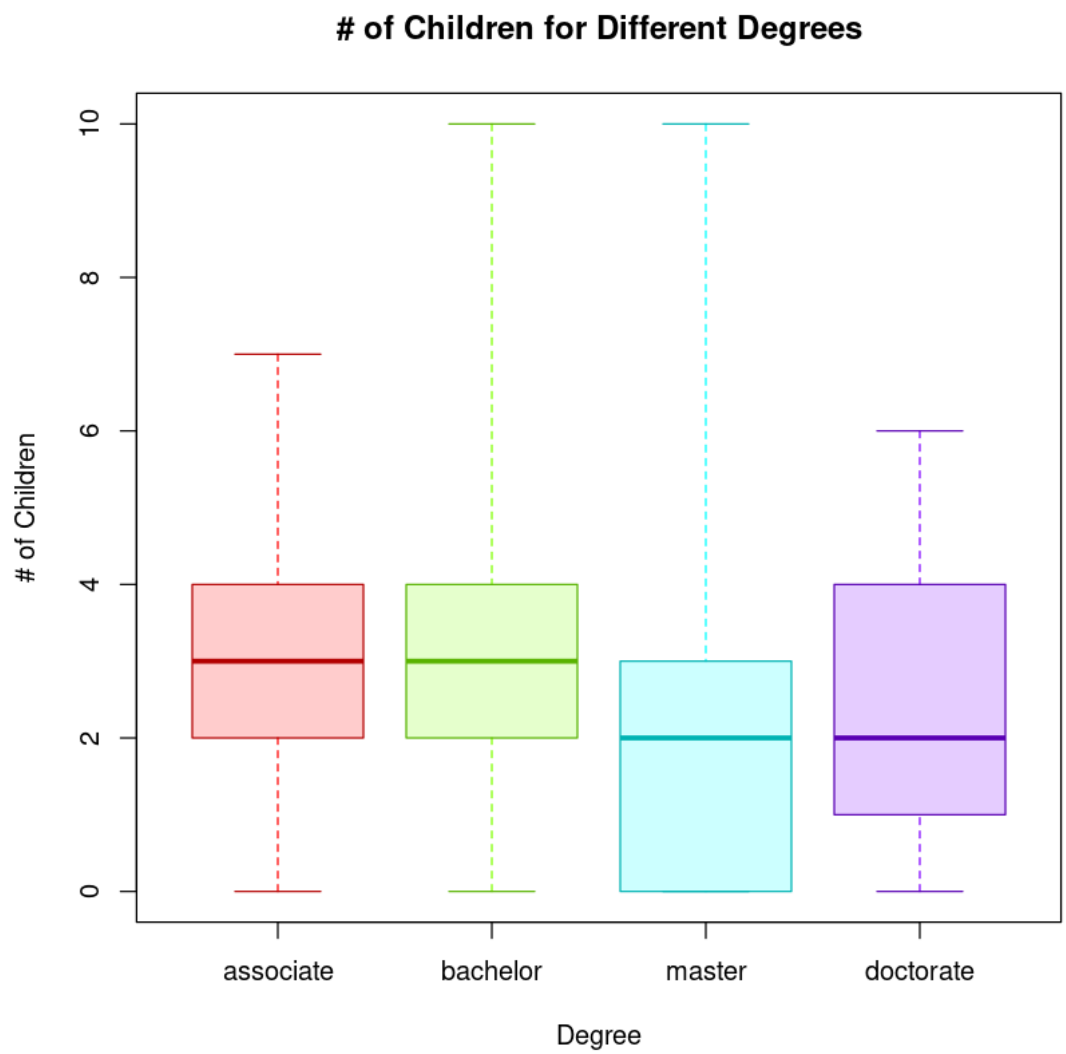
3

3

3

3

[10]:
```
#Just Female Degree vs. Children

WLS9 <- subset(WLS, Gender == 2)

associate6 <- subset(WLS9, Degree == 1)
bachelor6 <- subset(WLS9, Degree == 2)
master6 <- subset(WLS9, Degree == 3)
doctor6 <- subset(WLS9, Degree == 4)


c1 <- rainbow(4)
c2 <- rainbow(4, alpha=0.2)
c3 <- rainbow(4, v=0.7)

boxplot(Children~Degree,
        data = WLS9,
        names = c("associate", "bachelor", "master", "doctorate"),
        xlab = "Degree",
        ylab = "# of Children",
        main = "# of Children for Different Degrees",
        range = 8,
       col = c2,
       medcol = c3,
       whiskcol = c1,
       staplecol= c3,
       boxcol = c3,
       outcol = c3,
       pch = 23,
       cex = 2)
```



[11]:
```
t.test(master6$Children,bachelor6$Children,
       conf.level = 0.99,
       alternative = "l",
       var.equal = TRUE)

t.test(doctor6$Children,bachelor6$Children,
       conf.level = 0.99,
       var.equal = TRUE)

mean(associate6$Children, na.rm = TRUE)
mean(bachelor6$Children, na.rm = TRUE)
mean(master6$Children, na.rm = TRUE)
mean(doctor6$Children, na.rm = TRUE)
```

```r
median(associate6$Children, na.rm = TRUE)
median(bachelor6$Children, na.rm = TRUE)
median(master6$Children, na.rm = TRUE)
median(doctor6$Children, na.rm = TRUE)
```

```
        Two Sample t-test

data:  master6$Children and bachelor6$Children
t = -5.112, df = 647, p-value = 2.102e-07
alternative hypothesis: true difference in means is less than 0
99 percent confidence interval:
       -Inf -0.4630557
sample estimates:
mean of x mean of y
 1.946154  2.797688
```

```
        Two Sample t-test

data:  doctor6$Children and bachelor6$Children
t = -0.31664, df = 528, p-value = 0.7516
alternative hypothesis: true difference in means is not equal to 0
99 percent confidence interval:
 -1.478423  1.155774
sample estimates:
mean of x mean of y
 2.636364  2.797688
```

3.28301886792453

2.79768786127168

1.94615384615385

2.63636363636364

3

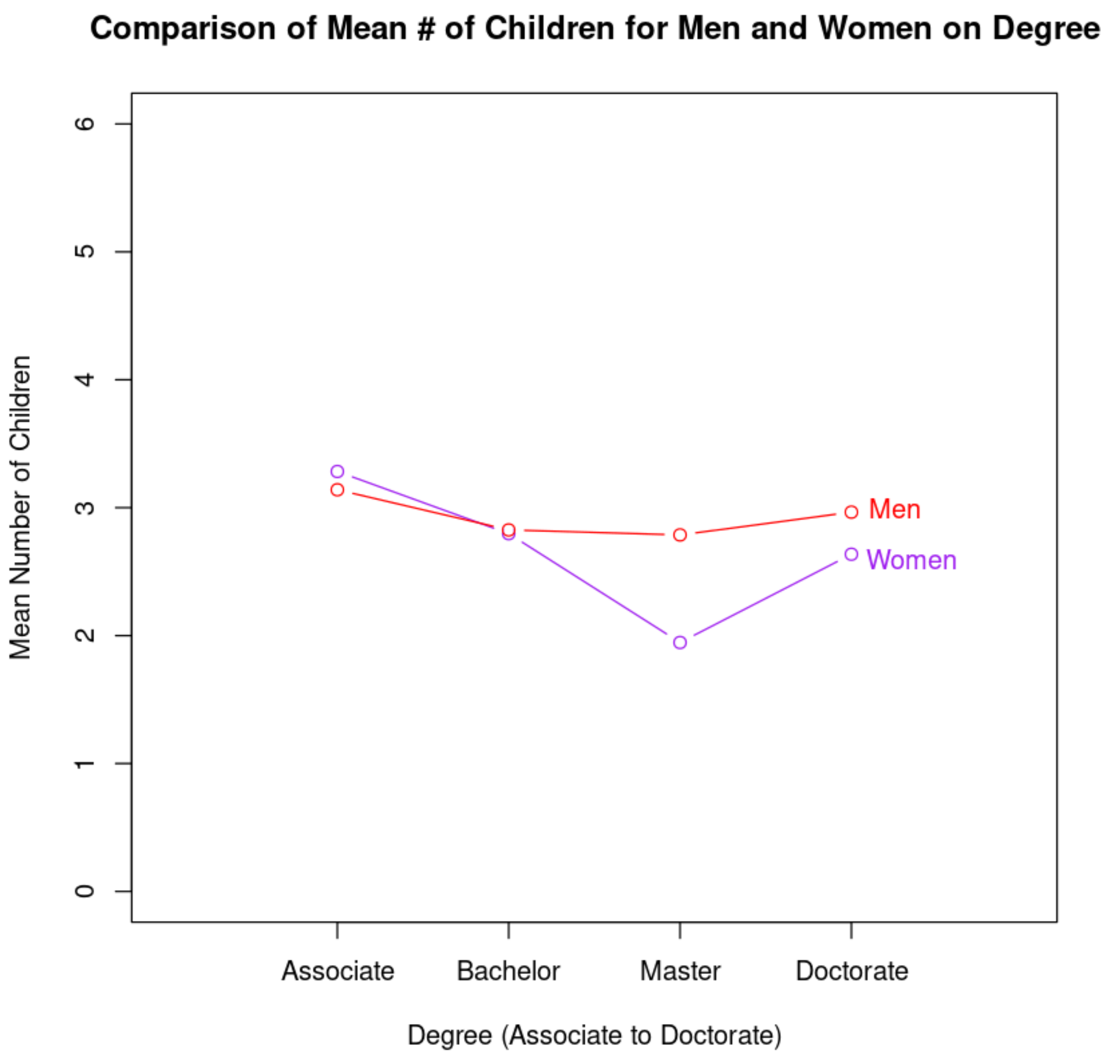
3

2

2

[19]:
```r
mean_values <- c(3.283, 2.797, 1.946, 2.636)
median_values <- c(3,3,2,2)

mean_values1 <- c(3.14,2.826,2.787,2.965)
```

```r
median_values1 <- c(2,2,2,2)

degree_labels <- c("Associate", "Bachelor", "Master", "Doctorate")

plot(c(1, 2, 3, 4),  # x values, made into a vector
     mean_values, # y values, the mean values for each degree
     type = "b",      # this means plot (b)oth lines and points
     ylim = c(0, 6), # make the y-axis go from 0 to 10, adjust as needed
     xlim = c(0, 5),  # make the x-axis go from 0 to 5
     xlab = "Degree (Associate to Doctorate)",
     ylab = "Mean Number of Children",
     main = "Comparison of Mean # of Children for Men and Women on Degree",
     xaxt = "n",
     col = "purple" # this means (n)o x-axis, because we will add our own with␣
 ↪the names of the degrees
)

# Add custom x-axis labels
axis(1, at = c(1, 2, 3, 4), labels = degree_labels)


lines(c(1, 2, 3, 4), mean_values1, col = "red", type = "b")
points(c(1, 2, 3, 4), mean_values1, col = "white", pch = 20)

text(4.25,3, # same x but y slightly lower
     labels = "Men",
     cex = 1,
     col = "red"
     )

text(4.35,2.6, # same x but y slightly lower
     labels = "Women",
     cex = 1,
     col = "purple"
     )
```



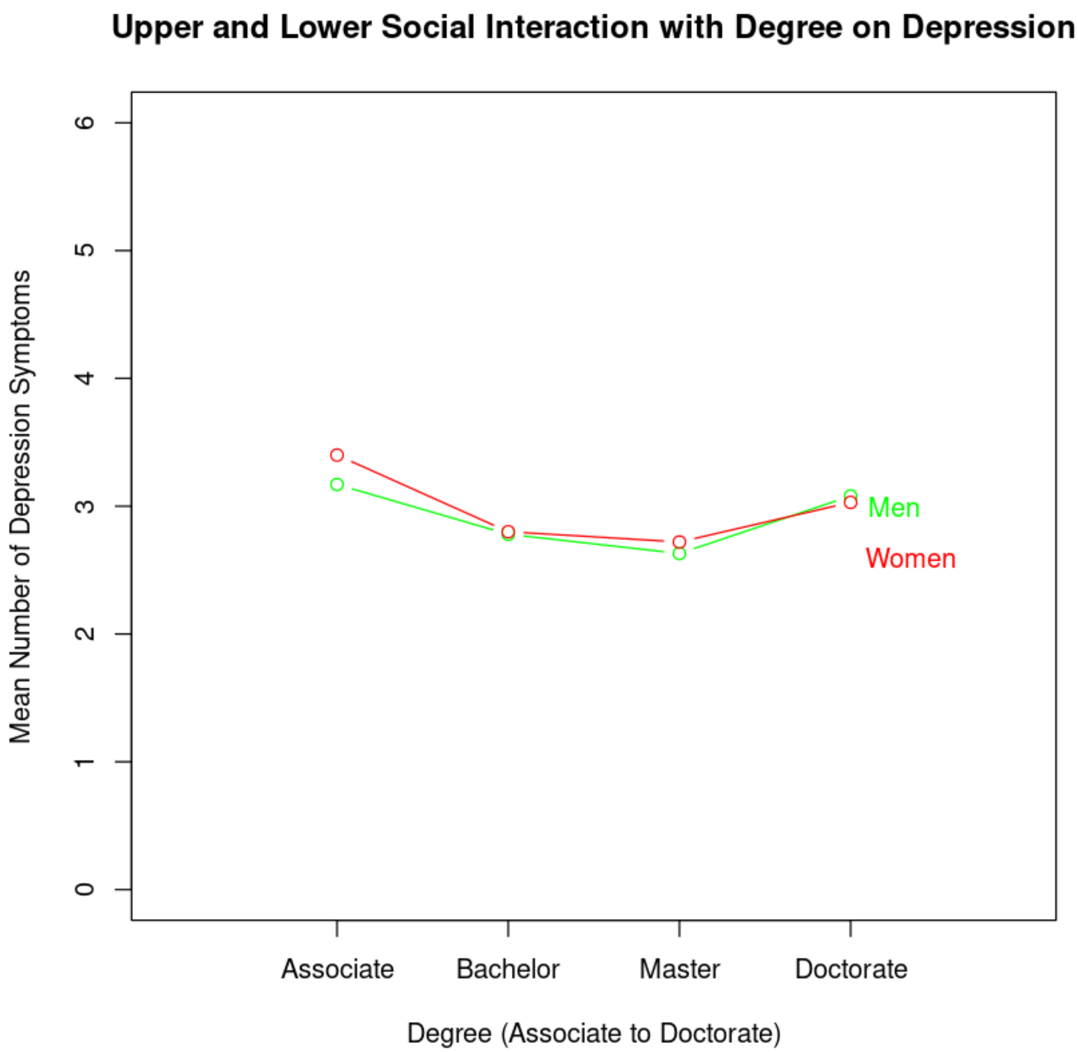
[62]:
```
prop.test(x = c(1.946, 2.767),
          n = c(10.662, 11.72),
          conf.level = 0.99,
          alternative = "l"
          )
```

```
Warning message in prop.test(x = c(1.946, 2.767), n = c(10.662, 11.72),
conf.level = 0.99, :
"Chi-squared approximation may be incorrect"


        2-sample test for equality of proportions with continuity correction

data:  c(1.946, 2.767) out of c(10.662, 11.72)
```

```
X-squared = 8.2279e-31, df = 1, p-value = 0.5
alternative hypothesis: less
99 percent confidence interval:
 -1.000000  0.398767
sample estimates:
   prop 1    prop 2
0.1825174 0.2360922
```

[14]:
```
mean_values3 <- c(3.17, 2.78, 2.63, 3.08)


mean_values4 <- c(3.4,2.8,2.72,3.03)


degree_labels1 <- c("Associate", "Bachelor", "Master", "Doctorate")

plot(c(1, 2, 3, 4),  # x values, made into a vector
     mean_values3, # y values, the mean values for each degree
     type = "b",     # this means plot (b)oth lines and points
     ylim = c(0, 6), # make the y-axis go from 0 to 10, adjust as needed
     xlim = c(0, 5),  # make the x-axis go from 0 to 5
     xlab = "Degree (Associate to Doctorate)",
     ylab = "Mean Number of Depression Symptoms",
     main = "Upper and Lower Social Interaction with Degree on Depression",
     xaxt = "n",
     col = "green" # this means (n)o x-axis, because we will add our own with
 ↪the names of the degrees
)

# Add custom x-axis labels
axis(1, at = c(1, 2, 3, 4), labels = degree_labels1)


lines(c(1, 2, 3, 4), mean_values4, col = "red", type = "b")
points(c(1, 2, 3, 4), mean_values4, col = "white", pch = 20)

text(4.25,3, # same x but y slightly lower
     labels = "Men",
     cex = 1,
     col = "green"
     )

text(4.35,2.6, # same x but y slightly lower
     labels = "Women",
     cex = 1,
     col = "red"
```

```
    )
```



Upper and Lower Social Interaction with Degree on Depression
Mean Number of Depression Symptoms
0
1
2
3
4
5
6
Men
Women
Associate
Bachelor
Master
Doctorate
Degree (Associate to Doctorate)