Preprocess survey data for metrics, scientific literacy and attitude

H Qin

May 18, 2020

Learning Objectives

- Load external tabular data from a .csv file into R.
- Manipulate string and categorical data in R
- Find iregular inputs and correct them
- Generate summary and clean results

References: ABC News: https://www.youtube.com/watch?v=1cPeZLCVWTw&feature=youtu.be

Productive failure http://manukapur.com/productive-failure/

Check files in the working directory

```
rm(list=ls())
list.files()
```

```
## [1] "Learning_R_by_metricExample.ppt" "metric_survey_data.csv"
## [3] "metric_survey_form.pdf"          "metric_v3.html"
## [5] "metric_v3.Rmd"                   "metric-attitude-literacy.csv"
```

Read the survey data in csv format

```
# colClass specify that all columns will be treated as characters for now.
# tb.ori = read.csv("metric_survey_data.csv", colClass=rep("character", 24))
tb.ori = read.csv("metric_survey_data.csv", stringsAsFactors = FALSE)
?str
str(tb.ori);
```

```
## 'data.frame':    318 obs. of  24 variables:
##  $ Timestamp
##  $ Please.indicate.your.gender
##  $ Please.indicate.your.age.category
##  $ What.is.the.highest.education.that.you.received.or.are.pursing.
##  $ Please.indicate.the.country.in.which.you.grew.up.
##  $ Light.is.both.a.wave.and.a.particle
##  $ A.man.is.2.16.meters.tall..Is.this.person.suited.to.be.a.good.professional.basketball.player.
##  $ A.30.year.old.scientist.found.a.6.million.year.old.fossil..When.this.scientist.becomes.35.years.ol
##  $ X.Kilo..means
##  $ X145.mm..._.__.m
##  $ Do.you.agree.that.organic.food.should.be.DNA.free.food.
##  $ A.person.s.pant.inseam.measures.35.centimeters.
##  $ The.weather.forecast.shows.a.high.of.32.degrees.Celcius..what.should.you.wear.
##  $ What.is.an.electron.attracted.to.
##  $ Early.human.once.lived.with.dinosaurs.
##  $ Lasers.work.by.focusing.sound.waves
##  $ The.continents.have.been.moving.their.location.for.millions.of.years.and.will.continue.to.move.
##  $ Antibiotics.kills.viruses.as.well.as.bacteria.
##  $ Electrons.are.smaller.than.atoms
##  $ The.center.of.the.earth.is.very.hot.
##  $ My.religious.views.are.more.important.than.scientific.views.
##  $ For.me..in.my.daily.life..it.is.not.important.to.know.about.science.
##  $ Science.and.technology.are.making.our.lives.healthier..easier.and.more.comfortable.
##  $ The.benefits.of.science.are.greater.than.any.harmful.effects.it.may.have.
```

```r
tb.ori$Timestamp
```

```
##   [1] "3/5/2013 14:34:19"  "3/5/2013 14:47:37"  "3/5/2013 14:53:48" 
##   [4] "3/5/2013 15:01:34"  "3/5/2013 15:03:33"  "3/5/2013 16:21:51" 
##   [7] "3/5/2013 16:22:06"  "3/5/2013 16:27:17"  "3/5/2013 18:07:57" 
##  [10] "3/5/2013 18:50:42"  "3/5/2013 19:39:08"  "3/7/2013 5:06:06"  
##  [13] "3/13/2013 14:58:00" "3/18/2013 12:21:55" "3/25/2013 15:19:50"
##  [16] "3/25/2013 15:29:20" "3/25/2013 15:29:24" "3/25/2013 16:41:29"
##  [19] "3/25/2013 17:19:04" "3/25/2013 17:19:05" "3/26/2013 10:41:15"
##  [22] "3/26/2013 10:46:11" "3/26/2013 10:47:41" "3/26/2013 10:49:55"
##  [25] "3/26/2013 11:16:14" "3/26/2013 12:52:59" "3/26/2013 13:23:58"
##  [28] "3/26/2013 14:00:27" "3/26/2013 14:01:52" "3/26/2013 14:02:02"
##  [31] "3/26/2013 14:05:07" "3/26/2013 14:05:36" "3/26/2013 14:05:44"
##  [34] "3/26/2013 14:07:44" "3/26/2013 14:08:59" "3/26/2013 14:09:27"
##  [37] "3/26/2013 14:09:53" "3/26/2013 14:12:19" "3/26/2013 14:12:37"
##  [40] "3/26/2013 14:13:35" "3/26/2013 14:14:55" "3/26/2013 14:15:45"
##  [43] "3/26/2013 14:16:45" "3/26/2013 14:18:45" "3/26/2013 14:18:53"
##  [46] "3/26/2013 14:19:07" "3/26/2013 14:22:21" "3/26/2013 14:22:34"
##  [49] "3/26/2013 14:22:44" "3/26/2013 14:23:16" "3/26/2013 14:25:20"
##  [52] "3/26/2013 14:25:56" "3/26/2013 14:26:39" "3/26/2013 14:28:30"
##  [55] "3/26/2013 17:21:50" "3/27/2013 16:39:01" "3/27/2013 19:37:56"
##  [58] "3/27/2013 19:41:55" "3/27/2013 19:42:17" "3/27/2013 19:42:57"
##  [61] "3/27/2013 19:45:11" "3/27/2013 19:45:56" "3/27/2013 19:51:11"
##  [64] "3/27/2013 19:56:34" "3/27/2013 19:58:53" "3/27/2013 20:07:18"
##  [67] "3/27/2013 20:12:17" "3/27/2013 20:20:15" "3/27/2013 20:22:25"
##  [70] "3/27/2013 20:36:51" "3/27/2013 21:15:36" "3/27/2013 21:19:55"
##  [73] "3/27/2013 21:43:03" "3/27/2013 21:43:45" "3/27/2013 21:43:54"
##  [76] "3/27/2013 22:33:37" "3/27/2013 22:36:12" "3/28/2013 0:57:41" 
##  [79] "3/28/2013 1:22:43"  "3/28/2013 1:41:10"  "3/28/2013 1:49:32" 
##  [82] "3/28/2013 2:15:08"  "3/28/2013 2:16:47"  "3/28/2013 4:19:13" 
##  [85] "3/28/2013 5:37:24"  "3/28/2013 5:37:49"  "3/28/2013 6:01:51" 
##  [88] "3/28/2013 6:03:21"  "3/28/2013 6:55:25"  "3/28/2013 6:57:09" 
##  [91] "3/28/2013 7:09:19"  "3/28/2013 7:15:53"  "3/28/2013 7:25:59" 
##  [94] "3/28/2013 7:29:47"  "3/28/2013 7:43:26"  "3/28/2013 7:44:15" 
##  [97] "3/28/2013 7:45:17"  "3/28/2013 7:53:09"  "3/28/2013 8:05:44" 
## [100] "3/28/2013 8:09:49"  "3/28/2013 8:09:51"  "3/28/2013 8:12:38" 
## [103] "3/28/2013 8:28:19"  "3/28/2013 9:07:03"  "3/28/2013 9:07:04" 
## [106] "3/28/2013 9:11:46"  "3/28/2013 9:46:50"  "3/28/2013 9:49:48" 
## [109] "3/28/2013 9:52:41"  "3/28/2013 9:55:38"  "3/28/2013 11:29:07"
## [112] "3/28/2013 11:44:43" "3/28/2013 11:54:38" "3/28/2013 12:25:10"
## [115] "3/28/2013 13:14:43" "3/28/2013 14:05:34" "3/28/2013 14:18:56"
## [118] "3/28/2013 17:24:58" "3/28/2013 17:46:15" "3/28/2013 23:28:48"
## [121] "3/28/2013 23:32:47" "3/29/2013 3:18:36"  "3/29/2013 4:39:40"
## [124] "3/29/2013 5:31:58"  "3/29/2013 5:42:30"  "3/29/2013 6:10:55"
## [127] "3/29/2013 6:34:41"  "3/29/2013 7:08:44"  "3/29/2013 7:22:27"
## [130] "3/29/2013 8:14:17"  "3/29/2013 8:23:10"  "3/29/2013 8:23:37"
## [133] "3/29/2013 9:03:08"  "3/29/2013 11:02:02" "3/29/2013 11:22:17"
## [136] "3/29/2013 13:25:25" "3/29/2013 14:12:33" "3/29/2013 14:33:45"
## [139] "3/29/2013 16:25:34" "3/29/2013 16:32:02" "3/29/2013 16:34:21"
## [142] "3/29/2013 17:21:24" "3/29/2013 18:33:49" "3/29/2013 20:10:51"
## [145] "3/29/2013 21:42:10" "3/30/2013 9:50:20"  "3/30/2013 11:41:27"
## [148] "3/30/2013 12:20:40" "3/30/2013 12:30:33" "3/30/2013 13:11:47"
## [151] "3/30/2013 13:12:05" "3/30/2013 13:18:26" "3/30/2013 13:39:11"
## [154] "3/30/2013 14:04:25" "3/30/2013 14:12:40" "3/30/2013 14:41:30"
## [157] "3/30/2013 14:43:02" "3/30/2013 16:26:35" "3/30/2013 16:47:20"
## [160] "3/30/2013 17:18:51" "3/30/2013 17:35:30" "3/30/2013 17:56:44"
## [163] "3/30/2013 18:11:15" "3/30/2013 18:25:03" "3/30/2013 18:30:31"
## [166] "3/30/2013 19:14:59" "3/30/2013 19:18:34" "3/30/2013 19:43:26"
## [169] "3/30/2013 19:47:37" "3/30/2013 20:13:35" "3/30/2013 22:01:05"
## [172] "3/30/2013 22:02:00" "3/30/2013 22:18:59" "3/31/2013 0:23:35"
## [175] "3/31/2013 1:45:07"  "3/31/2013 2:00:48"  "3/31/2013 2:46:50"
## [178] "3/31/2013 7:59:28"  "3/31/2013 17:21:59" "3/31/2013 17:25:53"
## [181] "3/31/2013 17:31:05" "3/31/2013 19:30:12" "3/31/2013 23:06:00"
## [184] "4/1/2013 0:07:04"   "4/1/2013 0:50:00"   "4/1/2013 10:31:58"
## [187] "4/1/2013 11:47:05"  "4/1/2013 12:02:16"  "4/1/2013 12:31:17"
## [190] "4/1/2013 15:30:22"  "4/1/2013 15:32:50"  "4/1/2013 15:35:07"
## [193] "4/1/2013 15:36:00"  "4/1/2013 15:37:11"  "4/1/2013 15:39:12"
## [196] "4/1/2013 16:25:24"  "4/2/2013 7:58:36"   "4/2/2013 8:01:36"
## [199] "4/2/2013 12:17:33"  "4/3/2013 6:12:57"   "4/3/2013 9:43:46"
## [202] "4/3/2013 17:02:56"  "4/3/2013 17:45:26"  "4/4/2013 7:10:39"
## [205] "4/4/2013 12:02:30"  "4/4/2013 12:39:09"  "4/5/2013 18:04:03"
## [208] "4/6/2013 11:17:50"  "4/6/2013 11:19:52"  "4/6/2013 11:21:32"
## [211] "4/6/2013 11:23:20"  "4/6/2013 11:24:51"  "4/6/2013 11:29:23"
## [214] "4/6/2013 12:38:21"  "4/7/2013 3:08:22"   "4/7/2013 15:55:55"
## [217] "4/7/2013 16:01:12"  "4/8/2013 19:33:38"  "4/10/2013 16:30:25"
## [220] "4/10/2013 20:00:02" "4/15/2013 2:22:03"  "4/15/2013 20:21:56"
## [223] "4/23/2013 18:19:59" "4/23/2013 18:23:48" "4/23/2013 18:25:57"
## [226] "4/23/2013 18:28:08" "4/23/2013 18:31:49" "4/23/2013 18:35:09"
## [229] "4/23/2013 18:37:13" "4/23/2013 18:39:15" "4/23/2013 18:41:17"
## [232] "4/23/2013 18:43:48" "4/24/2013 6:08:34"  "4/25/2013 17:05:31"
## [235] "4/25/2013 17:08:28" "4/25/2013 17:10:24" "4/25/2013 17:13:53"
## [238] "4/25/2013 17:15:45" "4/25/2013 17:17:27" "4/25/2013 17:19:44"
## [241] "4/25/2013 17:21:21" "4/25/2013 17:22:57" "4/25/2013 17:24:38"
## [244] "4/25/2013 17:26:24" "4/25/2013 17:28:16" "4/25/2013 17:29:54"
## [247] "4/25/2013 17:31:27" "4/25/2013 17:33:14" "4/25/2013 17:34:42"
## [250] "4/25/2013 17:36:30" "4/25/2013 17:40:24" "4/25/2013 17:43:31"
## [253] "4/25/2013 17:46:43" "4/25/2013 17:48:25" "5/2/2013 21:00:00"
## [256] "5/2/2013 21:01:48"  "5/2/2013 21:03:35"  "5/2/2013 21:05:05"
## [259] "5/2/2013 21:06:47"  "5/2/2013 21:08:22"  "5/2/2013 21:09:46"
## [262] "5/2/2013 21:11:16"  "5/2/2013 21:13:03"  "5/2/2013 21:14:40"
## [265] "5/2/2013 21:17:06"  "5/2/2013 21:18:30"  "5/2/2013 21:19:54"
## [268] "5/2/2013 21:21:07"  "5/2/2013 21:22:27"  "5/2/2013 21:24:02"
## [271] "5/2/2013 21:25:36"  "5/2/2013 21:27:01"  "5/2/2013 21:29:02"
## [274] "5/2/2013 21:30:24"  "5/2/2013 21:31:42"  "5/2/2013 21:33:45"
## [277] "5/2/2013 21:35:08"  "5/2/2013 21:36:28"  "5/2/2013 21:38:16"
## [280] "5/2/2013 21:39:45"  "5/2/2013 21:43:06"  "5/2/2013 21:44:46"
## [283] "5/2/2013 21:46:09"  "5/2/2013 21:47:33"  "5/2/2013 21:48:45"
## [286] "5/2/2013 21:50:00"  "5/2/2013 21:51:12"  "5/2/2013 21:52:25"
## [289] "5/2/2013 21:53:47"  "5/2/2013 21:55:07"  "5/2/2013 21:56:29"
## [292] "5/2/2013 21:57:45"  "5/2/2013 21:59:10"  "5/2/2013 22:00:27"
## [295] "5/2/2013 22:01:46"  "5/2/2013 22:02:51"  "5/2/2013 22:04:10"
## [298] "5/2/2013 22:05:49"  "5/2/2013 22:07:00"  "5/2/2013 22:08:13"
## [301] "5/29/2013 22:52:19" "6/12/2013 23:08:09" "6/21/2013 17:49:28"
## [304] "7/11/2013 7:29:26"  "8/5/2013 15:13:31"  "8/6/2013 9:50:14"
## [307] "8/6/2013 9:53:52"   "8/12/2013 14:12:00" "9/4/2013 12:11:37"
## [310] "9/6/2013 15:50:43"  "9/19/2013 18:12:37" "9/19/2013 18:14:44"
## [313] "9/19/2013 18:16:34" "9/19/2013 18:19:06" "9/19/2013 18:20:54"
## [316] "9/19/2013 18:22:47" "9/19/2013 18:24:28" "2/27/2014 18:01:44"
```

```
tb = tb.ori  #make a copy because we will modify the table.
```

Shorten columns names

```
names(tb.ori)
```

```
##  [1] "Timestamp"
##  [2] "Please.indicate.your.gender"
##  [3] "Please.indicate.your.age.category"
##  [4] "What.is.the.highest.education.that.you.received.or.are.pursing."
##  [5] "Please.indicate.the.country.in.which.you.grew.up."
##  [6] "Light.is.both.a.wave.and.a.particle"
##  [7] "A.man.is.2.16.meters.tall..Is.this.person.suited.to.be.a.good.professional.basketball.player."
##  [8] "A.30.year.old.scientist.found.a.6.million.year.old.fossil..When.this.scientist.becomes.35.years
##  [9] "X.Kilo..means"
## [10] "X145.mm...___.m"
## [11] "Do.you.agree.that.organic.food.should.be.DNA.free.food."
## [12] "A.person.s.pant.inseam.measures.35.centimeters."
## [13] "The.weather.forecast.shows.a.high.of.32.degrees.Celcius..what.should.you.wear."
## [14] "What.is.an.electron.attracted.to."
## [15] "Early.human.once.lived.with.dinosaurs."
## [16] "Lasers.work.by.focusing.sound.waves"
## [17] "The.continents.have.been.moving.their.location.for.millions.of.years.and.will.continue.to.move
## [18] "Antibiotics.kills.viruses.as.well.as.bacteria."
## [19] "Electrons.are.smaller.than.atoms"
## [20] "The.center.of.the.earth.is.very.hot."
## [21] "My.religious.views.are.more.important.than.scientific.views."
## [22] "For.me..in.my.daily.life..it.is.not.important.to.know.about.science."
## [23] "Science.and.technology.are.making.our.lives.healthier..easier.and.more.comfortable."
## [24] "The.benefits.of.science.are.greater.than.any.harmful.effects.it.may.have."
```

```
?names
#rename the columns with shortter names for convenience
names(tb) = c("time","gender", "age", "degree", "country", "light", "shaq", "fossil", "kilo", "mm",
        "food","inseam", "weather","electronCharge","earlyHuman",
        "laser", "continents", "antibiotics", "electronSize","earthCenter",
        "religiousView","dailyLife","SciOnLife", "SciEffect")
str(tb)
```

```
## 'data.frame':    318 obs. of  24 variables:
##  $ time          : chr  "3/5/2013 14:34:19" "3/5/2013 14:47:37" "3/5/2013 14:53:48" "3/5/2013 15:01:
##  $ gender        : chr  "Do not wish to answer" "Male" "Female" "Do not wish to answer" ...
##  $ age           : chr  "18-22" "18-22" "31-40" NA ...
##  $ degree        : chr  "Bachelor Degree in Science or equivalent" "High School or equivalent" "High
##  $ country       : chr  "United States" "United States" "United States" "United States" ...
##  $ light         : chr  "TRUE" "TRUE" "TRUE" "Wrong" ...
##  $ shaq          : chr  "Yes" "No" "No" "Yes" ...
##  $ fossil        : chr  "6 million and 5 years old" "6 million and 5 years old" "6 million and 5 yea
##  $ kilo          : chr  "1000 x" "1000 x" "100 x" "1000 x" ...
##  $ mm            : chr  "0.145" "0.145" "1.45" "0.145" ...
##  $ food          : chr  "I don't know" "Dis-agree" "Dis-agree" "Dis-agree" ...
##  $ inseam        : chr  "This person is tall" "This person is short" "This person is short" "This pe
##  $ weather       : chr  "A winter coat" "A Short sleeve shirt" "A light jacket" "A winter coat" ...
##  $ electronCharge: chr  "Negative charge" "Positive charge" "Positive charge" "Positive charge" ...
##  $ earlyHuman    : chr  "FALSE" "FALSE" "TRUE" "FALSE" ...
##  $ laser         : chr  "TRUE" "FALSE" "FALSE" "FALSE" ...
##  $ continents    : chr  "TRUE" "TRUE" "TRUE" "TRUE" ...
##  $ antibiotics   : chr  "FALSE" "FALSE" "FALSE" "FALSE" ...
##  $ electronSize  : chr  "True " "True " "True " "True " ...
##  $ earthCenter   : chr  "TRUE" "TRUE" "TRUE" "TRUE" ...
##  $ religiousView : chr  "Yes" "Yes" "Yes" "No" ...
##  $ dailyLife     : chr  "FALSE" "FALSE" "Neutral" "FALSE" ...
##  $ SciOnLife     : chr  "TRUE" "TRUE" "TRUE" "TRUE" ...
##  $ SciEffect     : chr  "TRUE" "TRUE" "FALSE" "Not sure" ...
```

```r
summary(tb)
```

```
##      time              gender              age               degree
##  Length:318         Length:318         Length:318         Length:318
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##    country             light               shaq              fossil
##  Length:318         Length:318         Length:318         Length:318
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##      kilo                mm                food              inseam
##  Length:318         Length:318         Length:318         Length:318
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##    weather          electronCharge      earlyHuman           laser
##  Length:318         Length:318         Length:318         Length:318
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##   continents        antibiotics        electronSize       earthCenter
##  Length:318         Length:318         Length:318         Length:318
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##  religiousView       dailyLife          SciOnLife          SciEffect
##  Length:318         Length:318         Length:318         Length:318
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
```

Visually check of the renamed columns.

cbind is to combine columns.

substr is to take a portion of string variables.

```
cbind (names(tb), substr(names(tb.ori), 1, 30))
```

```
##       [,1]             [,2]
##  [1,] "time"           "Timestamp"
##  [2,] "gender"         "Please.indicate.your.gender"
##  [3,] "age"            "Please.indicate.your.age.categ"
##  [4,] "degree"         "What.is.the.highest.education."
##  [5,] "country"        "Please.indicate.the.country.in"
##  [6,] "light"          "Light.is.both.a.wave.and.a.par"
##  [7,] "shaq"           "A.man.is.2.16.meters.tall..Is."
##  [8,] "fossil"         "A.30.year.old.scientist.found."
##  [9,] "kilo"           "X.Kilo..means"
## [10,] "mm"             "X145.mm...___.m"
## [11,] "food"           "Do.you.agree.that.organic.food"
## [12,] "inseam"         "A.person.s.pant.inseam.measure"
## [13,] "weather"        "The.weather.forecast.shows.a.h"
## [14,] "electronCharge" "What.is.an.electron.attracted."
## [15,] "earlyHuman"     "Early.human.once.lived.with.di"
## [16,] "laser"          "Lasers.work.by.focusing.sound."
## [17,] "continents"     "The.continents.have.been.movin"
## [18,] "antibiotics"    "Antibiotics.kills.viruses.as.w"
## [19,] "electronSize"   "Electrons.are.smaller.than.ato"
## [20,] "earthCenter"    "The.center.of.the.earth.is.ver"
## [21,] "religiousView"  "My.religious.views.are.more.im"
## [22,] "dailyLife"      "For.me..in.my.daily.life..it.i"
## [23,] "SciOnLife"      "Science.and.technology.are.mak"
## [24,] "SciEffect"      "The.benefits.of.science.are.gr"
```

```
?cbind
```

Change all missing values (skipped questions) to NA

doing repetive task one by one is tedious. we can use nested for-loops

```
# dealing with missing values, add 'NA' to empty answers
# nested for-loops
for( i in 1:length(tb[, 1])) {  #outter for-loop, i for row, from 1 to the last row
  for( j in 5:length(tb[1, ])) {  #inner for-loop, j for column, from 5th to the last column
    # print( paste("i=", i, "j=", j) )
  }
}

#if there is empty cell, we assign a missing value 'NA' there
for( i in 1:length(tb[, 1])) {  #outter for-loop
  for( j in 5:length(tb[1, ])) {  #inner for-loop
    if ( is.na(tb[i, j]) ) {
      # do nothing
    } else if (tb[i,j]=='') {
      tb[i,j] = NA
    }
  }
}
```

```
table(is.na(tb$age))
```

```
## 
## FALSE  TRUE 
##   317     1
```

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

```
##                   gender   age
## 1 Do not wish to answer 18-22
## 2                  Male 18-22
## 3                Female 31-40
## 4 Do not wish to answer  <NA>
## 5                Female 51-55
```

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

```
##   [1] "18-22"                  "18-22"                  "31-40"                 
##   [4] NA                       "51-55"                  "56-60"                 
##   [7] "18-22"                  "41-50"                  "31-40"                 
##  [10] "31-40"                  "18-22"                  "56-60"                 
##  [13] "More than 60 years old" "41-50"                  "23-30"                 
##  [16] "23-30"                  "23-30"                  "18-22"                 
##  [19] "23-30"                  "23-30"                  "23-30"                 
##  [22] "23-30"                  "18-22"                  "23-30"                 
##  [25] "31-40"                  "23-30"                  "23-30"                 
##  [28] "18-22"                  "18-22"                  "23-30"                 
##  [31] "18-22"                  "18-22"                  "18-22"                 
##  [34] "18-22"                  "18-22"                  "18-22"                 
##  [37] "18-22"                  "18-22"                  "18-22"                 
##  [40] "18-22"                  "23-30"                  "18-22"                 
##  [43] "18-22"                  "18-22"                  "18-22"                 
##  [46] "18-22"                  "23-30"                  "23-30"                 
##  [49] "18-22"                  "23-30"                  "23-30"                 
##  [52] "23-30"                  "18-22"                  "18-22"                 
##  [55] "31-40"                  "23-30"                  "18-22"                 
##  [58] "18-22"                  "18-22"                  "18-22"                 
##  [61] "18-22"                  "18-22"                  "23-30"                 
##  [64] "18-22"                  "18-22"                  "18-22"                 
##  [67] "18-22"                  "18-22"                  "18-22"                 
##  [70] "31-40"                  "18-22"                  "18-22"                 
##  [73] "18-22"                  "18-22"                  "18-22"                 
##  [76] "23-30"                  "18-22"                  "18-22"                 
##  [79] "18-22"                  "18-22"                  "18-22"                 
##  [82] "41-50"                  "23-30"                  "56-60"                 
##  [85] "31-40"                  "18-22"                  "18-22"                 
##  [88] "56-60"                  "18-22"                  "31-40"                 
##  [91] "23-30"                  "23-30"                  "18-22"                 
##  [94] "More than 60 years old" "51-55"                  "23-30"                 
##  [97] "More than 60 years old" "23-30"                  "18-22"                 
## [100] "23-30"                  "18-22"                  "51-55"                 
## [103] "18-22"                  "56-60"                  "41-50"                 
## [106] "More than 60 years old" "18-22"                  "18-22"                 
## [109] "18-22"                  "41-50"                  "More than 60 years old"
## [112] "56-60"                  "51-55"                  "18-22"
## [115] "18-22"                  "41-50"                  "23-30"
## [118] "18-22"                  "51-55"                  "More than 60 years old"
## [121] "41-50"                  "More than 60 years old" "More than 60 years old"
## [124] "More than 60 years old" "31-40"                  "More than 60 years old"
## [127] "31-40"                  "31-40"                  "56-60"
## [130] "56-60"                  "56-60"                  "56-60"
## [133] "56-60"                  "41-50"                  "41-50"
## [136] "More than 60 years old" "51-55"                  "More than 60 years old"
## [139] "31-40"                  "31-40"                  "More than 60 years old"
## [142] "51-55"                  "41-50"                  "41-50"
## [145] "18-22"                  "31-40"                  "18-22"
## [148] "51-55"                  "41-50"                  "41-50"
## [151] "41-50"                  "41-50"                  "More than 60 years old"
## [154] "More than 60 years old" "18-22"                  "56-60"
## [157] "41-50"                  "More than 60 years old" "51-55"
## [160] "18-22"                  "31-40"                  "56-60"
## [163] "56-60"                  "51-55"                  "41-50"
## [166] "31-40"                  "23-30"                  "51-55"
## [169] "31-40"                  "31-40"                  "18-22"
## [172] "18-22"                  "23-30"                  "23-30"
## [175] "51-55"                  "31-40"                  "31-40"
## [178] "31-40"                  "18-22"                  "More than 60 years old"
## [181] "31-40"                  "41-50"                  "18-22"
## [184] "More than 60 years old" "56-60"                  "More than 60 years old"
## [187] "23-30"                  "18-22"                  "18-22"
## [190] "18-22"                  "18-22"                  "31-40"
## [193] "23-30"                  "18-22"                  "18-22"
## [196] "More than 60 years old" "18-22"                  "31-40"
## [199] "23-30"                  "More than 60 years old" "More than 60 years old"
## [202] "18-22"                  "23-30"                  "41-50"
## [205] "More than 60 years old" "More than 60 years old" "18-22"
## [208] "41-50"                  "31-40"                  "31-40"
## [211] "31-40"                  "51-55"                  "41-50"
## [214] "18-22"                  "More than 60 years old" "56-60"
## [217] "More than 60 years old" "18-22"                  "56-60"
## [220] "Do not wish to answer"  "23-30"                  "41-50"
## [223] "18-22"                  "18-22"                  "18-22"
## [226] "18-22"                  "18-22"                  "18-22"
## [229] "18-22"                  "18-22"                  "18-22"
## [232] "18-22"                  "23-30"                  "18-22"
## [235] "18-22"                  "18-22"                  "18-22"
## [238] "18-22"                  "18-22"                  "18-22"
## [241] "18-22"                  "23-30"                  "18-22"
## [244] "18-22"                  "18-22"                  "18-22"
## [247] "18-22"                  "18-22"                  "18-22"
## [250] "18-22"                  "18-22"                  "18-22"
## [253] "18-22"                  "18-22"                  "18-22"
## [256] "18-22"                  "18-22"                  "18-22"
## [259] "18-22"                  "18-22"                  "18-22"
## [262] "18-22"                  "18-22"                  "18-22"
## [265] "18-22"                  "18-22"                  "18-22"
## [268] "18-22"                  "18-22"                  "18-22"
## [271] "18-22"                  "18-22"                  "18-22"
## [274] "18-22"                  "18-22"                  "18-22"
## [277] "18-22"                  "18-22"                  "18-22"
## [280] "18-22"                  "23-30"                  "18-22"
## [283] "18-22"                  "18-22"                  "18-22"
## [286] "18-22"                  "18-22"                  "18-22"
## [289] "18-22"                  "18-22"                  "18-22"
## [292] "18-22"                  "18-22"                  "18-22"
## [295] "18-22"                  "18-22"                  "18-22"
## [298] "18-22"                  "18-22"                  "18-22"
## [301] "51-55"                  "More than 60 years old" "More than 60 years old"
## [304] "18-22"                  "18-22"                  "18-22"
## [307] "18-22"                  "51-55"                  "31-40"
## [310] "31-40"                  "31-40"                  "31-40"
## [313] "18-22"                  "18-22"                  "18-22"
## [316] "18-22"                  "18-22"                  "23-30"
```

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

#correct some input errors
# If there is no input of 'age'
tb$age[is.na(tb$age)] = 'Do not wish to answer'
table(tb$age)
```

```
##
##                  18-22                  23-30                  31-40
##                    164                     39                     31
##                  41-50                  51-55                  56-60
##                     22                     15                     17
##  Do not wish to answer More than 60 years old
##                      2                     28
```

```
?table
# If there is no input of 'age'

tb$degree [is.na(tb$degree)] = 'Do not wish to answer'
table(tb$degree)
```

```
##
##    Bachelor Degree in Arts or equivalent
##                                       46
## Bachelor Degree in Science or equivalent
##                                      125
##                High School or equivalent
##                                       67
##                       M.D. or equivalent
##                                        2
##              Master Degree or equivalent
##                                       31
##                      Ph.D. or equivalent
##                                       47
```

```
tb$gender[tb$gender=='']='Do not wish to answer'
table(tb$gender)
```

```
##
## Do not wish to answer                Female                   Male
##                      4                   208                    106
```

Now, we need to conver survey data in chacracters into numeric scores

First, convert age categories into numeric values

```
##### create a second table, convert character values to numerical values
tb2 = tb[ ,c(2,4,5)]  #this is the score table, empty space before comma indicate every row
head(tb2)
```

```
##                   gender                                   degree       country
## 1 Do not wish to answer Bachelor Degree in Science or equivalent United States
## 2                  Male                High School or equivalent United States
## 3                Female                High School or equivalent United States
## 4 Do not wish to answer Bachelor Degree in Science or equivalent United States
## 5                Female                High School or equivalent United States
## 6                Female    Bachelor Degree in Arts or equivalent United States
```

```
#calculate the average age for each category
?grep #This is not GRE prep. This is pattern match.
# grep(pattern, x, ignore.case = FALSE, perl = FALSE, value = FALSE,
#     fixed = FALSE, useBytes = FALSE, invert = FALSE)

tb2$age = NA
tb2$age[grep("18-22", tb$age)] = 18/2 + 22/2
tb2$age[grep("23-30", tb$age)] = 23/2 + 30/2
tb2$age[grep("31-40", tb$age)] = 31/2 + 40/2
tb2$age[grep("41-50", tb$age)] = 41/2 + 50/2
tb2$age[grep("51-55", tb$age)] = 51/2 + 55/2
tb2$age[grep("56-60", tb$age)] = 56/2 + 60/2
#> grep("56-60", tb$age)
# [1]   6  12  84  88 104 112 129 130 131 132 133 156 162 163 185 216 219
tb2$age[grep("More than 60 years", tb$age)] = 65
```

Check the age responses

```
table(tb$age) #table is a very useful function (command) for tabulation
```

```
## 
##                  18-22                  23-30                  31-40 
##                    164                     39                     31 
##                  41-50                  51-55                  56-60 
##                     22                     15                     17 
##  Do not wish to answer More than 60 years old 
##                      2                     28
```

```
table(tb2$age)
```

```
## 
##   20 26.5 35.5 45.5   53   58   65 
##  164   39   31   22   15   17   28
```

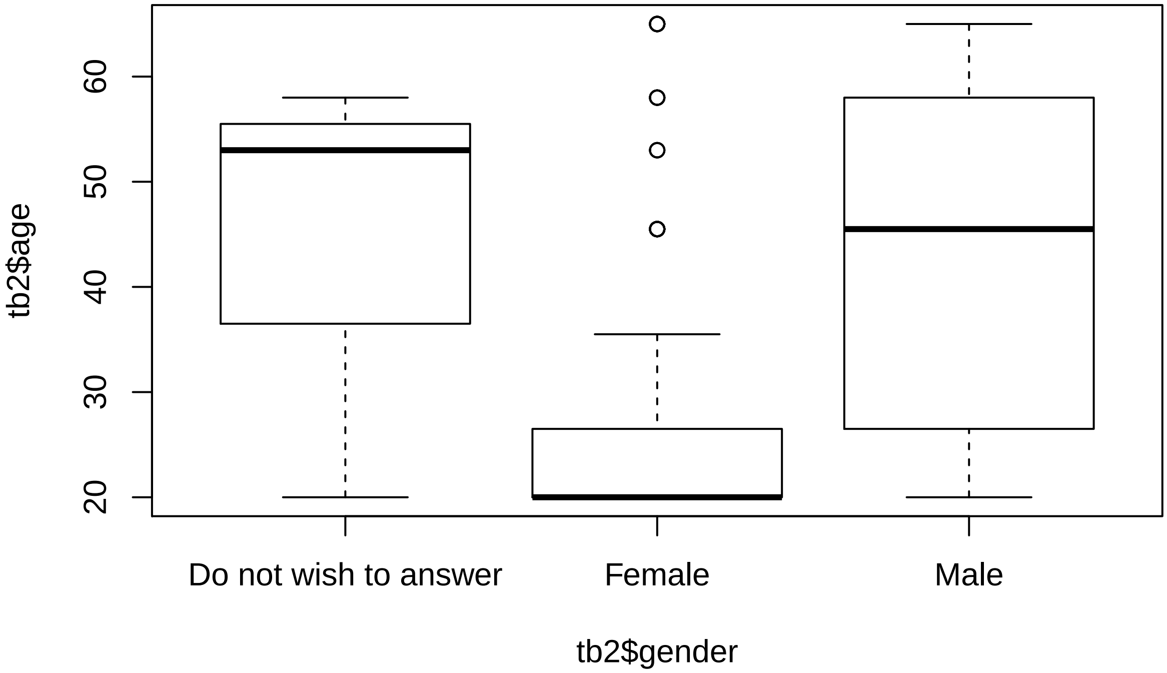
```
#summary(tb2$age)
```

Visualize the age values

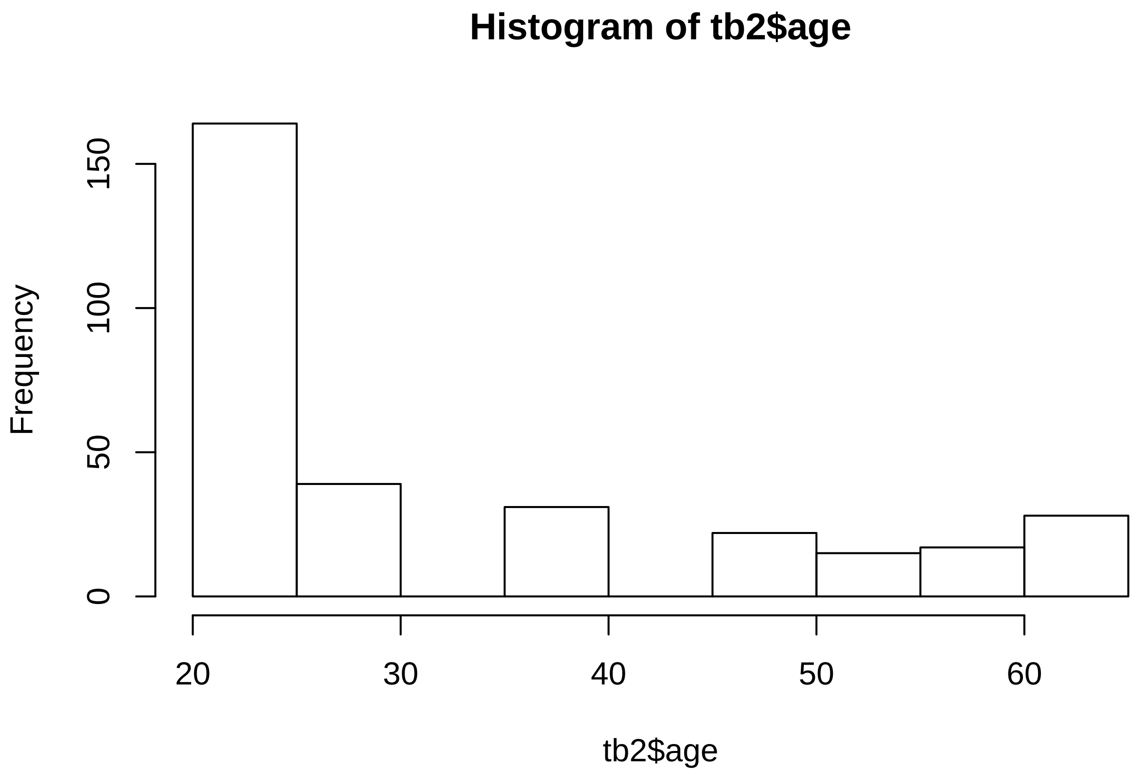
```
table(tb2$age, tb2$gender)
```

```
##
##         Do not wish to answer Female Male
##    20                       1    144   19
##    26.5                     0     24   15
##    35.5                     0     15   16
##    45.5                     0     10   12
##    53                       1      3   11
##    58                       1      5   11
##    65                       0      7   21
```

```
boxplot( tb2$age ~ tb2$gender)
```



```
#histogram of age
hist(tb2$age)
```



Convert country responses into values

```
table( tb$country )  #All the inputed 'countries'
```

```
##
##            Armenia          Australia            Bahamas             Canada
##                  1                  5                  2                  3
##              China            Croatia            Estonia           Ethiopia
##                  3                  1                  1                  1
##             France            Germany              Ghana             Guyana
##                  1                  1                  1                  1
##              India            Jamaica              Kenya            Lebanon
##                  1                  1                  2                  1
##             Mexico        New Zealand             Norway             Poland
##                  1                  1                  1                  3
## Russian Federation             Rwanda            Senegal       South Africa
##                  2                  2                  1                  2
##              Syria  Trinidad & Tobago     United Kingdom      United States
##                  1                  3                  9                264
```

```
tb2$country = 0  #for non-USA countries
tb2$country[tb$country=='United States'] = 1
table( tb2$country )
```

```
##
##   0   1
##  54 264
```

```r
#have a look at some entries
head(tb2)
```

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

```r
#double-check the columns
names(tb2)
```

```
## [1] "gender"  "degree"  "country" "age"
```

The survey contains by 3 categories of questions

1) Metric proficiency
2) Scientific literacy
3) Attitude toward science

Tocalculate the score of each categoriy separately, we need to identify these columns.

```r
### Here are the columns for the 3 categories
metrics = c("shaq", "kilo", "mm", "inseam", "weather")
sciLiteracy = c("light", "fossil", "food", "electronCharge",
                "earlyHuman", "laser", "continents", "antibiotics",
                "electronSize", "earthCenter")
sciAttitude = c("religiousView", "dailyLife", "SciOnLife", "SciEffect")
```

Calculate the metric-proficiency scores

```r
tb2$shaq = 0
tb2$shaq[ tb$shaq=='Yes' ] = 1
tb2$shaq[ tb$shaq=='No' ] = 0
table(tb2$shaq)
```

```
##
##   0   1
##  91 227
```

```r
tb2$kilo = 0
tb2$kilo[ tb$kilo=='1000 x' ] = 1
table(tb2$kilo)
```

```
##
##   0   1
##  31 287
```

```r
tb2$mm=0
tb2$mm[ tb$mm==0.145 ] = 1
table(tb2$mm)
```

```
##
##   0   1
## 118 200
```

```
table(tb$mm)
```

```
##
##          0.0145           0.145            1.45          145000 I do not know.
##              35             200              72               7               1
```

```
tb2$inseam = 0
tb2$inseam[tb$inseam=="This person is short"] = 1
tb2$inseam[tb$inseam=="This person is tall"] = 0
table(tb2$inseam)
```

```
##
##   0   1
## 112 206
```

```
tb2$weather = 0
tb2$weather[tb$weather=="A Short sleeve shirt"] = 1
#tb2$weather[tb$weather=="A winter coat"] = 0
#tb2$weather[tb$weather=="A light jacket"] = 0
table(tb$weather)
```

```
##
##       A light jacket A Short sleeve shirt        A winter coat
##                   44                  204                   40
##         I don't know
##                   29
```

```
table(tb2$weather)
```

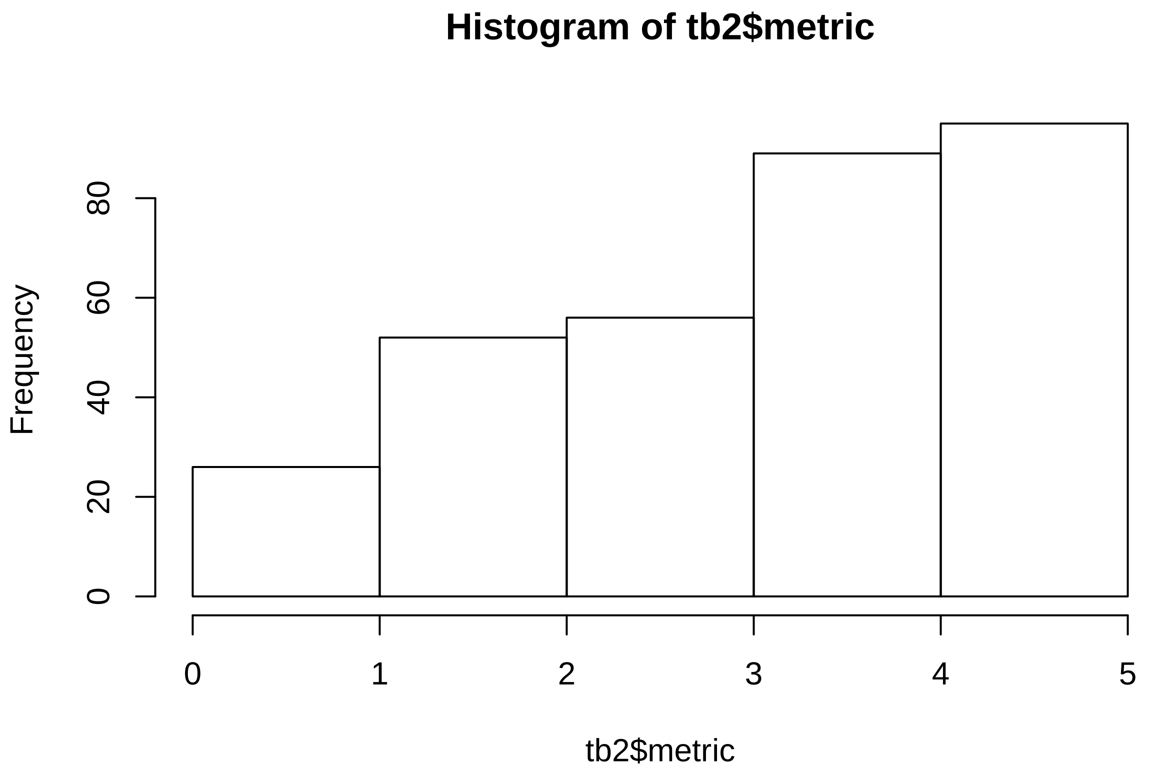
```
##
##   0   1
## 114 204
```

Summarize the metric proficiency score by rows.

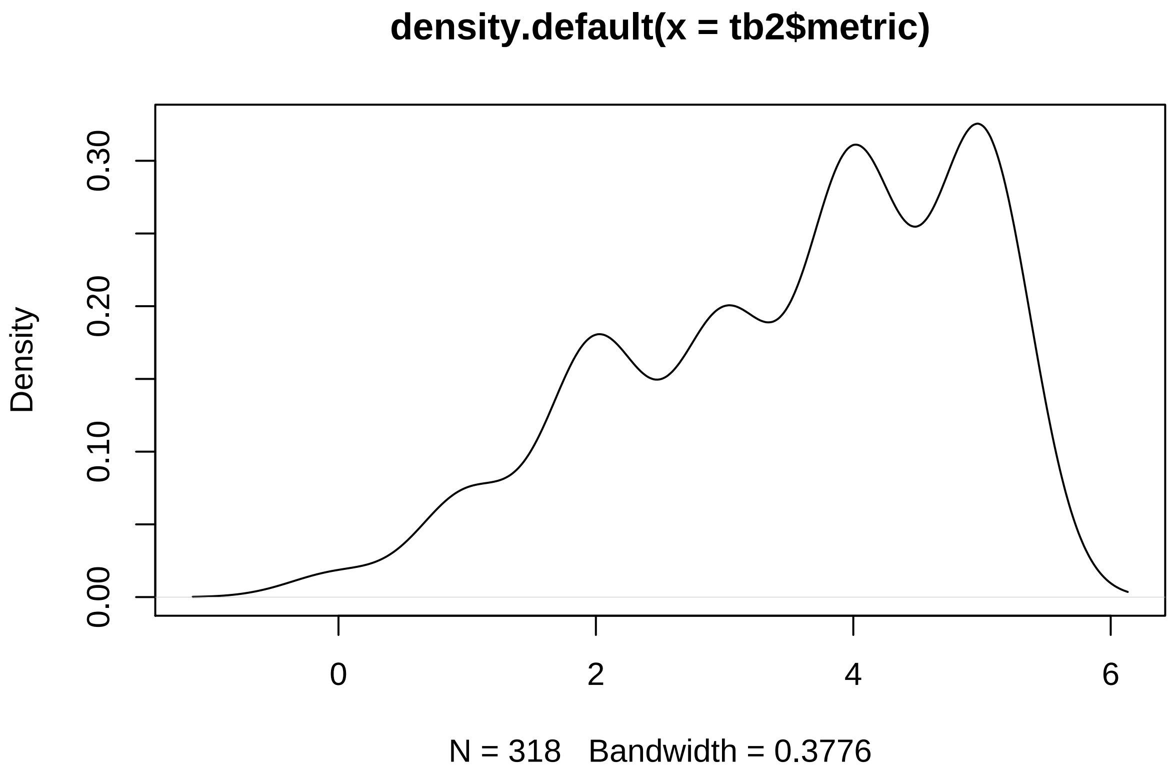
```
# metrics = c("shaq", "kilo", "mm", "inseam", "weather")
# metric total score
print(paste("metrics are: ", metrics));
```

```
## [1] "metrics are:  shaq"    "metrics are:  kilo"    "metrics are:  mm"
## [4] "metrics are:  inseam"  "metrics are:  weather"
```

```
tb2$metric = apply( tb2[ , metrics], MARGIN=1, FUN=sum )
hist(tb2$metric, br=5 )
```



```
plot( density(tb2$metric) )
```



```
#hist(tb2$metric, br=5, probability = TRUE )
```

Calcualte the science attitude scores

```
#sciAttitude = c("religiousView", "dailyLife", "SciOnLife", "SciEffect")
# "My religious views are more important than scientific views
tb2$religiousView = 0
tb2$religiousView[grep("No", tb$religiousView)] = 1
tb2$religiousView[grep("Yes", tb$religiousView)] = 0
table(tb2$religiousView)
```

```
##
##   0   1
## 162 156
```

```
table(tb$religiousView)
```

```
##
## I do not know            No           Yes
##            29           156           130
```

```
# "For me, in my daily life, it is not important to know about science"
tb2$dailyLife = 0
tb2$dailyLife[ tb$dailyLife=='TRUE' ] = 0
tb2$dailyLife[ tb$dailyLife=='FALSE' ] = 1
table(tb2$dailyLife)
```

```
##
##   0   1
##  90 228
```

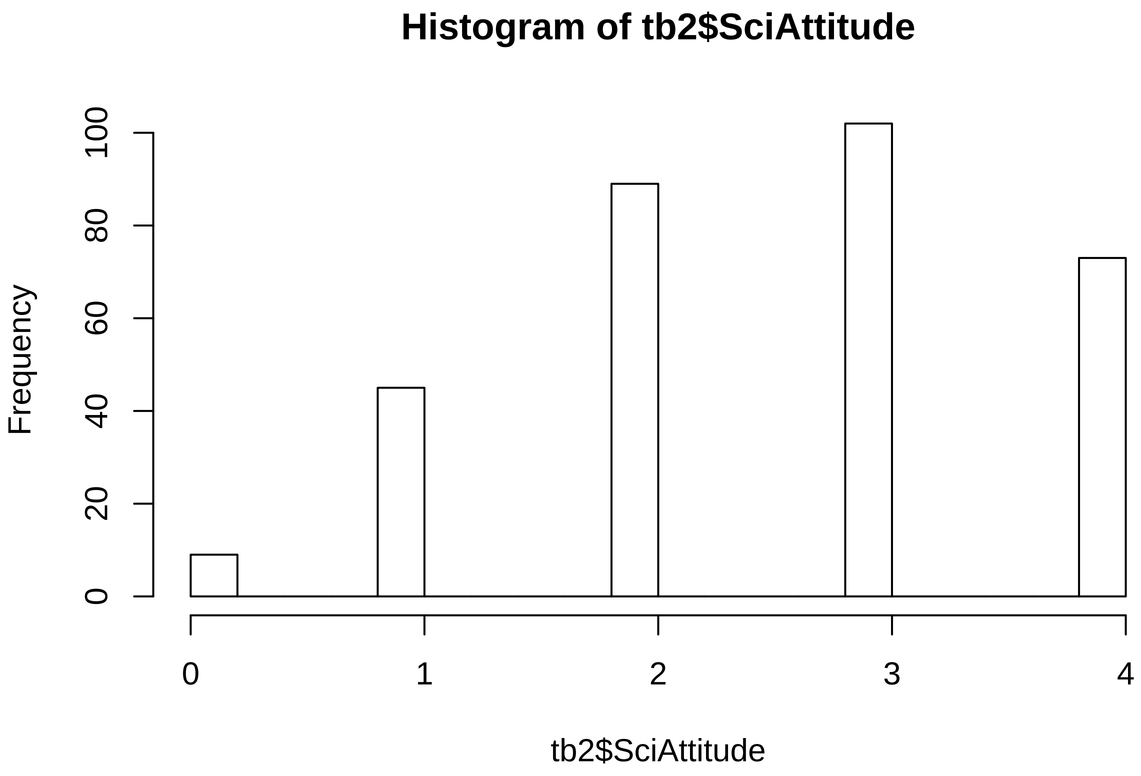
```
# "Science and technology are making our lives healthiers, easiers and more comfortable."
tb2$SciOnLife = 0
tb2$SciOnLife[ tb$SciOnLife=='TRUE' ] = 1
tb2$SciOnLife[ tb$SciOnLife=='FALSE' ] = 0
table(tb2$SciOnLife)
```

```
##
##   0   1
##  47 271
```

```
# "The benefits of sciences are greaters than any harmful effects that it may have."
tb2$SciEffect = 0
tb2$SciEffect[ tb$SciEffect=='TRUE' ] = 1
tb2$SciEffect[ tb$SciEffect=='FALSE' ] = 0
table( tb2$SciEffect )
```

```
##
##   0   1
## 152 166
```

```
#sciAttitude = c("religiousView", "dailyLife", "SciOnLife", "SciEffect")
#Attitude total score
tb2$SciAttitude = apply( tb2[, sciAttitude], MARGIN=1, FUN=sum)
hist(tb2$SciAttitude, br=20)
```

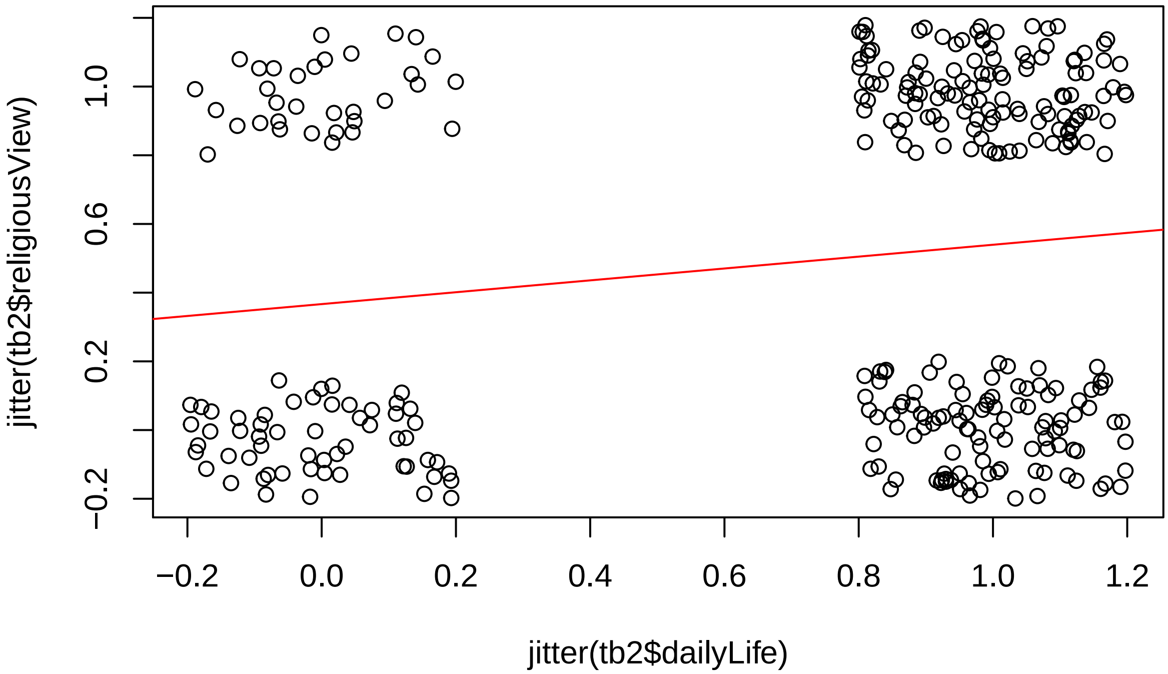


Do responses to religeonus questions correlate?

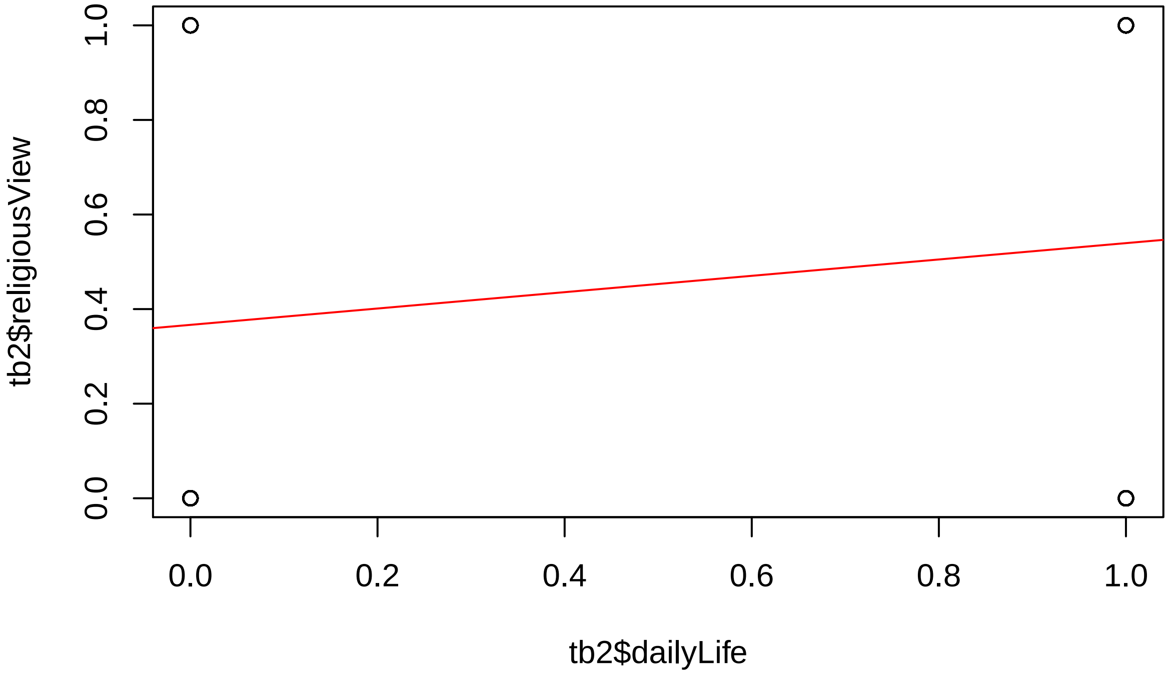
```
m =  lm ( tb2$religiousView ~ tb2$dailyLife)
summary( m )
```

```
##
## Call:
## lm(formula = tb2$religiousView ~ tb2$dailyLife)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.5395 -0.5395 -0.3667  0.4605  0.6333
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.36667    0.05222   7.022 1.34e-11 ***
## tb2$dailyLife  0.17281    0.06167   2.802  0.00539 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4954 on 316 degrees of freedom
## Multiple R-squared:  0.02425,    Adjusted R-squared:  0.02116
## F-statistic: 7.852 on 1 and 316 DF,  p-value: 0.005388
```

```
plot( jitter(tb2$religiousView) ~ jitter(tb2$dailyLife))
abline(m, col="red")
```



```
plot( tb2$religiousView  ~ tb2$dailyLife )
abline(m, col="red")
```



```
table(tb2$religiousView ,  tb2$dailyLife )
```

```
##
##       0   1
##   0  57 105
##   1  33 123
```

Fisher exact test on 2x2 table

```r
table(tb2$religiousView ,  tb2$dailyLife )
```

```
##
##       0   1
##   0  57 105
##   1  33 123
```

```r
RVTable =  as.matrix( table(tb2$religiousView ,  tb2$dailyLife ) )
str(RVTable)
```

```
##  'table' int [1:2, 1:2] 57 33 105 123
##  - attr(*, "dimnames")=List of 2
##   ..$ : chr [1:2] "0" "1"
##   ..$ : chr [1:2] "0" "1"
```

```r
fisher.test(RVTable)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  RVTable
## p-value = 0.006176
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  1.191508 3.461041
## sample estimates:
## odds ratio
##   2.018893
```

Calculate scientific literacy

```r
#sciLiteracy = c("light", "fossil", "food", "electronCharge",
#                 "earlyHuman", "laser", "continents", "antibiotics", "electronSize", "earthCenter")
tb2$light = 0
tb2$light[ tb$light=='TRUE' ] =1
tb2$light[ tb$light=='Wrong' ] =0
table(tb$light)
```

```
##
## I don't know.          TRUE         Wrong
##            23           245            48
```

```r
table(tb2$light)
```

```
##
##   0   1
##  73 245
```

```r
tb2$fossil = 0
tb2$fossil[ tb$fossil=='6 million and 5 years old' ] = 0
tb2$fossil[grep('Still', tb$fossil)] = 1;
table(tb$fossil)
```

```
##
```

```
##        6 million and 5 years old                     I don't know
##                              117                               17
## Still about 6 million years old.
##                              182
```

```
table(tb2$fossil)
```

```
##
##   0   1
## 136 182
```

```
tb2$food = 0
tb2$food[ tb$food=='Dis-agree' ] = 1
tb2$food[grep('Agree', tb$food)] = 0;
table(tb$food)
```

```
##
##        Agree    Dis-agree I don't know
##           49          179           90
```

```
table(tb2$food)
```

```
##
##   0   1
## 139 179
```

```
tb2$electronCharge = 0
tb2$electronCharge[grep('Positive', tb$electronCharge)] = 1;
table(tb$electronCharge)
```

```
##
##    Electricity  Negative charge         Neutron Positive charge
##              9               47              31             230
```

```
table(tb2$electronCharge)
```

```
##
##   0   1
##  88 230
```

```
tb2$earlyHuman = 0
tb2$earlyHuman[grep('TRUE', tb$earlyHuman)] = 0;
tb2$earlyHuman[grep('FALSE', tb$earlyHuman)] = 1;
table(tb$earlyHuman)
```

```
##
##          FALSE I do not know.           TRUE
##            229             37             52
```

```
table(tb2$earlyHuman)
```

```
##
##   0   1
##  89 229
```

```
tb2$earlyHuman = 0
tb2$earlyHuman[grep('TRUE', tb$earlyHuman)] = 0;
tb2$earlyHuman[grep('FALSE', tb$earlyHuman)] = 1;
table(tb$earlyHuman)
```

```
##
##          FALSE I do not know.           TRUE
##            229             37             52
```

```
table(tb2$earlyHuman)
```

```
##
##    0   1
##   89 229
```

```
tb2$laser = 0
tb2$laser[grep('TRUE', tb$laser)] = 0;
tb2$laser[grep('FALSE', tb$laser)] = 1;
table(tb$laser)
```

```
##
##           FALSE I do not know.             TRUE
##             208              69              41
```

```
table(tb2$laser)
```

```
##
##    0   1
##  110 208
```

```
tb2$continents = 0
tb2$continents[grep('TRUE', tb$continents)] = 1;
tb2$continents[grep('FALSE', tb$continents)] = 0;
table(tb$continents)
```

```
##
##           FALSE I do not know.             TRUE
##              11              16             290
```

```
table(tb2$continents)
```

```
##
##    0   1
##   28 290
```

```
tb2$antibiotics = 0
tb2$antibiotics[grep('TRUE', tb$antibiotics)] = 0;
tb2$antibiotics[grep('FALSE', tb$antibiotics)] = 1;
table(tb$antibiotics)
```

```
##
##          FALSE I do not know.           TRUE
##            221             19             78
```

```
table(tb2$antibiotics)
```

```
##
##    0   1
##   97 221
```

```
tb2$electronSize = 0
tb2$electronSize[grep('True', tb$electronSize)] = 1;
tb2$electronSize[grep('FALSE', tb$electronSize)] = 0;
table(tb$electronSize)
```

```
##
```

```
##           FALSE I do no know.           True
##             61             22            234
```

```
table(tb2$electronSize)
```

```
##
##    0   1
##   84 234
```

```
tb2$earthCenter = 0
tb2$earthCenter[grep('TRUE', tb$earthCenter)] = 1;
tb2$earthCenter[grep('FALSE', tb$earthCenter)] = 0;
table(tb$earthCenter)
```

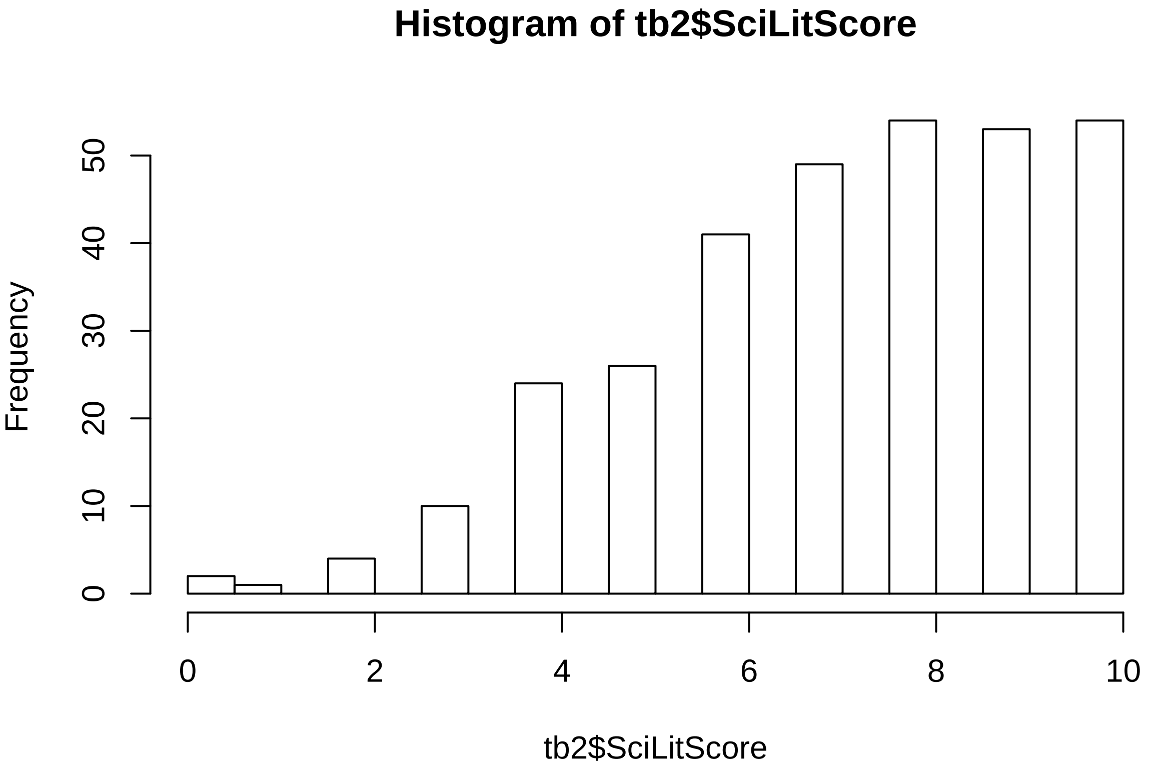
```
##
##           FALSE I do not know.           TRUE
##             14             18            285
```

```
table(tb2$earthCenter)
```

```
##
##    0   1
##   33 285
```

```
#sciLiteracy = c("light", "fossil", "food", "electronCharge",
#                "earlyHuman", "laser", "continents", "antibiotics", "electronSize", "earthCenter")

tb2$SciLitScore = apply( tb2[, sciLiteracy], MARGIN=1, FUN=sum ) #by row
hist(tb2$SciLitScore, br=20)
```



```
str(tb2)
```

```
## 'data.frame':    318 obs. of  26 variables:
```

```
##  $ gender        : chr  "Do not wish to answer" "Male" "Female" "Do not wish to answer" ...
##  $ degree        : chr  "Bachelor Degree in Science or equivalent" "High School or equivalent" "High
##  $ country       : num  1 1 1 1 1 1 1 1 1 0 ...
##  $ age           : num  20 20 35.5 NA 53 58 20 45.5 35.5 35.5 ...
##  $ shaq          : num  1 0 0 1 0 0 0 0 1 0 ...
##  $ kilo          : num  1 1 0 1 0 1 0 1 1 1 ...
##  $ mm            : num  1 1 0 1 1 1 0 1 1 0 ...
##  $ inseam        : num  0 1 1 1 0 1 0 1 1 1 ...
##  $ weather       : num  0 1 0 0 0 1 0 1 1 1 ...
##  $ metric        : num  3 4 1 4 1 4 0 4 5 3 ...
##  $ religiousView : num  0 0 0 1 1 1 1 1 1 1 ...
##  $ dailyLife     : num  1 1 0 1 1 1 0 1 1 1 ...
##  $ SciOnLife     : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ SciEffect     : num  1 1 0 0 0 1 0 0 1 1 ...
##  $ SciAttitude   : num  3 3 1 3 3 4 2 3 4 4 ...
##  $ light         : num  1 1 1 0 0 0 1 0 1 1 ...
##  $ fossil        : num  0 0 0 1 0 1 0 1 1 0 ...
##  $ food          : num  0 1 1 1 1 1 0 1 1 1 ...
##  $ electronCharge: num  0 1 1 1 0 1 1 1 1 1 ...
##  $ earlyHuman    : num  1 1 0 1 1 1 1 1 1 1 ...
##  $ laser         : num  0 1 1 1 1 1 0 1 0 1 ...
##  $ continents    : num  1 1 1 1 0 1 1 1 1 1 ...
##  $ antibiotics   : num  1 1 1 1 0 1 1 1 1 1 ...
##  $ electronSize  : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ earthCenter   : num  1 1 1 1 0 1 1 1 1 1 ...
##  $ SciLitScore   : num  6 9 8 9 4 9 7 9 9 9 ...
```

Output the 'cleaned' data to a csv file

```
tb3 = tb2[, c("gender", "age", "country", "degree", "metric", "SciAttitude", "SciLitScore")]
head(tb3)
```

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

```
write.csv(tb3, file = "metric-attitude-literacy.csv", row.names = FALSE, quote=TRUE)
```