Analysis of survey data for metrics, scientific literacy and attitude

H Qin

5/30/2019

Learning Objectives

- Load external tabular data from a .csv file into R.
- Describe what an R data frame is.
- Summarize the contents of a data frame in R.
- Manipulate categorical data in R using factors.

Check files in the working directory

```
rm( list=ls()) #clean up worksapce
list.files()
```

```
## [1] "Learning_R_by_metricExample.ppt"
## [2] "learningR_through_metric_example.html"
## [3] "learningR_through_metric_example.pdf"
## [4] "learningR_through_metric_example.Rmd"
## [5] "metric_survey_form.pdf"
## [6] "metric-attitude-literacy.csv"
## [7] "simpleR.html"
## [8] "simpleR.Rmd"
```

Read the survey data in csv format

```
tb = read.csv("metric-attitude-literacy.csv", stringsAsFactors = FALSE)
```

What is a data frame?

Looking for helps

```
?str
help(str)
str(tb);
```

```
## 'data.frame':    316 obs. of  7 variables:
##  $ gender     : chr  "Do not wish to answer" "Male" "Female" "Female" ...
##  $ age        : num  20 20 35.5 53 58 20 45.5 35.5 35.5 20 ...
##  $ country    : int  1 1 1 1 1 1 1 1 1 0 1 ...
##  $ degree     : chr  "Bachelor Degree in Science or equivalent" "High School or equivalent" "High Sc
##  $ metric     : int  3 4 1 1 4 0 4 5 3 2 ...
##  $ SciAttitude: int  3 3 1 3 4 2 3 4 4 2 ...
##  $ SciLitScore: int  6 9 8 4 9 7 9 9 9 4 ...
```

Indexing an element in a dataframe

```r
#indexing features of R
tb[1:5, 2:3]
```

```
##     age country
## 1 20.0       1
## 2 20.0       1
## 3 35.5       1
## 4 53.0       1
## 5 58.0       1
```

```r
tb$age #what does mean?
```

```
##     [1] 20.0 20.0 35.5 53.0 58.0 20.0 45.5 35.5 35.5 20.0 58.0 65.0 45.5 26.5
##    [15] 26.5 26.5 20.0 26.5 26.5 26.5 26.5 20.0 26.5 35.5 26.5 26.5 20.0 20.0
##    [29] 26.5 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 26.5 20.0 20.0
##    [43] 20.0 20.0 20.0 26.5 26.5 20.0 26.5 26.5 26.5 20.0 20.0 35.5 26.5 20.0
##    [57] 20.0 20.0 20.0 20.0 20.0 26.5 20.0 20.0 20.0 20.0 20.0 20.0 35.5 20.0
##    [71] 20.0 20.0 20.0 20.0 26.5 20.0 20.0 20.0 20.0 20.0 45.5 26.5 58.0 35.5
##    [85] 20.0 20.0 58.0 20.0 35.5 26.5 26.5 20.0 65.0 53.0 26.5 65.0 26.5 20.0
##    [99] 26.5 20.0 53.0 20.0 58.0 45.5 65.0 20.0 20.0 20.0 45.5 65.0 58.0 53.0
##   [113] 20.0 20.0 45.5 26.5 20.0 53.0 65.0 45.5 65.0 65.0 65.0 35.5 65.0 35.5
##   [127] 35.5 58.0 58.0 58.0 58.0 58.0 45.5 45.5 65.0 53.0 65.0 35.5 35.5 65.0
##   [141] 53.0 45.5 45.5 20.0 35.5 20.0 53.0 45.5 45.5 45.5 45.5 65.0 65.0 20.0
##   [155] 58.0 45.5 65.0 53.0 20.0 35.5 58.0 58.0 53.0 45.5 35.5 26.5 53.0 35.5
##   [169] 35.5 20.0 20.0 26.5 26.5 53.0 35.5 35.5 35.5 20.0 65.0 35.5 45.5 20.0
##   [183] 65.0 58.0 65.0 26.5 20.0 20.0 20.0 20.0 35.5 26.5 20.0 20.0 65.0 20.0
##   [197] 35.5 26.5 65.0 65.0 20.0 26.5 45.5 65.0 65.0 20.0 45.5 35.5 35.5 35.5
##   [211] 53.0 45.5 20.0 65.0 58.0 65.0 20.0 58.0 26.5 45.5 20.0 20.0 20.0 20.0
##   [225] 20.0 20.0 20.0 20.0 20.0 20.0 26.5 20.0 20.0 20.0 20.0 20.0 20.0 20.0
##   [239] 20.0 26.5 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0
##   [253] 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0
##   [267] 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 26.5 20.0
##   [281] 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0 20.0
##   [295] 20.0 20.0 20.0 20.0 53.0 65.0 65.0 20.0 20.0 20.0 20.0 53.0 35.5 35.5
##   [309] 35.5 35.5 20.0 20.0 20.0 20.0 20.0 26.5
```

```r
#tb$age[?] #try for 5th row in age

# what is the 5th and 9th rows in age and SciLitScore:
tb[c(5,9), c(7,2)]
```

```
##   SciLitScore  age
## 5           9 58.0
## 9           9 35.5
```

```r
tb[c(5,9), c('age', 'SciLitScore')]
```

```
##    age SciLitScore
## 5 58.0           9
## 9 35.5           9
```

Overview of the data

```r
summary(tb)
```

```
##     gender               age          country        degree
```

```
##   Length:316        Min.   :20.0   Min.   :0.0000   Length:316
##   Class :character  1st Qu.:20.0   1st Qu.:1.0000   Class :character
##   Mode  :character  Median :20.0   Median :1.0000   Mode  :character
##                     Mean   :31.7   Mean   :0.8323
##                     3rd Qu.:45.5   3rd Qu.:1.0000
##                     Max.   :65.0   Max.   :1.0000
##      metric        SciAttitude     SciLitScore
##   Min.   :0.000   Min.   :0.000   Min.   : 0.000
##   1st Qu.:3.000   1st Qu.:2.000   1st Qu.: 6.000
##   Median :4.000   Median :3.000   Median : 8.000
##   Mean   :3.544   Mean   :2.582   Mean   : 7.259
##   3rd Qu.:5.000   3rd Qu.:3.000   3rd Qu.: 9.000
##   Max.   :5.000   Max.   :4.000   Max.   :10.000
```

Exercise: Find out help information of summary().

```
head(tb)
```

```
##                  gender  age country
## 1 Do not wish to answer 20.0       1
## 2                  Male 20.0       1
## 3                Female 35.5       1
## 4                Female 53.0       1
## 5                Female 58.0       1
## 6                Female 20.0       1
##                                    degree metric SciAttitude SciLitScore
## 1 Bachelor Degree in Science or equivalent      3           3           6
## 2             High School or equivalent      4           3           9
## 3             High School or equivalent      1           1           8
## 4             High School or equivalent      1           3           4
## 5    Bachelor Degree in Arts or equivalent      4           4           9
## 6             High School or equivalent      0           2           7
```

Look at first 2 rows of columns 2 and 3

```
head( tb[, 2:3], n=2)
```

```
##   age country
## 1  20       1
## 2  20       1
```

The survey provide results for three types of questions

1) Metric proficiency
2) Scientific literacy
3) Attitude toward science
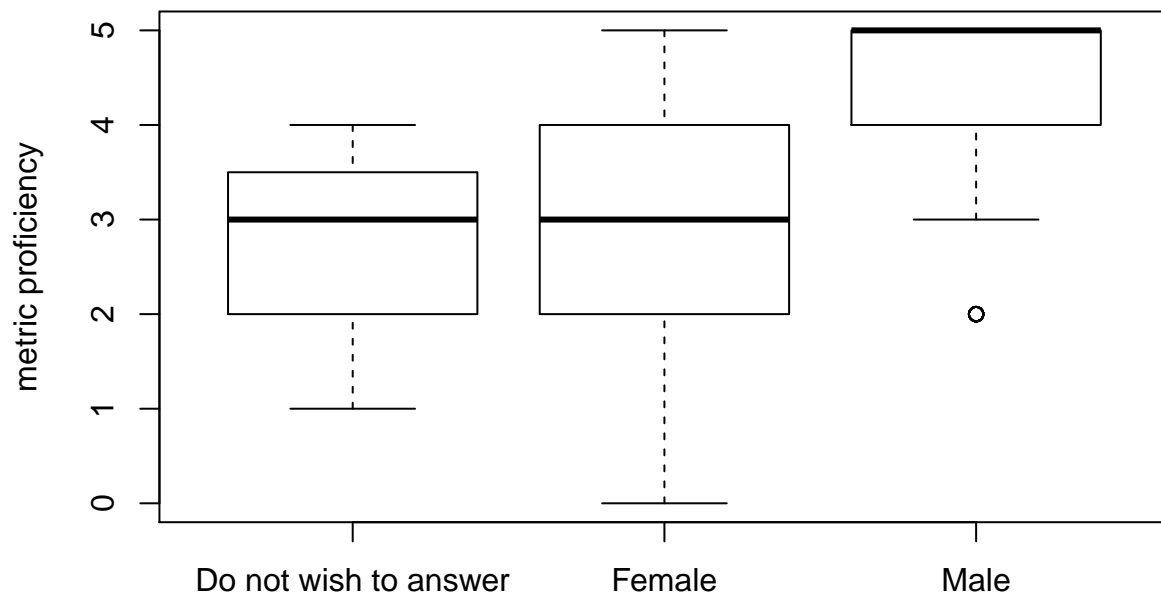
What does the 'country' coded for?

```
table(tb$country)
```

```
##
##   0   1
##  53 263
```

So, USA is 1, other countries are coded as zeros.

Is gender associated with metric proficiency?

```r
boxplot( tb$metric ~ tb$gender, ylab="metric proficiency" )
```



```r
table(tb$metric, tb$gender )
```

```
##
##      Do not wish to answer Female Male
##   0                      0      4    0
##   1                      1     20    0
##   2                      0     40   12
##   3                      1     43   12
##   4                      1     62   25
##   5                      0     39   56
```

How to pick a subset of data? Pick "Male" metric proficiency data?

```r
tb$gender=='Male'
```

```
##   [1] FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE
##  [12] FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE
##  [23]  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE
##  [34]  TRUE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [45] FALSE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE FALSE  TRUE FALSE FALSE
##  [56] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [67] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [78] FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE
##  [89] FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE
## [100] FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE
## [111]  TRUE  TRUE FALSE  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE  TRUE
## [122]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE  TRUE
## [133]  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE FALSE FALSE  TRUE  TRUE
## [144]  TRUE  TRUE FALSE  TRUE FALSE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE
## [155]  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [166]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [177]  TRUE FALSE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE
```

```
## [188] FALSE  TRUE FALSE  TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE  TRUE
## [199]  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE
## [210] FALSE  TRUE FALSE FALSE  TRUE FALSE  TRUE FALSE  TRUE  TRUE FALSE
## [221] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [232] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [243] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [254] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [265] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [276] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [287] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [298] FALSE FALSE  TRUE FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE FALSE
## [309]  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE
```

```r
tb$metric[tb$gender=='Male']
```

```
##   [1] 4 3 3 3 4 2 2 2 5 5 2 2 4 2 2 2 4 3 5 3 4 5 5 5 4 5 5 2 5 5 5 5 5 4 5
##  [36] 4 5 5 4 5 5 4 5 4 5 5 4 5 5 5 5 5 4 4 4 3 3 5 4 5 5 5 5 5 4 4 3 5 5 4
##  [71] 5 5 3 5 5 5 3 5 5 2 5 5 3 2 4 5 5 5 5 2 3 4 4 5 5 4 4 5 5 4 5 5 5 5 5
```

```r
tb$metric[tb$gender=='Female']
```

```
##   [1] 1 1 4 0 4 5 2 4 4 4 4 3 2 4 2 3 5 5 4 3 1 3 2 4 3 1 1 4 2 3 3 3 1 1 2
##  [36] 2 5 5 4 4 5 5 1 1 3 3 5 3 4 3 1 4 1 2 5 4 2 2 4 3 1 4 2 4 2 3 2 2 4 4
##  [71] 2 5 4 4 5 3 2 2 5 2 5 4 0 3 4 4 3 4 5 2 5 5 4 3 1 2 4 2 5 3 5 3 2 0 3
## [106] 1 3 4 5 3 2 5 5 5 5 2 3 1 5 2 1 2 4 1 4 4 0 3 2 5 2 4 4 2 4 4 2 4 4 5
## [141] 3 4 5 1 2 3 5 5 2 5 3 2 5 2 1 4 4 4 2 2 5 3 3 4 4 4 4 4 5 2 4 3 3 3 4
## [176] 1 3 4 5 4 4 3 5 4 5 4 2 5 4 4 3 4 4 3 4 4 3 3 2 3 3 5 4 5 2 4 3 3
```

What is t-test?

```r
t.test(tb$metric[tb$gender=='Female'], tb$metric[tb$gender=='Male'], alternative = 'less')
```

```
##
##  Welch Two Sample t-test
##
## data:  tb$metric[tb$gender == "Female"] and tb$metric[tb$gender == "Male"]
## t = -7.0175, df = 257.55, p-value = 9.969e-12
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##        -Inf -0.7339468
## sample estimates:
## mean of x mean of y
##  3.230769  4.190476
```

```r
# Does this mean that females are more uncomfortable with metric usage?
```

What does p-value mean?

```r
# Female participants tend to be younger
boxplot( tb$age ~ tb$gender, ylab='age')
```
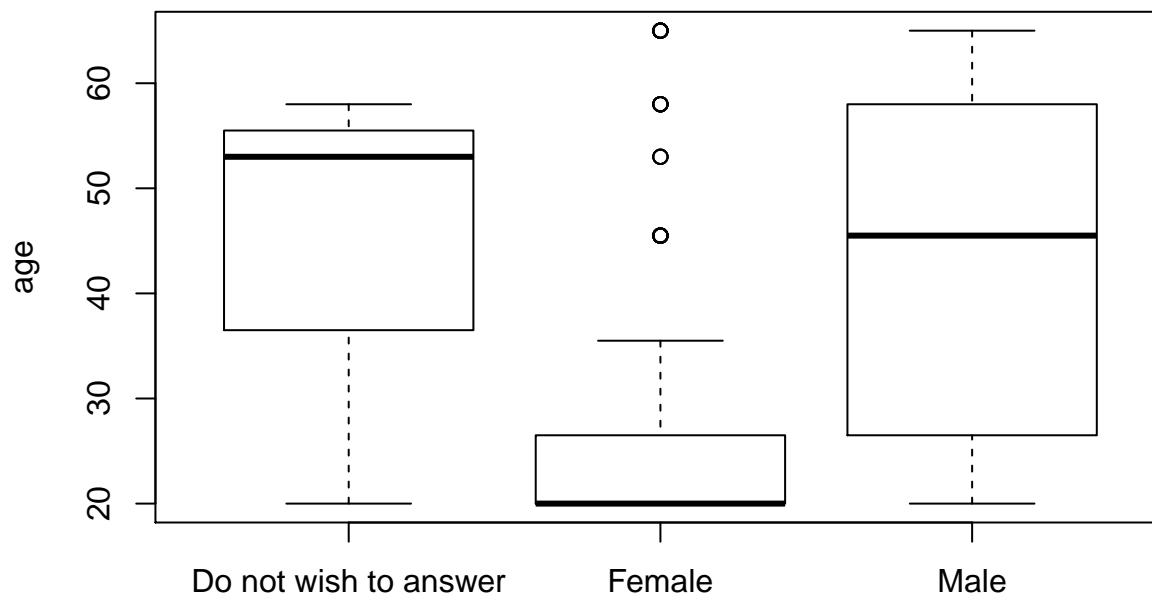
```
# More female participants with Bachelor degrees
table( tb$gender, tb$degree )
```

```
##
##                           Bachelor Degree in Arts or equivalent
##   Do not wish to answer                                        0
##   Female                                                      29
##   Male                                                        16
##
##                           Bachelor Degree in Science or equivalent
##   Do not wish to answer                                          1
##   Female                                                        96
##   Male                                                          27
##
##                           High School or equivalent M.D. or equivalent
##   Do not wish to answer                         0                     0
##   Female                                       44                     0
##   Male                                         23                     2
##
##                           Master Degree or equivalent Ph.D. or equivalent
##   Do not wish to answer                           0                     2
##   Female                                         13                    26
##   Male                                           18                    19
```

```
boxplot( tb$metric ~ tb$degree, ylab='metric proficiency')
```
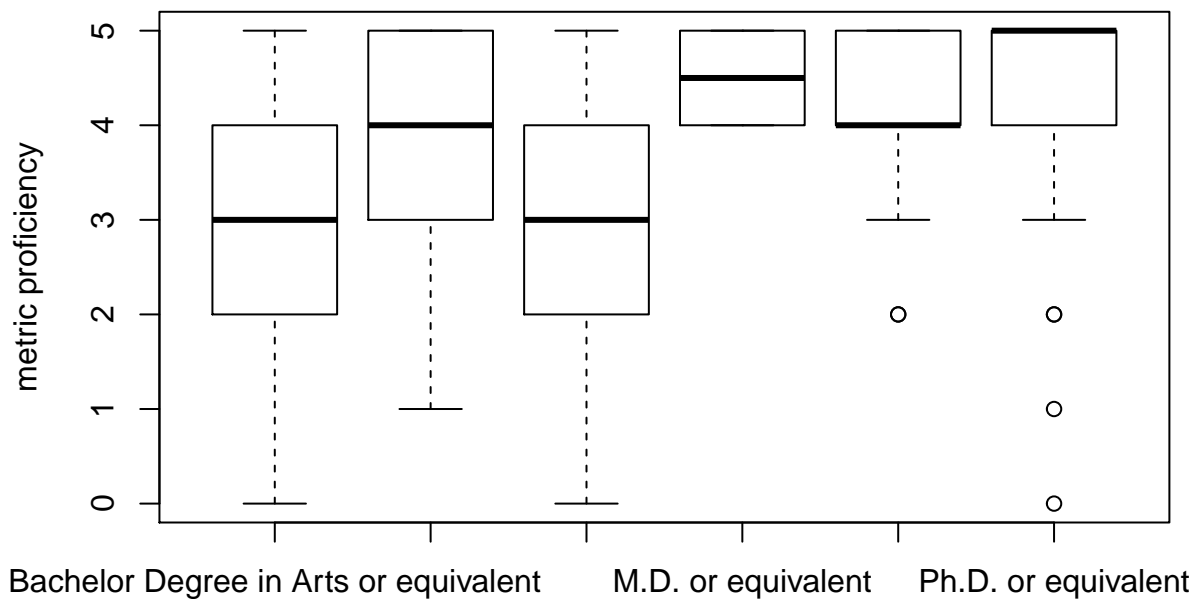
```
# how age in boxplot
# ???

m1 = lm( tb$metric ~ tb$gender )
summary(m1)
```

```
##
## Call:
## lm(formula = tb$metric ~ tb$gender)
##
## Residuals:
##     Min     1Q  Median      3Q     Max
## -3.2308 -1.1905  0.5513  0.8095  1.7692
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       2.6667     0.7148   3.731 0.000227 ***
## tb$genderFemale   0.5641     0.7199   0.784 0.433904
## tb$genderMale     1.5238     0.7249   2.102 0.036355 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.238 on 313 degrees of freedom
## Multiple R-squared:  0.1219, Adjusted R-squared:  0.1163
## F-statistic: 21.72 on 2 and 313 DF,  p-value: 1.463e-09
```

```
#summary( lm( tb$metric ~ tb$gender ) )

m2 = lm( tb$metric ~ tb$age )
summary(m2)
```

```
##
## Call:
## lm(formula = tb$metric ~ tb$age)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -4.3338 -1.2670  0.3655  0.8321  1.7330
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.792915   0.160471  17.404  < 2e-16 ***
## tb$age      0.023706   0.004538   5.224 3.19e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.265 on 314 degrees of freedom
## Multiple R-squared:  0.07997,    Adjusted R-squared:  0.07704
## F-statistic: 27.29 on 1 and 314 DF,  p-value: 3.189e-07
```

```
m3 = lm( tb$metric ~ tb$degree )
summary(m3)
```

```
##
## Call:
## lm(formula = tb$metric ~ tb$degree)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1702 -0.8667  0.1333  0.8387  2.1333
##
## Coefficients:
##                                             Estimate Std. Error
## (Intercept)                                   2.8667     0.1861
## tb$degreeBachelor Degree in Science or equivalent  0.7301     0.2172
## tb$degreeHigh School or equivalent            0.2826     0.2405
## tb$degreeM.D. or equivalent                   1.6333     0.9019
## tb$degreeMaster Degree or equivalent          1.2946     0.2913
## tb$degreePh.D. or equivalent                  1.3035     0.2603
##                                             t value Pr(>|t|)
## (Intercept)                                  15.408  < 2e-16 ***
## tb$degreeBachelor Degree in Science or equivalent  3.361 0.000873 ***
## tb$degreeHigh School or equivalent            1.175 0.240992
## tb$degreeM.D. or equivalent                   1.811 0.071115 .
## tb$degreeMaster Degree or equivalent          4.444 1.23e-05 ***
## tb$degreePh.D. or equivalent                  5.008 9.26e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.248 on 310 degrees of freedom
## Multiple R-squared:  0.1162, Adjusted R-squared:  0.102
## F-statistic: 8.153 on 5 and 310 DF,  p-value: 3.017e-07
```

```
m4 = lm( tb$metric~ tb$gender + tb$age + tb$degree )
summary(m4)
```

```
##
## Call:
## lm(formula = tb$metric ~ tb$gender + tb$age + tb$degree)
##
## Residuals:
```

```
##       Min       1Q   Median       3Q      Max
## -4.0431  -0.7871   0.2290   0.6649   2.5024
##
## Coefficients:
##                                                Estimate Std. Error
## (Intercept)                                    1.231733   0.735745
## tb$genderFemale                                1.111723   0.690190
## tb$genderMale                                  1.932072   0.692157
## tb$age                                         0.007706   0.005601
## tb$degreeBachelor Degree in Science or equivalent 0.897745   0.205772
## tb$degreeHigh School or equivalent             0.289489   0.225330
## tb$degreeM.D. or equivalent                    0.948954   0.852364
## tb$degreeMaster Degree or equivalent           0.984652   0.284637
## tb$degreePh.D. or equivalent                   1.198766   0.256987
##                                                t value Pr(>|t|)
## (Intercept)                                      1.674 0.095123 .
## tb$genderFemale                                  1.611 0.108262
## tb$genderMale                                    2.791 0.005577 **
## tb$age                                           1.376 0.169880
## tb$degreeBachelor Degree in Science or equivalent  4.363 1.76e-05 ***
## tb$degreeHigh School or equivalent               1.285 0.199854
## tb$degreeM.D. or equivalent                      1.113 0.266442
## tb$degreeMaster Degree or equivalent             3.459 0.000618 ***
## tb$degreePh.D. or equivalent                     4.665 4.62e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.169 on 307 degrees of freedom
## Multiple R-squared:  0.2322, Adjusted R-squared:  0.2122
## F-statistic: 11.61 on 8 and 307 DF,  p-value: 2.026e-14
```

Caudal regression analyses

```
#########################
#summary(tb); str(tb)

#remove rows with missing age from analysis. Missing age can cause bugs in anova model comparisons.
tb = tb[!is.na(tb$age), ]

summary(tb)
```

```
##     gender               age             country          degree
## Length:316         Min.   :20.0    Min.   :0.0000    Length:316
## Class :character   1st Qu.:20.0    1st Qu.:1.0000    Class :character
## Mode  :character   Median :20.0    Median :1.0000    Mode  :character
##                    Mean   :31.7    Mean   :0.8323
##                    3rd Qu.:45.5    3rd Qu.:1.0000
##                    Max.   :65.0    Max.   :1.0000
##     metric         SciAttitude       SciLitScore
## Min.   :0.000   Min.   :0.000    Min.   : 0.000
## 1st Qu.:3.000   1st Qu.:2.000    1st Qu.: 6.000
## Median :4.000   Median :3.000    Median : 8.000
## Mean   :3.544   Mean   :2.582    Mean   : 7.259
```
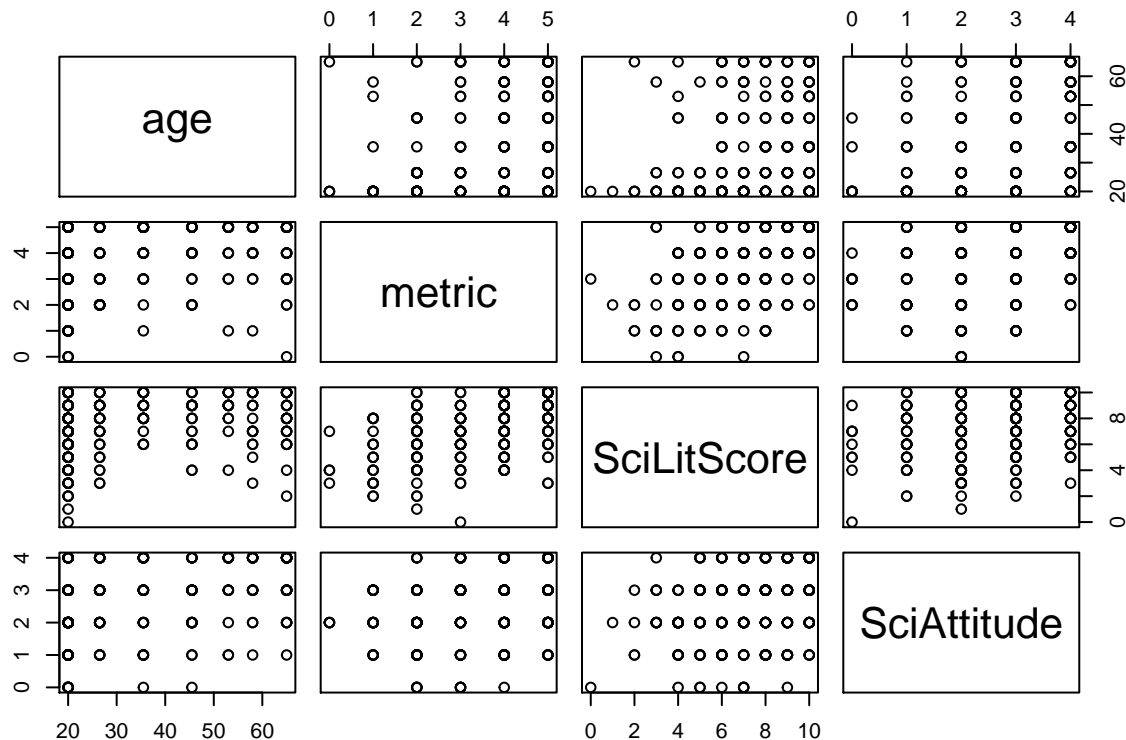
```
##   3rd Qu.:5.000    3rd Qu.:3.000    3rd Qu.: 9.000
##   Max.   :5.000    Max.   :4.000    Max.   :10.000
```

```r
str(tb);
```

```
## 'data.frame':    316 obs. of  7 variables:
##  $ gender     : chr  "Do not wish to answer" "Male" "Female" "Female" ...
##  $ age        : num  20 20 35.5 53 58 20 45.5 35.5 35.5 20 ...
##  $ country    : int  1 1 1 1 1 1 1 1 1 0 1 ...
##  $ degree     : chr  "Bachelor Degree in Science or equivalent" "High School or equivalent" "High Sc
##  $ metric     : int  3 4 1 1 4 0 4 5 3 2 ...
##  $ SciAttitude: int  3 3 1 3 4 2 3 4 4 2 ...
##  $ SciLitScore: int  6 9 8 4 9 7 9 9 9 4 ...
```

```r
pairs(tb[, c("age", "metric", "SciLitScore", "SciAttitude")])
```



```r
summary(lm(tb$SciLitScore ~ tb$metric )) #significant
```

```
##
## Call:
## lm(formula = tb$SciLitScore ~ tb$metric)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.7959 -0.9442  0.2041  1.3894  4.0558
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.24071    0.29382   14.43   <2e-16 ***
## tb$metric    0.85173    0.07772   10.96   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## 
## Residual standard error: 1.817 on 314 degrees of freedom
## Multiple R-squared:  0.2767, Adjusted R-squared:  0.2743
## F-statistic: 120.1 on 1 and 314 DF,  p-value: < 2.2e-16
```

```r
summary(lm(tb$SciAttitude ~ tb$metric )) #significant
```

```
## 
## Call:
## lm(formula = tb$SciAttitude ~ tb$metric)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -2.71893 -0.71893 -0.01881  0.98119  1.88083 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  1.51941    0.16269   9.340  < 2e-16 ***
## tb$metric    0.29988    0.04303   6.968 1.89e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.006 on 314 degrees of freedom
## Multiple R-squared:  0.1339, Adjusted R-squared:  0.1312
## F-statistic: 48.56 on 1 and 314 DF,  p-value: 1.892e-11
```

```r
summary(lm(tb$SciAttitude ~ tb$SciLitScore )) #significant
```

```
## 
## Call:
## lm(formula = tb$SciAttitude ~ tb$SciLitScore)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -2.8741 -0.7065  0.1259  0.9582  2.1320 
## 
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)    
## (Intercept)     1.36493    0.20383   6.696 9.86e-11 ***
## tb$SciLitScore  0.16769    0.02694   6.224 1.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.02 on 314 degrees of freedom
## Multiple R-squared:  0.1098, Adjusted R-squared:  0.107
## F-statistic: 38.74 on 1 and 314 DF,  p-value: 1.55e-09
```

```r
summary(lm(tb$SciAttitude ~ tb$age )) #significant
```

```
## 
## Call:
## lm(formula = tb$SciAttitude ~ tb$age)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -2.90581 -0.46049 -0.08159  0.69185  1.69185 
```

```
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.839391   0.128861   14.274  < 2e-16 ***
## tb$age      0.023438   0.003644    6.432 4.68e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.016 on 314 degrees of freedom
## Multiple R-squared:  0.1164, Adjusted R-squared:  0.1136
## F-statistic: 41.38 on 1 and 314 DF,  p-value: 4.68e-10
```

```r
summary(lm(tb$metric ~ tb$age )) #significant
```

```
##
## Call:
## lm(formula = tb$metric ~ tb$age)
##
## Residuals:
##     Min     1Q  Median     3Q     Max
## -4.3338 -1.2670  0.3655  0.8321  1.7330
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.792915   0.160471   17.404  < 2e-16 ***
## tb$age      0.023706   0.004538    5.224 3.19e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.265 on 314 degrees of freedom
## Multiple R-squared:  0.07997,    Adjusted R-squared:  0.07704
## F-statistic: 27.29 on 1 and 314 DF,  p-value: 3.189e-07
```

```r
summary(lm(tb$SciAttitude ~ tb$SciLitScore + tb$metric )) #significant
```

```
##
## Call:
## lm(formula = tb$SciAttitude ~ tb$SciLitScore + tb$metric)
##
## Residuals:
##     Min     1Q  Median     3Q     Max
## -2.6560 -0.6931  0.1269  0.8354  1.9726
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.10736    0.20690   5.352 1.68e-07 ***
## tb$SciLitScore   0.09717    0.03081   3.154  0.00177 **
## tb$metric        0.21712    0.04989   4.352 1.83e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9919 on 313 degrees of freedom
## Multiple R-squared:  0.1606, Adjusted R-squared:  0.1552
## F-statistic: 29.94 on 2 and 313 DF,  p-value: 1.261e-12
```

```
## metric -> SciAttitude and SciLitScore
```

```
summary(lm(tb$SciAttitude ~ tb$metric + tb$age + tb$gender + tb$country  )) #only metric is significant
```

```
##
## Call:
## lm(formula = tb$SciAttitude ~ tb$metric + tb$age + tb$gender +
##     tb$country)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.7560 -0.6046  0.1531  0.6321  2.0073
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.511175   0.598878   0.854  0.39401
## tb$metric        0.203966   0.044651   4.568 7.11e-06 ***
## tb$age           0.012800   0.004014   3.189  0.00157 **
## tb$genderFemale  0.738602   0.562717   1.313  0.19030
## tb$genderMale    1.175483   0.564200   2.083  0.03803 *
## tb$country       0.078949   0.150618   0.524  0.60054
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9561 on 310 degrees of freedom
## Multiple R-squared:  0.2275, Adjusted R-squared:  0.215
## F-statistic: 18.26 on 5 and 310 DF,  p-value: 6.96e-16
```

```
summary(lm(tb$SciLitScore ~ tb$metric + tb$age + tb$gender + tb$country  )) #only metric is significant
```

```
##
## Call:
## lm(formula = tb$SciLitScore ~ tb$metric + tb$age + tb$gender +
##     tb$country)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.4536 -1.1718  0.1338  1.2435  4.2646
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      5.751977   1.103787   5.211 3.43e-07 ***
## tb$metric        0.718190   0.082295   8.727  < 2e-16 ***
## tb$age           0.008642   0.007398   1.168    0.244
## tb$genderFemale -1.559046   1.037138  -1.503    0.134
## tb$genderMale   -0.692518   1.039872  -0.666    0.506
## tb$country      -0.066754   0.277603  -0.240    0.810
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.762 on 310 degrees of freedom
## Multiple R-squared:  0.328,  Adjusted R-squared:  0.3172
## F-statistic: 30.27 on 5 and 310 DF,  p-value: < 2.2e-16
```

```r
summary(lm(tb$SciLitScore ~ tb$country)) #p=0.0009, but seems due to metric?
```

```
##
## Call:
## lm(formula = tb$SciLitScore ~ tb$country)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.0951 -1.0951 -0.0755  1.9049  2.9049
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.0755     0.2890  27.939  < 2e-16 ***
## tb$country   -0.9804     0.3168  -3.094  0.00215 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.104 on 314 degrees of freedom
## Multiple R-squared:  0.02959,    Adjusted R-squared:  0.0265
## F-statistic: 9.576 on 1 and 314 DF,  p-value: 0.002149
```

```r
summary(lm(tb$SciLitScore ~ tb$metric + tb$country  )) #only metric is significant
```

```
##
## Call:
## lm(formula = tb$SciLitScore ~ tb$metric + tb$country)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.7537 -1.0765  0.2463  1.4150  4.0776
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.57786    0.41813  10.948   <2e-16 ***
## tb$metric    0.83127    0.07976  10.422   <2e-16 ***
## tb$country  -0.31797    0.28070  -1.133    0.258
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.816 on 313 degrees of freedom
## Multiple R-squared:  0.2796, Adjusted R-squared:  0.275
## F-statistic: 60.74 on 2 and 313 DF,  p-value: < 2.2e-16
```

```r
summary(lm(tb$SciAttitude ~ tb$country)) #p=0.0127, but seems due to metric?
```

```
##
## Call:
## lm(formula = tb$SciAttitude ~ tb$country)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.5323 -0.5323  0.1698  0.4677  1.4677
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)    2.8302     0.1477    19.16    <2e-16 ***
## tb$country    -0.2979     0.1619    -1.84     0.0667 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.075 on 314 degrees of freedom
## Multiple R-squared:  0.01067,    Adjusted R-squared:  0.007518
## F-statistic: 3.386 on 1 and 314 DF,  p-value: 0.0667
```

```r
summary(lm(tb$SciAttitude ~ tb$country + tb$metric)) #country not significant when controled for metric
```

```
##
## Call:
## lm(formula = tb$SciAttitude ~ tb$country + tb$metric)
##
## Residuals:
##       Min       1Q    Median       3Q       Max
## -2.70670 -0.70670 -0.00259  0.93534  1.88507
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.58524    0.23193   6.835 4.30e-11 ***
## tb$country  -0.06208    0.15570  -0.399     0.69
## tb$metric    0.29589    0.04424   6.688 1.04e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.007 on 313 degrees of freedom
## Multiple R-squared:  0.1344, Adjusted R-squared:  0.1288
## F-statistic: 24.29 on 2 and 313 DF,  p-value: 1.557e-10
```

```r
plot( tb$SciLitScore ~ jitter(tb$metric), xlab='Metric Proficiency', ylab='Scientific Literacy', ylim=c
m1 = lm(tb$SciLitScore ~ tb$metric )
abline(m1, col='red')
summary(m1)
```

```
##
## Call:
## lm(formula = tb$SciLitScore ~ tb$metric)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.7959 -0.9442  0.2041  1.3894  4.0558
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.24071    0.29382   14.43   <2e-16 ***
## tb$metric    0.85173    0.07772   10.96   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.817 on 314 degrees of freedom
## Multiple R-squared:  0.2767, Adjusted R-squared:  0.2743
## F-statistic: 120.1 on 1 and 314 DF,  p-value: < 2.2e-16
```

```
text(2, 2.5, "SciLit ~ Metric, R2=0.26, p<0.001", col="red", pos=4)
```
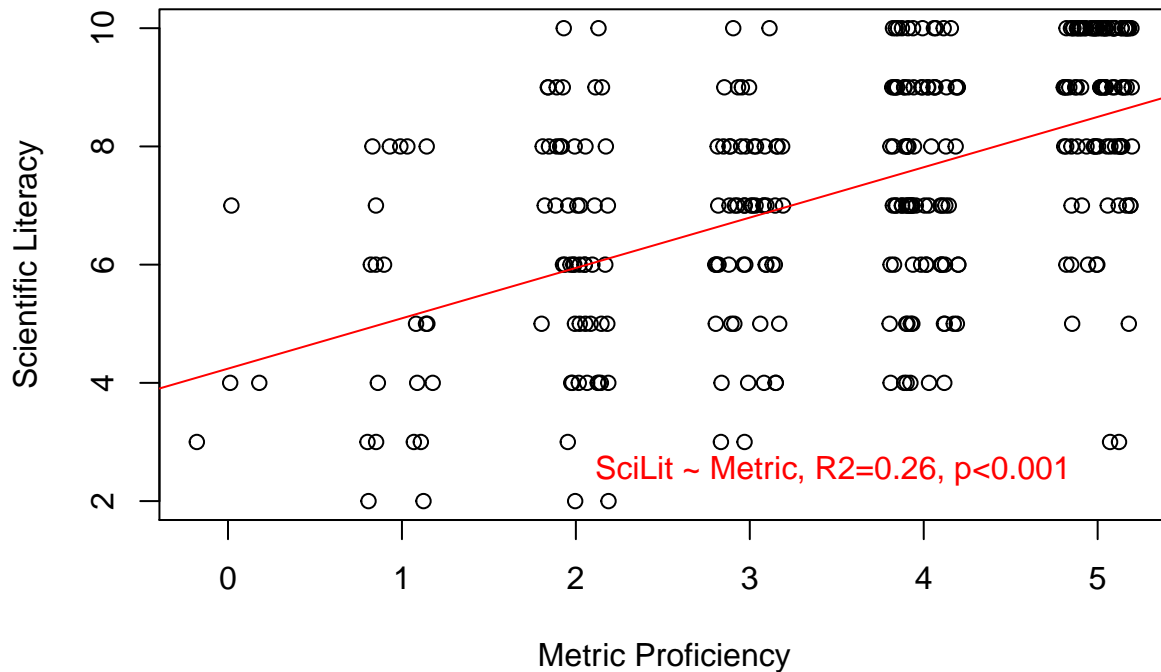


```
#abline(m2, col='blue')
summary(m2)
```

```
##
## Call:
## lm(formula = tb$metric ~ tb$age)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.3338 -1.2670  0.3655  0.8321  1.7330
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.792915   0.160471   17.404  < 2e-16 ***
## tb$age      0.023706   0.004538    5.224 3.19e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.265 on 314 degrees of freedom
## Multiple R-squared:  0.07997,    Adjusted R-squared:  0.07704
## F-statistic: 27.29 on 1 and 314 DF,  p-value: 3.189e-07
```

```
m2 = lm(tb$SciLitScore ~ tb$metric + tb$age)
anova(m1, m2)
```

```
## Analysis of Variance Table
##
## Model 1: tb$SciLitScore ~ tb$metric
## Model 2: tb$SciLitScore ~ tb$metric + tb$age
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1    314 1036.4
```

```
## 2     313 1005.0  1    31.388 9.7759 0.001934 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
m3 = lm(tb$SciLitScore ~ tb$metric + tb$age + tb$gender)
summary(m3)
```

```
##
## Call:
## lm(formula = tb$SciLitScore ~ tb$metric + tb$age + tb$gender)
##
## Residuals:
##      Min      1Q  Median      3Q     Max
## -6.4593 -1.1806  0.1239  1.2225  4.2619
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      5.697118   1.078316   5.283 2.39e-07 ***
## tb$metric        0.721252   0.081181   8.885  < 2e-16 ***
## tb$age           0.008692   0.007384   1.177    0.240
## tb$genderFemale -1.575399   1.033337  -1.525    0.128
## tb$genderMale   -0.699676   1.037870  -0.674    0.501
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.76 on 311 degrees of freedom
## Multiple R-squared:  0.3279, Adjusted R-squared:  0.3193
## F-statistic: 37.93 on 4 and 311 DF,  p-value: < 2.2e-16
```

```r
anova(m2,m3)
```

```
## Analysis of Variance Table
##
## Model 1: tb$SciLitScore ~ tb$metric + tb$age
## Model 2: tb$SciLitScore ~ tb$metric + tb$age + tb$gender
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1    313 1004.97
## 2    311  962.91  2    42.058 6.7919 0.001297 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
m4 = lm(tb$SciLitScore ~ tb$metric + tb$age + tb$country)
anova(m2, m4)
```

```
## Analysis of Variance Table
##
## Model 1: tb$SciLitScore ~ tb$metric + tb$age
## Model 2: tb$SciLitScore ~ tb$metric + tb$age + tb$country
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    313 1005.0
## 2    312 1002.9  1    2.0903 0.6503 0.4206
```

```r
#text(2, 2, "SciLit ~ Metric + Age, R2=0.29, p=2.8E-14", col="blue", pos=4)

plot( tb$SciAttitude ~ jitter(tb$metric), ylim=c(0.5,4), xlab='Metric Proficiency', ylab='Attitude towa
m1 = lm( tb$SciAttitude ~ tb$metric )
m2 = lm( tb$SciAttitude ~ tb$metric + tb$age )
```

```
abline(m1, col='red')
abline(m2, col='blue')
```

```
## Warning in abline(m2, col = "blue"): only using the first two of 3
## regression coefficients
```

```
summary(m1)
```

```
##
## Call:
## lm(formula = tb$SciAttitude ~ tb$metric)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -2.71893 -0.71893 -0.01881  0.98119  1.88083
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.51941    0.16269   9.340  < 2e-16 ***
## tb$metric    0.29988    0.04303   6.968 1.89e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.006 on 314 degrees of freedom
## Multiple R-squared:  0.1339, Adjusted R-squared:  0.1312
## F-statistic: 48.56 on 1 and 314 DF,  p-value: 1.892e-11
```

```
summary(m2)
```
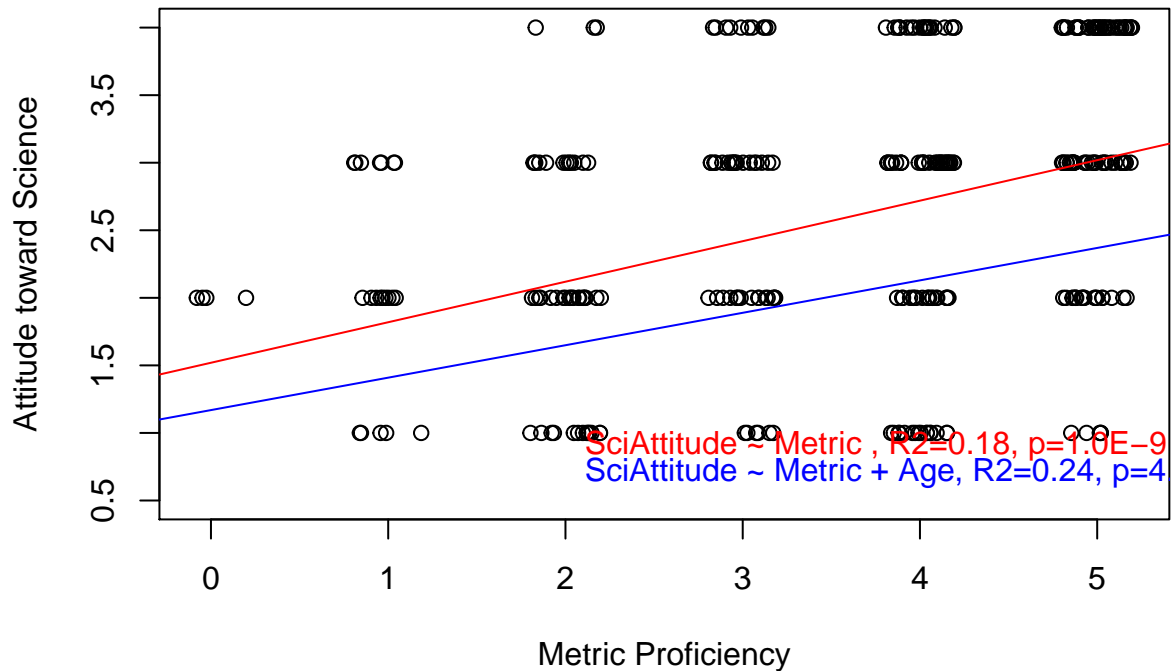
```
##
## Call:
## lm(formula = tb$SciAttitude ~ tb$metric + tb$age)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.7592 -0.6237  0.1654  0.7173  1.9959
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.169078   0.172640   6.772 6.30e-11 ***
## tb$metric    0.240005   0.043314   5.541 6.39e-08 ***
## tb$age       0.017748   0.003631   4.888 1.63e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9711 on 313 degrees of freedom
## Multiple R-squared:  0.1954, Adjusted R-squared:  0.1902
## F-statistic:    38 on 2 and 313 DF,  p-value: 1.686e-15
```

```
anova(m1, m2)
```

```
## Analysis of Variance Table
##
## Model 1: tb$SciAttitude ~ tb$metric
## Model 2: tb$SciAttitude ~ tb$metric + tb$age
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
```

```
## 1      314 317.73
## 2      313 295.19  1    22.534 23.893 1.63e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
text(2, 0.9, "SciAttitude ~ Metric , R2=0.18, p=1.0E-9", col="red", pos=4)
text(2, 0.7, "SciAttitude ~ Metric + Age, R2=0.24, p=4.7E-12", col="blue", pos=4)
```



```
plot( tb$SciAttitude ~ jitter(tb$age), ylab='Attitude toward Science', xlab='Age')
m2 = lm( tb$SciAttitude ~ tb$age + tb$metric)
abline(m2, col='blue')
```

```
## Warning in abline(m2, col = "blue"): only using the first two of 3
## regression coefficients
```

```
text(30, 1.7, "SciAttitude ~ Metric + Age, R2=0.24, p=4.7E-12", col="blue", pos=4)
```
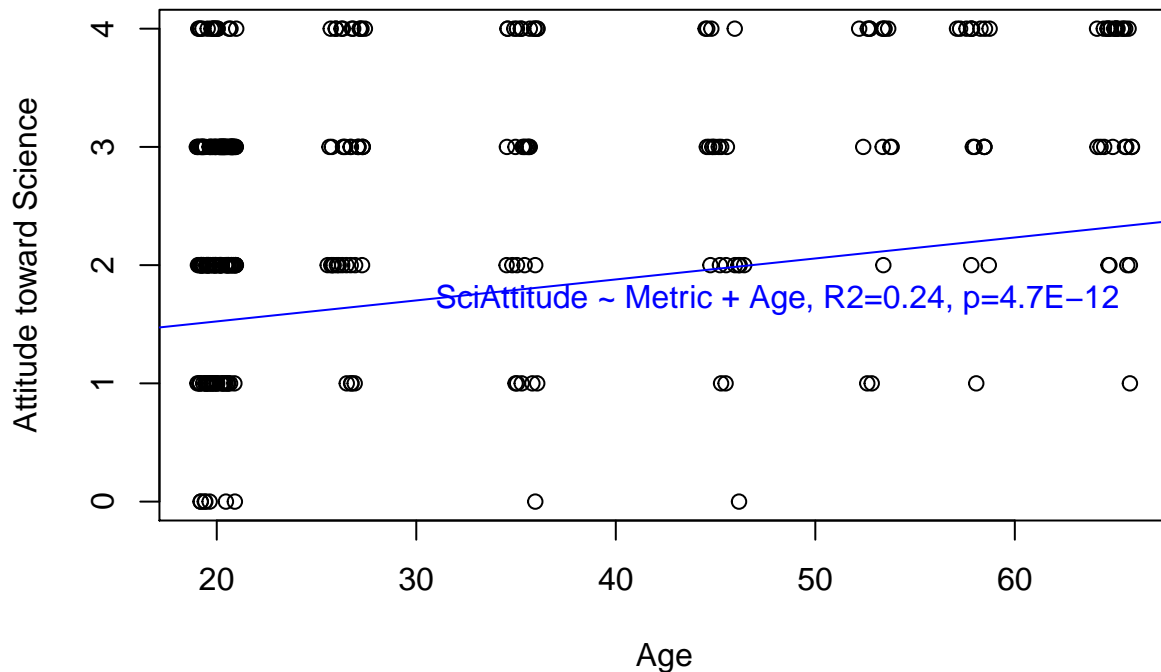
```
summary(lm(tb$SciAttitude ~ tb$metric + tb$age + tb$gender + tb$country  )) #age is signicant!!!
```

```
##
## Call:
## lm(formula = tb$SciAttitude ~ tb$metric + tb$age + tb$gender +
##     tb$country)
##
## Residuals:
##     Min     1Q  Median     3Q     Max
## -2.7560 -0.6046  0.1531  0.6321  2.0073
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     0.511175   0.598878   0.854  0.39401
## tb$metric       0.203966   0.044651   4.568 7.11e-06 ***
## tb$age          0.012800   0.004014   3.189  0.00157 **
## tb$genderFemale 0.738602   0.562717   1.313  0.19030
## tb$genderMale   1.175483   0.564200   2.083  0.03803 *
## tb$country      0.078949   0.150618   0.524  0.60054
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9561 on 310 degrees of freedom
## Multiple R-squared:  0.2275, Adjusted R-squared:  0.215
## F-statistic: 18.26 on 5 and 310 DF,  p-value: 6.96e-16
```

```
#but this might be a bias in the sample
# 1) there is many faculty
# 2) people took the sample may be interested in the metric and science at the first place?!
summary(lm(tb$SciAttitude ~ tb$metric + tb$age + tb$gender + tb$country + tb$degree  )) #age is signica
```

```
##
## Call:
```

```
## lm(formula = tb$SciAttitude ~ tb$metric + tb$age + tb$gender +
##     tb$country + tb$degree)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6181 -0.5932  0.1187  0.7031  1.9256
##
## Coefficients:
##                                                   Estimate Std. Error
## (Intercept)                                        0.580169   0.619008
## tb$metric                                          0.193036   0.047126
## tb$age                                             0.013517   0.004596
## tb$genderFemale                                    0.709856   0.568183
## tb$genderMale                                      1.134403   0.573535
## tb$country                                         0.069300   0.152319
## tb$degreeBachelor Degree in Science or equivalent  0.058631   0.173781
## tb$degreeHigh School or equivalent                -0.166874   0.186299
## tb$degreeM.D. or equivalent                        0.702902   0.700118
## tb$degreeMaster Degree or equivalent               0.146384   0.238620
## tb$degreePh.D. or equivalent                      -0.126356   0.218304
##                                                   t value Pr(>|t|)
## (Intercept)                                         0.937  0.34937
## tb$metric                                           4.096 5.39e-05 ***
## tb$age                                              2.941  0.00352 **
## tb$genderFemale                                     1.249  0.21250
## tb$genderMale                                       1.978  0.04884 *
## tb$country                                          0.455  0.64946
## tb$degreeBachelor Degree in Science or equivalent   0.337  0.73606
## tb$degreeHigh School or equivalent                 -0.896  0.37110
## tb$degreeM.D. or equivalent                         1.004  0.31619
## tb$degreeMaster Degree or equivalent                0.613  0.54003
## tb$degreePh.D. or equivalent                       -0.579  0.56315
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9562 on 305 degrees of freedom
## Multiple R-squared:  0.2399, Adjusted R-squared:  0.215
## F-statistic: 9.626 on 10 and 305 DF,  p-value: 5.893e-14
```

```
summary(lm(tb$SciAttitude ~ tb$SciLitScore))
```

```
##
## Call:
## lm(formula = tb$SciAttitude ~ tb$SciLitScore)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8741 -0.7065  0.1259  0.9582  2.1320
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     1.36493    0.20383   6.696 9.86e-11 ***
## tb$SciLitScore  0.16769    0.02694   6.224 1.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## 
## Residual standard error: 1.02 on 314 degrees of freedom
## Multiple R-squared:  0.1098, Adjusted R-squared:  0.107
## F-statistic: 38.74 on 1 and 314 DF,  p-value: 1.55e-09
```

```
summary(lm(tb$SciAttitude ~ tb$SciLitScore + tb$metric))
```

```
## 
## Call:
## lm(formula = tb$SciAttitude ~ tb$SciLitScore + tb$metric)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6560 -0.6931  0.1269  0.8354  1.9726
## 
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     1.10736    0.20690   5.352 1.68e-07 ***
## tb$SciLitScore  0.09717    0.03081   3.154  0.00177 **
## tb$metric       0.21712    0.04989   4.352 1.83e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.9919 on 313 degrees of freedom
## Multiple R-squared:  0.1606, Adjusted R-squared:  0.1552
## F-statistic: 29.94 on 2 and 313 DF,  p-value: 1.261e-12
```
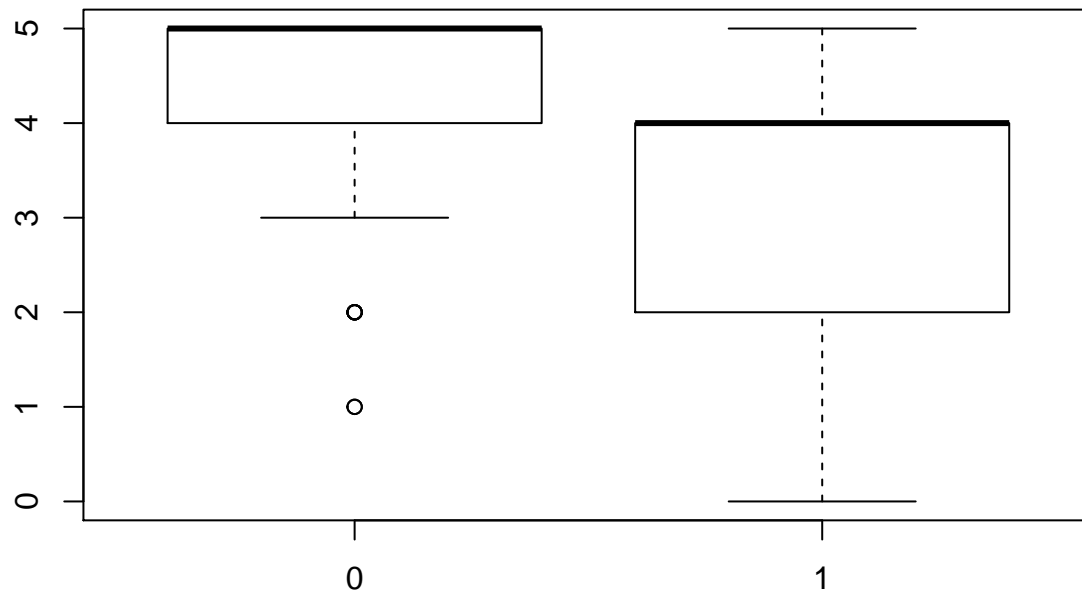
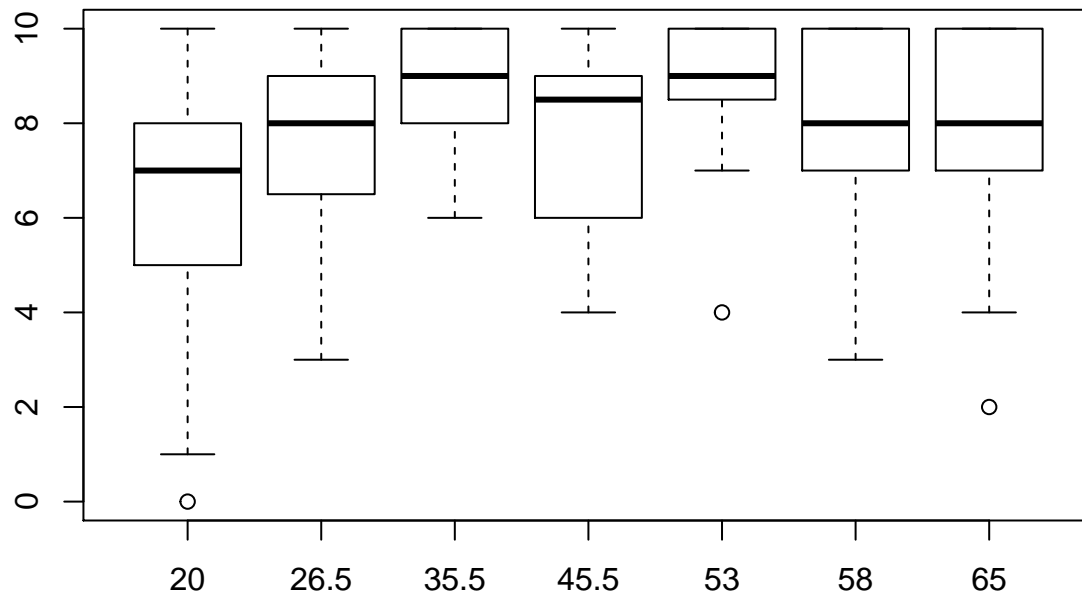```
boxplot(tb$SciAttitude ~ tb$country)
```
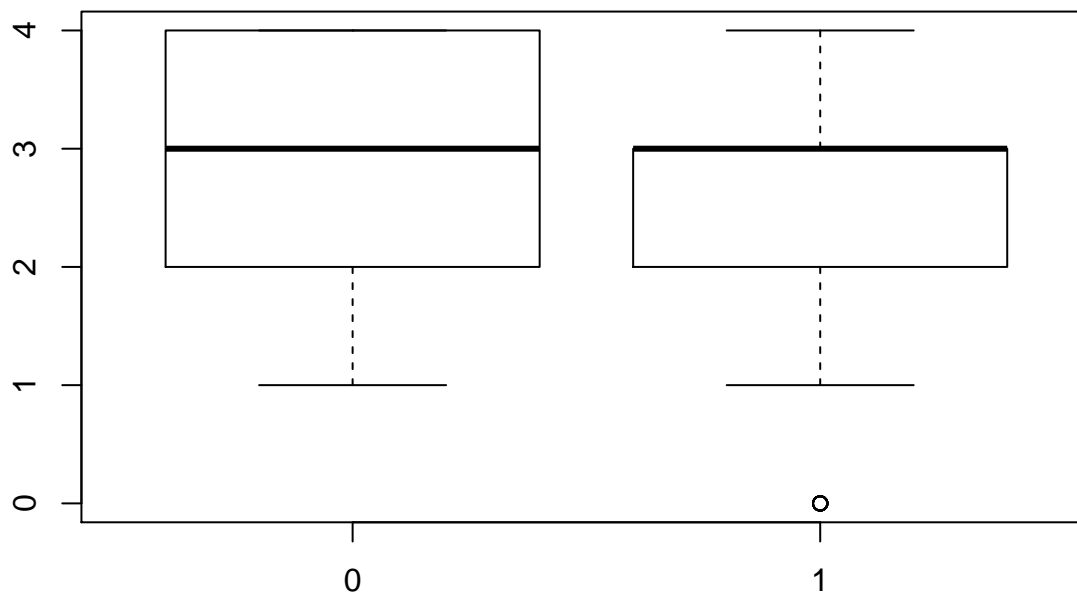


```
boxplot(tb$SciLitScore ~ tb$country)
```
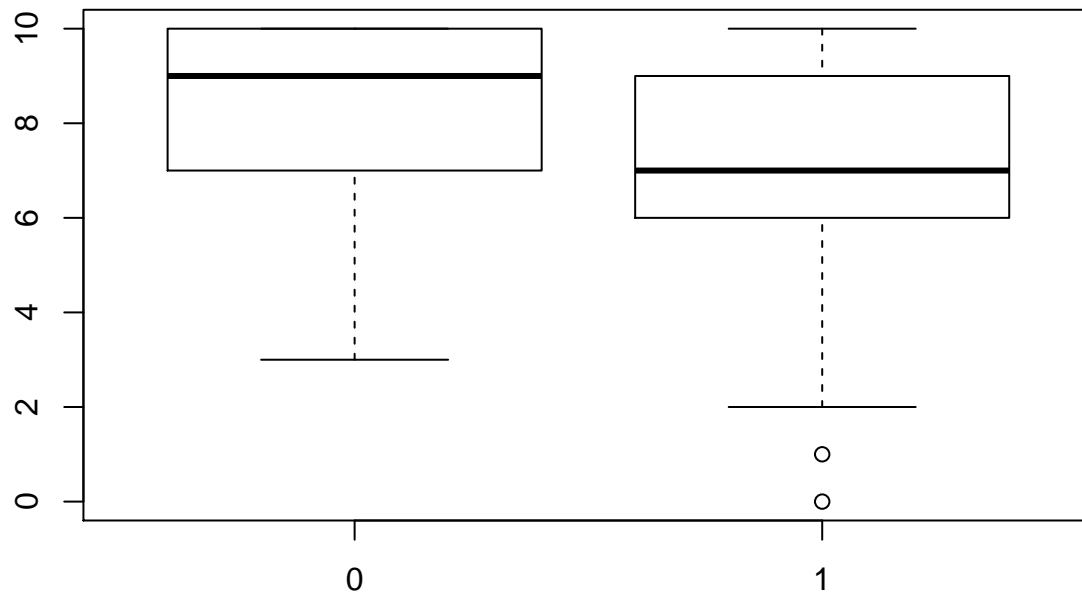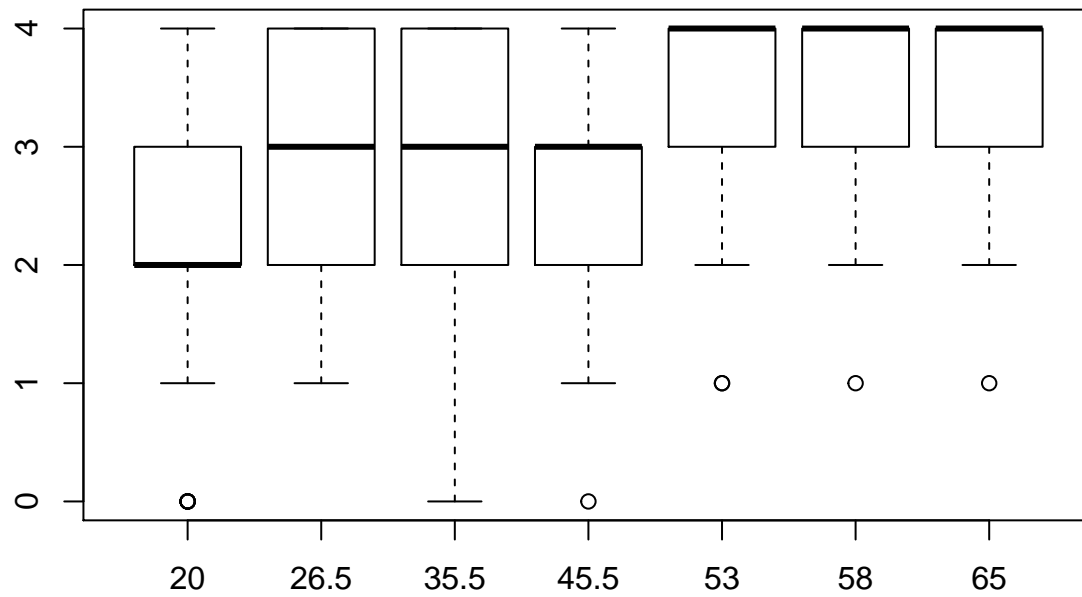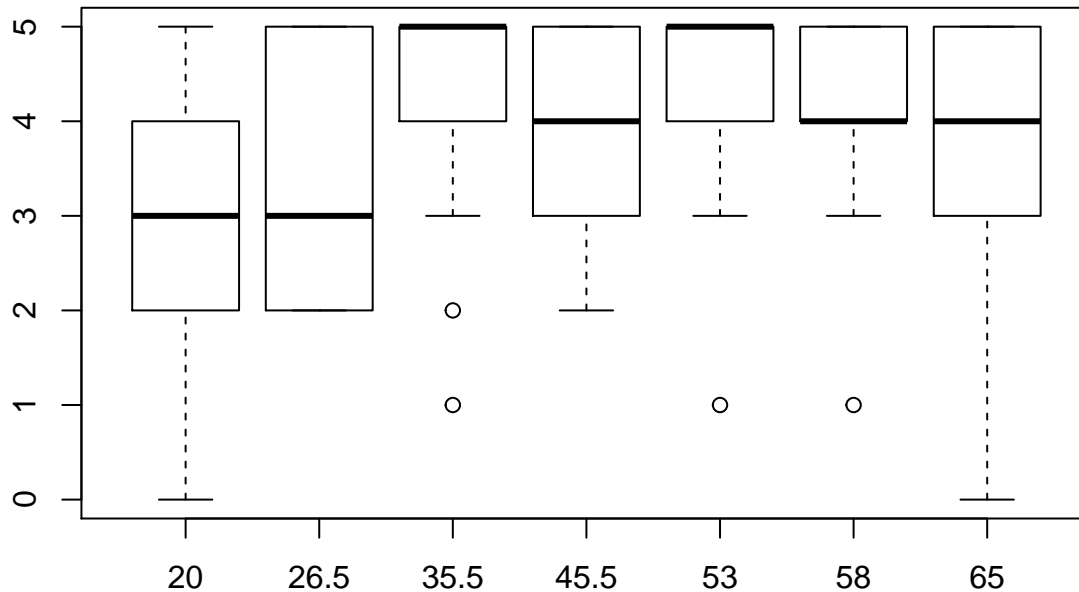
```
boxplot(tb$metric ~ tb$country)
```



```
boxplot( tb$SciLitScore ~ tb$age )
```

```r
boxplot( tb$SciAttitude ~ tb$age )
```



```r
boxplot( tb$metric ~ tb$age )
```

```
###########
# remove phD from the samples
#
summary(tb[, 1:5])
```

```
##     gender               age          country          degree
## Length:316         Min.   :20.0   Min.   :0.0000   Length:316
## Class :character   1st Qu.:20.0   1st Qu.:1.0000   Class :character
## Mode  :character   Median :20.0   Median :1.0000   Mode  :character
##                    Mean   :31.7   Mean   :0.8323
##                    3rd Qu.:45.5   3rd Qu.:1.0000
##                    Max.   :65.0   Max.   :1.0000
##     metric
## Min.   :0.000
## 1st Qu.:3.000
## Median :4.000
## Mean   :3.544
## 3rd Qu.:5.000
## Max.   :5.000
```

```
tb3 = tb[ - grep('Ph.D.', tb$degree)  , ]
summary(tb3)
```

```
##     gender               age           country           degree
## Length:269         Min.   :20.00   Min.   :0.0000   Length:269
## Class :character   1st Qu.:20.00   1st Qu.:1.0000   Class :character
## Mode  :character   Median :20.00   Median :1.0000   Mode  :character
##                    Mean   :29.44   Mean   :0.8439
##                    3rd Qu.:35.50   3rd Qu.:1.0000
##                    Max.   :65.00   Max.   :1.0000
##     metric         SciAttitude     SciLitScore
## Min.   :0.000   Min.   :0.00   Min.   : 0.000
## 1st Qu.:2.000   1st Qu.:2.00   1st Qu.: 6.000
## Median :4.000   Median :3.00   Median : 7.000
## Mean   :3.435   Mean   :2.55   Mean   : 7.074
## 3rd Qu.:5.000   3rd Qu.:3.00   3rd Qu.: 9.000
```

```
##  Max.   :5.000   Max.   :4.00   Max.   :10.000
```

```r
summary(lm(tb3$SciAttitude ~ tb3$metric + tb3$age + tb3$gender + tb3$country + tb3$degree  ))
```

```
##
## Call:
## lm(formula = tb3$SciAttitude ~ tb3$metric + tb3$age + tb3$gender +
##     tb3$country + tb3$degree)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6750 -0.5363  0.1161  0.6397  1.8845
##
## Coefficients:
##                                                 Estimate Std. Error
## (Intercept)                                     2.005844   0.994526
## tb3$metric                                      0.162736   0.051019
## tb3$age                                         0.015241   0.005106
## tb3$genderFemale                               -0.721754   0.949597
## tb3$genderMale                                 -0.319354   0.960012
## tb3$country                                     0.120446   0.170650
## tb3$degreeBachelor Degree in Science or equivalent  0.080689   0.173760
## tb3$degreeHigh School or equivalent            -0.150878   0.184590
## tb3$degreeM.D. or equivalent                    0.755129   0.692802
## tb3$degreeMaster Degree or equivalent           0.175136   0.237208
##                                                 t value Pr(>|t|)
## (Intercept)                                       2.017  0.04474 *
## tb3$metric                                        3.190  0.00160 **
## tb3$age                                           2.985  0.00311 **
## tb3$genderFemale                                 -0.760  0.44791
## tb3$genderMale                                   -0.333  0.73966
## tb3$country                                       0.706  0.48094
## tb3$degreeBachelor Degree in Science or equivalent   0.464  0.64277
## tb3$degreeHigh School or equivalent              -0.817  0.41447
## tb3$degreeM.D. or equivalent                      1.090  0.27674
## tb3$degreeMaster Degree or equivalent             0.738  0.46099
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9451 on 259 degrees of freedom
## Multiple R-squared:  0.2354, Adjusted R-squared:  0.2089
## F-statistic: 8.861 on 9 and 259 DF,  p-value: 1.249e-11
```

```r
#age is still signicant after PhD are removed from the sample
```