George_Smith_IST719_Lab2

George Smith

10/17/2021

Author George Smith

Purpose: Lab 2, data interogration or data exploration

Data exploration Lab 2

this code only works for interactive files as this file is on my PC it is not relevant

install package

```
#install.packages("vioplot")
#install.packages("plotrix")
```

```
#fname <- file.choose()
```

supplemented the above code with the following line

```
tips <- read.csv("C:/Users/GeorgeSmith/Documents/Syracuse/IST 719/tips.csv"
                 , header = TRUE
                #, StringASFactors = FALSE
                )
```

view the name of columns

```
colnames(tips)
```

```
## [1] "X"          "total_bill" "tip"        "sex"        "smoker"
## [6] "day"        "time"       "size"
```

lets you edit data in spreadsheet view

```
fix(tips)
```

lets you view data cant edit

```
#view(tips)
```

structure of the data

```
str(tips)
```

```
## 'data.frame':    244 obs. of  8 variables:
##  $ X         : num  1 2 3 4 5 6 7 8 9 10 ...
##  $ total_bill: num  17 10.3 21 23.7 24.6 ...
##  $ tip       : num  1.01 1.66 3.5 3.31 3.61 4.71 2 3.12 1.96 3.23 ...
##  $ sex       : chr  "Female" "Male" "Male" "Male" ...
##  $ smoker    : chr  "No" "No" "No" "No" ...
##  $ day       : chr  "Sun" "Sun" "Sun" "Sun" ...
##  $ time      : chr  "Dinner" "Dinner" "Dinner" "Dinner" ...
##  $ size      : num  2 3 3 2 4 4 2 4 2 2 ...
```

call rows and columns

```
tips[1,]
```

```
##   X total_bill  tip    sex smoker day   time size
## 1 1      16.99 1.01 Female     No Sun Dinner    2
```

```
tips[,1]
```

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

```
## [199] 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216
## [217] 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234
## [235] 235 236 237 238 239 240 241 242 243 244
```

```
tips[3,3]
```

```
## [1] 3.5
```

```
tips[1:3,]
```

```
##   X total_bill  tip    sex smoker day   time size
## 1 1      16.99 1.01 Female     No Sun Dinner    2
## 2 2      10.34 1.66   Male     No Sun Dinner    3
## 3 3      21.01 3.50   Male     No Sun Dinner    3
```

#tells us how many variables are in this subset

```
length(tips [1:3, 2])
```

```
## [1] 3
```

number of dimensions in a data set

```
dim(tips)
```

```
## [1] 244   8
```

subsets the dim function

```
dim(tips)[1]
```

```
## [1] 244
```

#call a column

```
tips$time
```

```
##   [1] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
##   [9] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
##  [17] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
##  [25] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
##  [33] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
##  [41] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
##  [49] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
```

```
##  [57] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
##  [65] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
##  [73] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Lunch"  "Lunch"  "Lunch"
##  [81] "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"
##  [89] "Lunch"  "Lunch"  "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
##  [97] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
## [105] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
## [113] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Lunch"  "Lunch"  "Lunch"
## [121] "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"
## [129] "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"
## [137] "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"
## [145] "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Dinner" "Dinner"
## [153] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
## [161] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
## [169] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
## [177] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
## [185] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Lunch"
## [193] "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"
## [201] "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Dinner" "Dinner"
## [209] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
## [217] "Dinner" "Dinner" "Dinner" "Dinner" "Lunch"  "Lunch"  "Lunch"  "Lunch"
## [225] "Lunch"  "Lunch"  "Lunch"  "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
## [233] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
## [241] "Dinner" "Dinner" "Dinner" "Dinner"
```

differnt way to call the columns

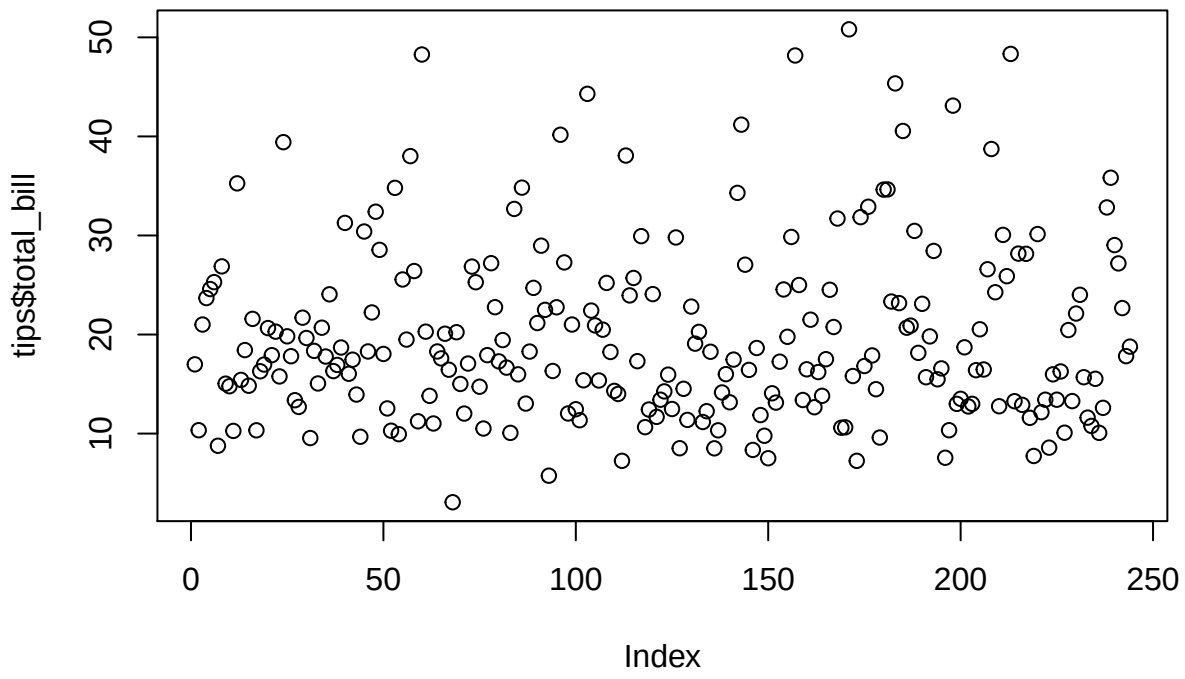
```
tips[,"time"]
```

```
##   [1] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
##   [9] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
##  [17] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
##  [25] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
##  [33] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
##  [41] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
##  [49] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
##  [57] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
##  [65] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
##  [73] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Lunch"  "Lunch"  "Lunch"
##  [81] "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"
##  [89] "Lunch"  "Lunch"  "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
##  [97] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
## [105] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
## [113] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Lunch"  "Lunch"  "Lunch"
## [121] "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"
## [129] "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"
## [137] "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"
## [145] "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Dinner" "Dinner"
## [153] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
## [161] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
## [169] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
```

```
## [177] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
## [185] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Lunch"
## [193] "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"
## [201] "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Lunch"  "Dinner" "Dinner"
## [209] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
## [217] "Dinner" "Dinner" "Dinner" "Dinner" "Lunch"  "Lunch"  "Lunch"  "Lunch"
## [225] "Lunch"  "Lunch"  "Lunch"  "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
## [233] "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner" "Dinner"
## [241] "Dinner" "Dinner" "Dinner" "Dinner"
```

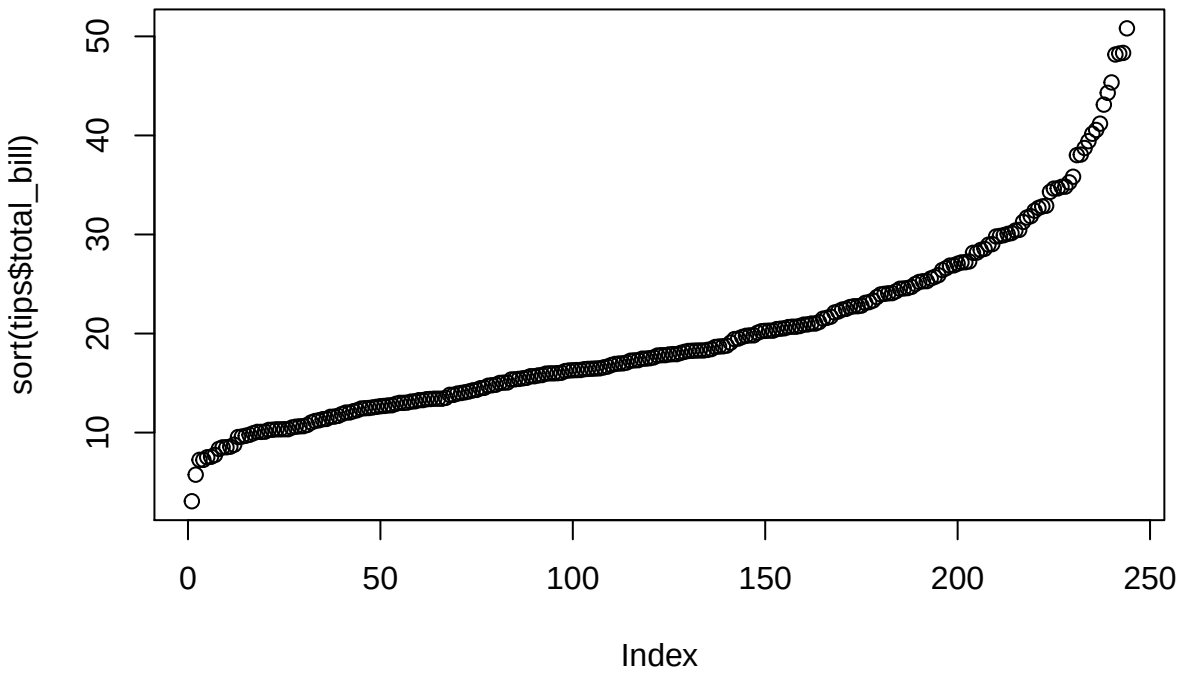
plot the total bull column

```
plot(tips$total_bill)
```



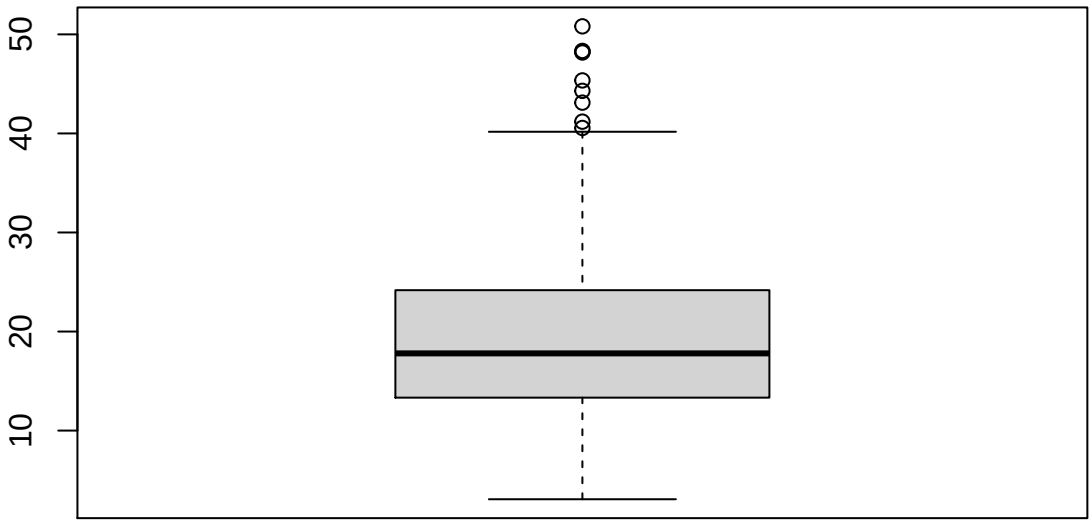
sorts the data

```
plot(sort(tips$total_bill))
```



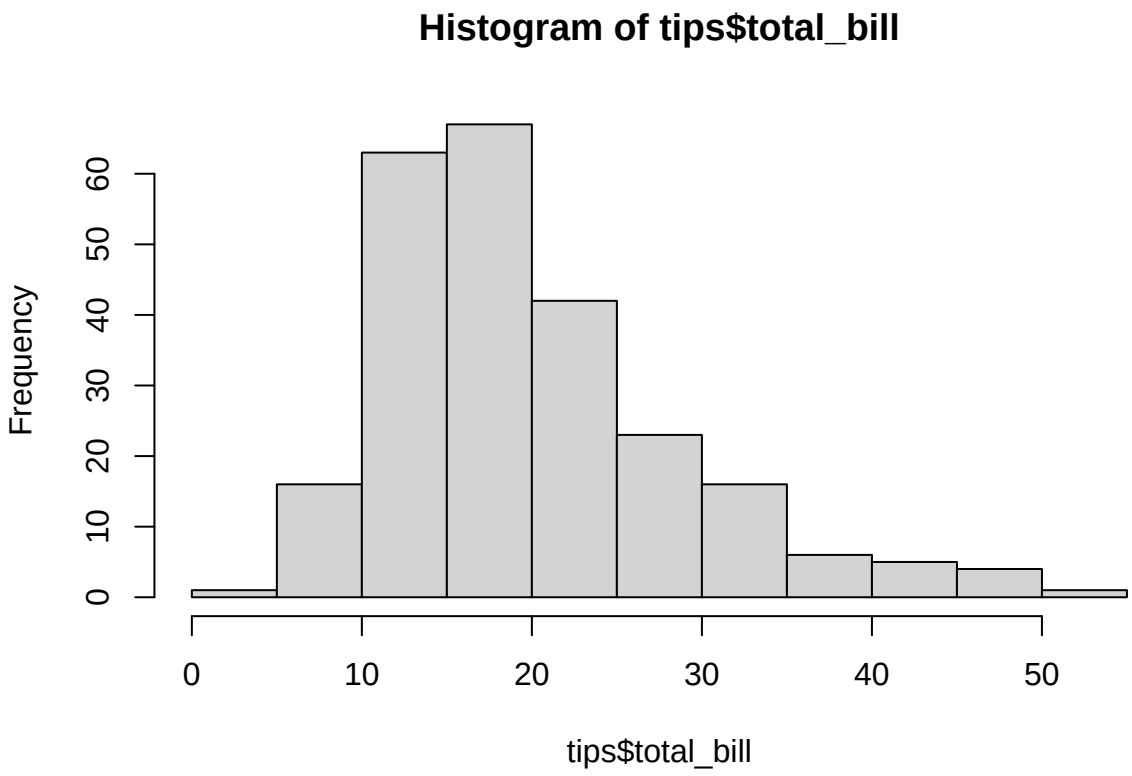
boxplot

```
boxplot(tips$total_bill)
```



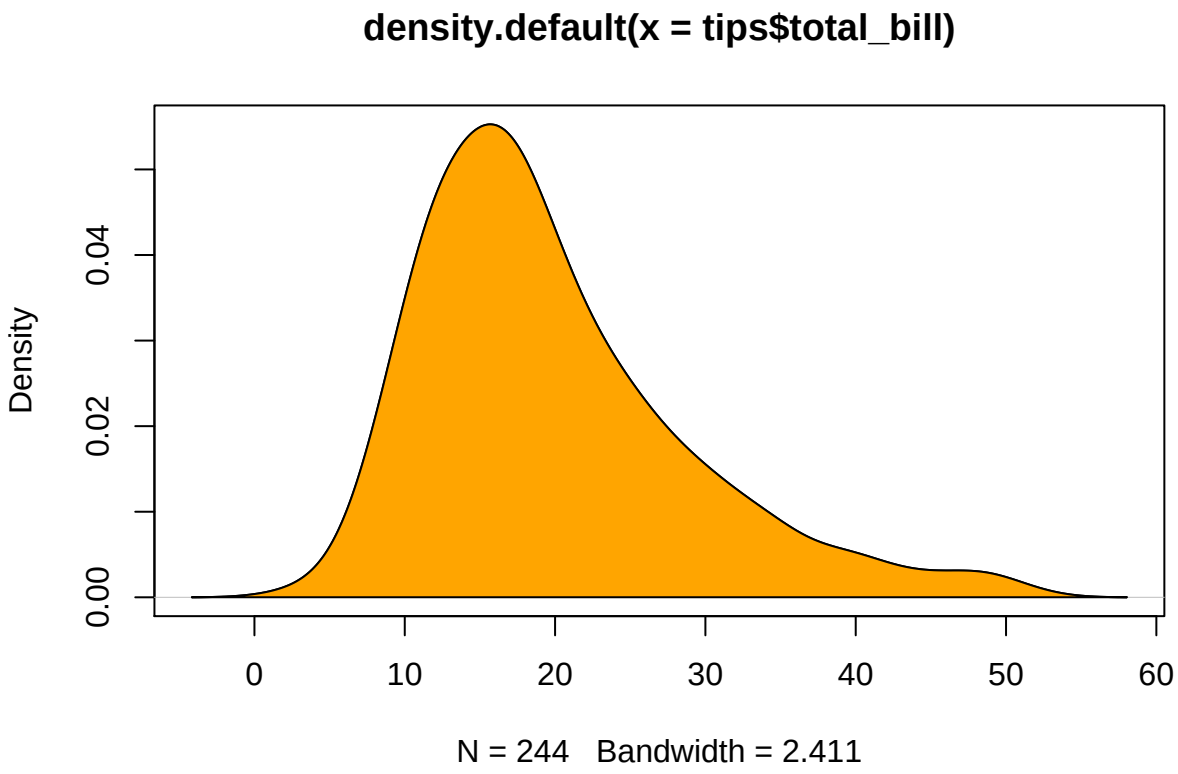
histogram

```
hist(tips$total_bill)
```



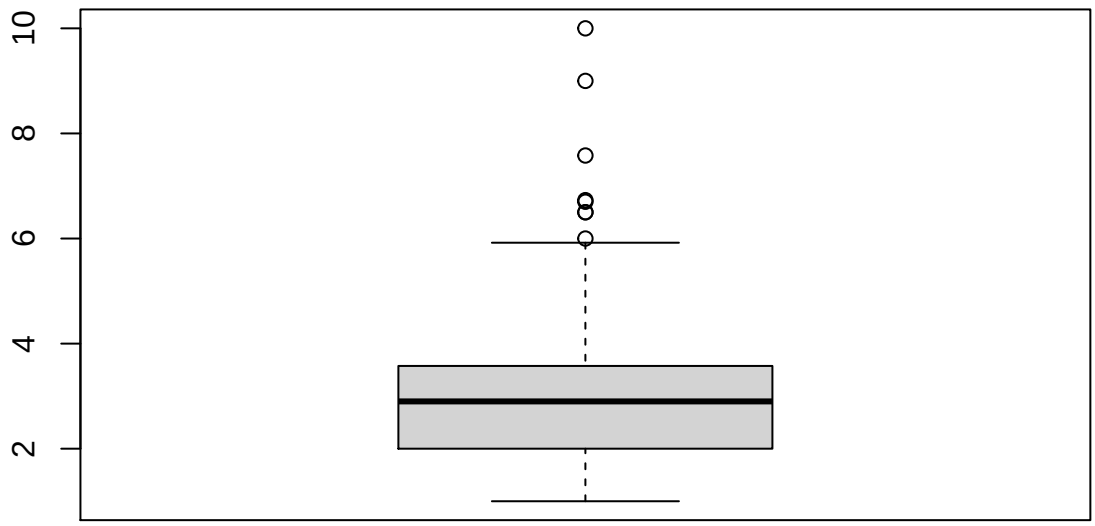
density plot

```
d <- density(tips$total_bill)
plot(d)
polygon(d, col = "orange")
```



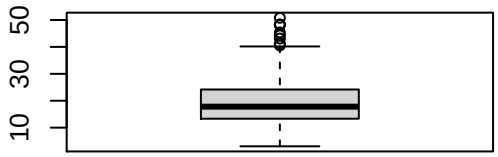
boxplot of tips for asynch

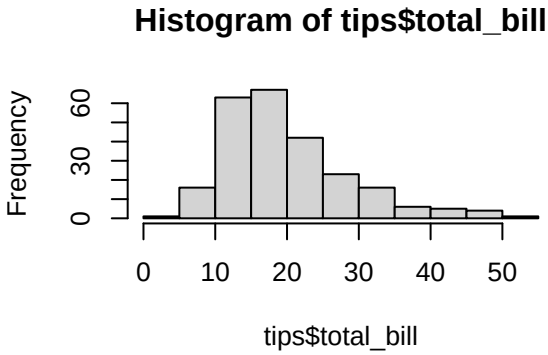
```
boxplot(tips$tip)
```

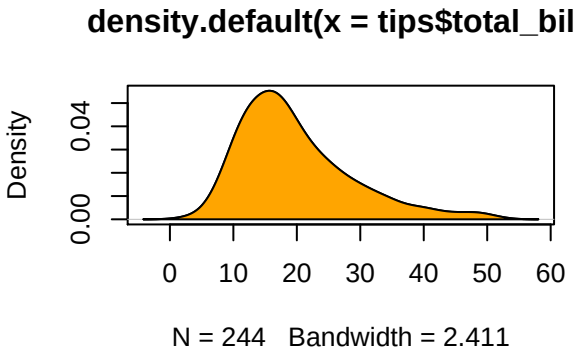


have multiple chars on the same plot

```
par(mfrow = c(2,2))
boxplot(tips$total_bill)
hist(tips$total_bill)
d <- density(tips$total_bill)
plot(d)
polygon(d, col = "orange")
```







add a vioplot

```
library(vioplot)
```

```
## Loading required package: sm
```

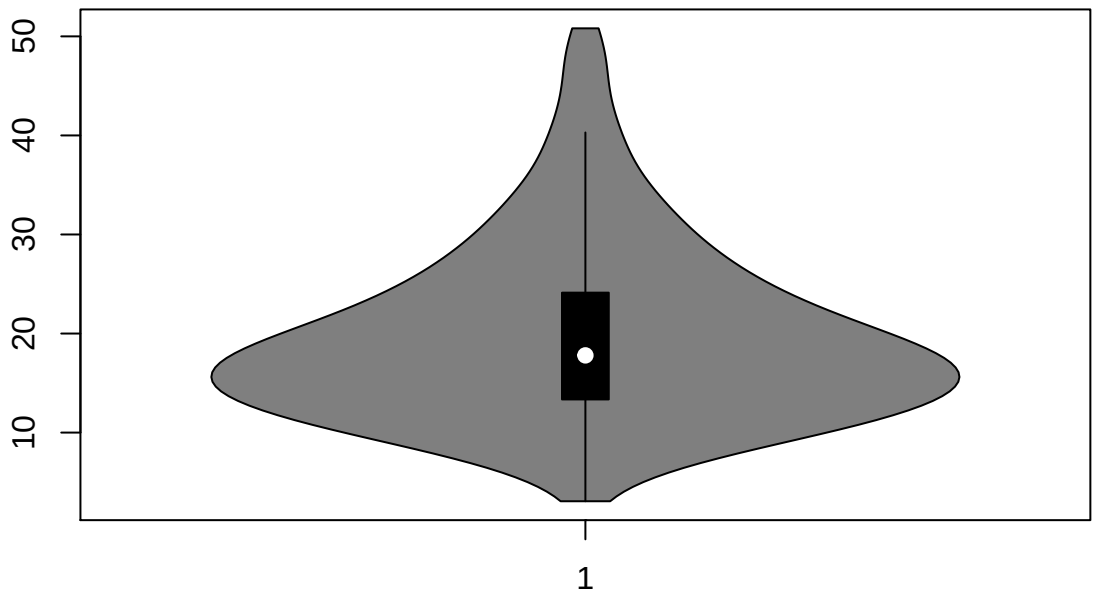
```
## Package 'sm', version 2.2-5.7: type help(sm) for summary information
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
vioplot(tips$total_bill)
```




```
unique(tips$sex)
```

```
## [1] "Female" "Male"
```

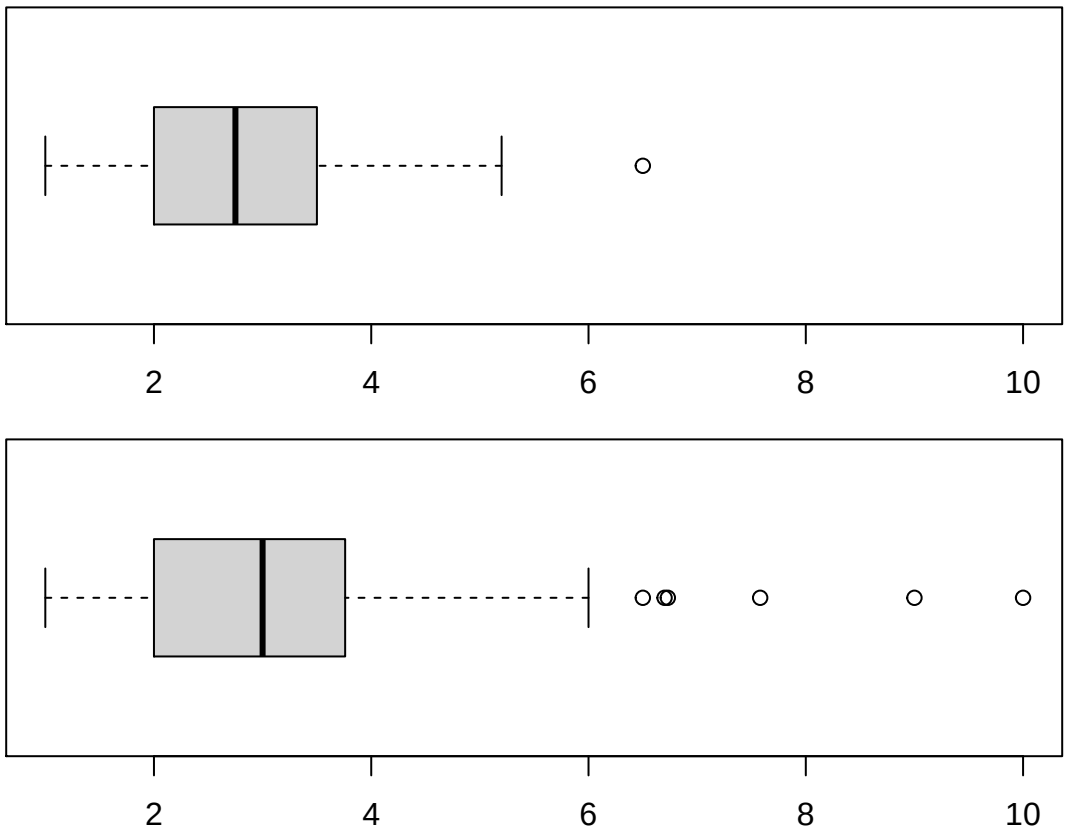
#subet for just males

```
tips.M <- tips[tips$sex == "Male", ]
```

subet only females

```
tips.F <- tips[tips$sex == "Female", ]
```

```
par(mfrow = c(2,1), mar = c(2,3,1,2))
boxplot(tips.F$tip, horizontal = T, ylim = c(1,10))
boxplot(tips.M$tip, horizontal = T, ylim = c(1,10))
```



working with JSON files

```
fname <- "C:/Users/GeorgeSmith/Documents/Syracuse/IST 719/tweet.formated.json"
```

import library

```
library(jsonlite)
```

read in data

```
raw.tweet<- fromJSON(fname, flatten = FALSE)
```

analyze data

```
str(raw.tweet)
```

```
## List of 21
##  $ text                     : chr "We Are All Human Microphones http://t.co/Ge2JYANI #OWS"
##  $ created_at               : chr "Wed Oct 19 00:42:11 +0000 2011"
##  $ retweeted                : logi FALSE
##  $ in_reply_to_status_id_str: NULL
##  $ entities                 :List of 3
##   ..$ hashtags     :'data.frame':    1 obs. of  2 variables:
##   .. ..$ text   : chr "OWS"
##   .. ..$ indices:List of 1
##   .. .. ..$ : int [1:2] 50 54
##   ..$ user_mentions: list()
##   ..$ urls         :'data.frame':    1 obs. of  4 variables:
##   .. ..$ indices     :List of 1
##   .. .. ..$ : int [1:2] 29 49
##   .. ..$ url         : chr "http://t.co/Ge2JYANI"
##   .. ..$ expanded_url: chr "http://n.pr/nMb97t"
##   .. ..$ display_url : chr "n.pr/nMb97t"
##  $ geo                      : NULL
##  $ place                    : NULL
##  $ possibly_sensitive       : logi FALSE
##  $ in_reply_to_user_id_str  : NULL
##  $ id_str                   : chr "126458082875281409"
##  $ source                   : chr "<a href=\"http://twitter.com/tweetbutton\" rel=\"nofollow\">Tweet
##  $ contributors             : NULL
##  $ coordinates              : NULL
##  $ in_reply_to_status_id    : NULL
##  $ retweet_count            : int 0
##  $ in_reply_to_user_id      : NULL
##  $ favorited                : logi FALSE
##  $ truncated                : logi FALSE
##  $ user                     :List of 38
##   ..$ default_profile                  : logi FALSE
##   ..$ created_at                       : chr "Sun Nov 28 04:30:39 +0000 2010"
##   ..$ geo_enabled                      : logi TRUE
##   ..$ profile_use_background_image     : logi TRUE
##   ..$ follow_request_sent              : NULL
##   ..$ lang                             : chr "en"
##   ..$ profile_background_image_url_https: chr "https://si0.twimg.com/profile_background_images/341549
##   ..$ profile_text_color               : chr "666666"
##   ..$ followers_count                  : int 498
##   ..$ profile_background_image_url     : chr "http://a1.twimg.com/profile_background_images/3415490
##   ..$ url                              : chr "http://www.scribd.com/dan_schell"
##   ..$ description                      : chr "Poetry/Fiction Writer in a Brave New World. Namaste!\
##   ..$ screen_name                      : chr "WordEngineer"
##   ..$ id_str                           : chr "220563286"
##   ..$ profile_link_color               : chr "2FC2EF"
##   ..$ is_translator                    : logi FALSE
##   ..$ following                        : NULL
##   ..$ favourites_count                 : int 1
##   ..$ listed_count                     : int 11
##   ..$ friends_count                    : int 1028
```

```
##   ..$ profile_background_color          : chr "1A1B1F"
##   ..$ location                          : chr "Saginaw, Michigan"
##   ..$ notifications                     : NULL
##   ..$ profile_background_tile           : logi TRUE
##   ..$ protected                         : logi FALSE
##   ..$ profile_image_url_https           : chr "https://si0.twimg.com/profile_images/1387603076/22779
##   ..$ show_all_inline_media             : logi FALSE
##   ..$ contributors_enabled              : logi FALSE
##   ..$ statuses_count                    : int 6089
##   ..$ verified                          : logi FALSE
##   ..$ profile_sidebar_fill_color        : chr "252429"
##   ..$ name                              : chr "Dan Schell"
##   ..$ profile_image_url                 : chr "http://a0.twimg.com/profile_images/1387603076/227792_
##   ..$ id                                : int 220563286
##   ..$ default_profile_image             : logi FALSE
##   ..$ time_zone                         : chr "Eastern Time (US & Canada)"
##   ..$ utc_offset                        : int -18000
##   ..$ profile_sidebar_border_color      : chr "181A1E"
##  $ id                        : num 1.26e+17
##  $ in_reply_to_screen_name   : NULL
```

```
names(raw.tweet)
```

```
##  [1] "text"                     "created_at"
##  [3] "retweeted"                "in_reply_to_status_id_str"
##  [5] "entities"                 "geo"
##  [7] "place"                    "possibly_sensitive"
##  [9] "in_reply_to_user_id_str"  "id_str"
## [11] "source"                   "contributors"
## [13] "coordinates"              "in_reply_to_status_id"
## [15] "retweet_count"            "in_reply_to_user_id"
## [17] "favorited"                "truncated"
## [19] "user"                     "id"
## [21] "in_reply_to_screen_name"
```

```
raw.tweet$text
```

```
## [1] "We Are All Human Microphones http://t.co/Ge2JYANI #OWS"
```

```
raw.tweet$user$followers_count
```

```
## [1] 498
```

```
raw.tweet[["user"]]
```

```
## $default_profile
## [1] FALSE
##
## $created_at
## [1] "Sun Nov 28 04:30:39 +0000 2010"
##
```

```
## $geo_enabled
## [1] TRUE
##
## $profile_use_background_image
## [1] TRUE
##
## $follow_request_sent
## NULL
##
## $lang
## [1] "en"
##
## $profile_background_image_url_https
## [1] "https://si0.twimg.com/profile_background_images/341549012/twilk_background_4e8c772cce037.jpg"
##
## $profile_text_color
## [1] "666666"
##
## $followers_count
## [1] 498
##
## $profile_background_image_url
## [1] "http://a1.twimg.com/profile_background_images/341549012/twilk_background_4e8c772cce037.jpg"
##
## $url
## [1] "http://www.scribd.com/dan_schell"
##
## $description
## [1] "Poetry/Fiction Writer in a Brave New World. Namaste!\r\n(Warning: I tweet/retweet a lot)."
##
## $screen_name
## [1] "WordEngineer"
##
## $id_str
## [1] "220563286"
##
## $profile_link_color
## [1] "2FC2EF"
##
## $is_translator
## [1] FALSE
##
## $following
## NULL
##
## $favourites_count
## [1] 1
##
## $listed_count
## [1] 11
##
## $friends_count
## [1] 1028
##
```

```
## $profile_background_color
## [1] "1A1B1F"
##
## $location
## [1] "Saginaw, Michigan"
##
## $notifications
## NULL
##
## $profile_background_tile
## [1] TRUE
##
## $protected
## [1] FALSE
##
## $profile_image_url_https
## [1] "https://si0.twimg.com/profile_images/1387603076/227792_671181140970_51903108_35123681_4713128_n
##
## $show_all_inline_media
## [1] FALSE
##
## $contributors_enabled
## [1] FALSE
##
## $statuses_count
## [1] 6089
##
## $verified
## [1] FALSE
##
## $profile_sidebar_fill_color
## [1] "252429"
##
## $name
## [1] "Dan Schell"
##
## $profile_image_url
## [1] "http://a0.twimg.com/profile_images/1387603076/227792_671181140970_51903108_35123681_4713128_n_n
##
## $id
## [1] 220563286
##
## $default_profile_image
## [1] FALSE
##
## $time_zone
## [1] "Eastern Time (US & Canada)"
##
## $utc_offset
## [1] -18000
##
## $profile_sidebar_border_color
## [1] "181A1E"
```

```
raw.tweet[["user"]]$followers_count
```

```
## [1] 498
```

```
raw.tweet[["user"]][["follower_count"]]
```

```
## NULL
```

read in new file

based on error message appears to be issues with the underlying file

```
fname3 <- "C:/Users/GeorgeSmith/Documents/Syracuse/IST 719/tweets5814.json"
#con <- file(fname3, open ="r")
#tweets <- stream_in(con)
#close(con)
```

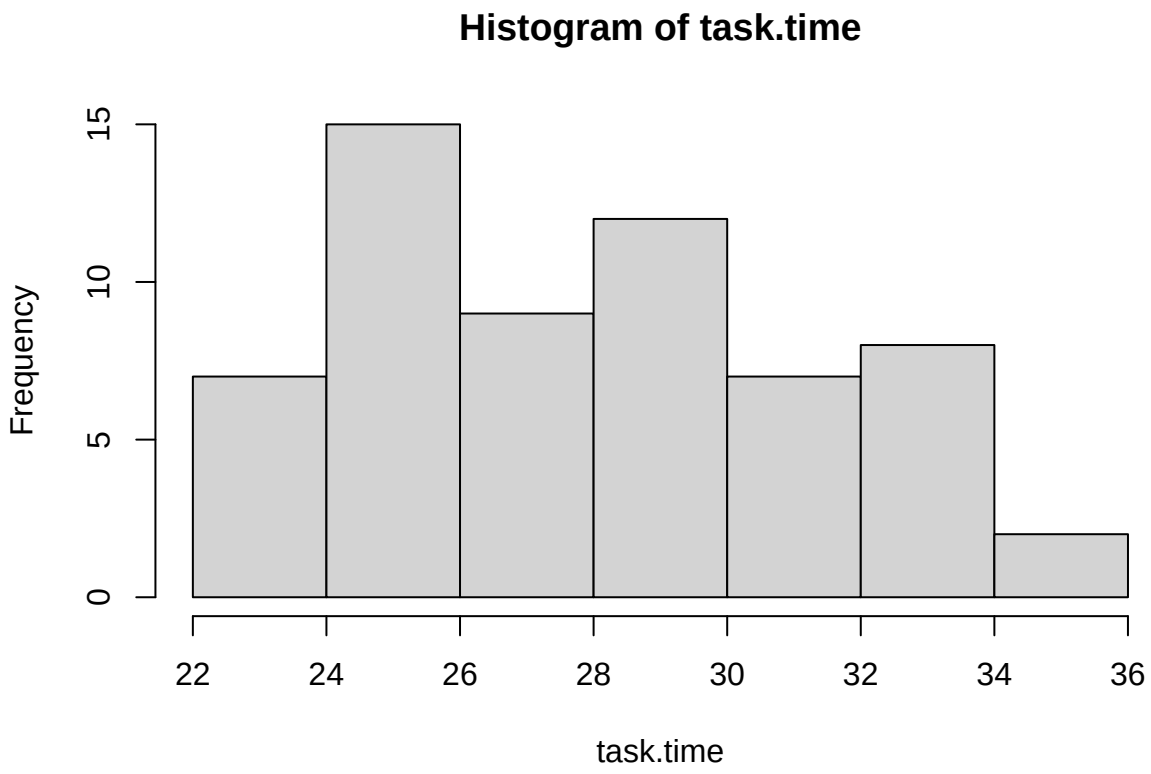
```
#dim(tweets)
```

```
#tweets$text [1:3]
```

```
#boxplot(log10(tweets$user$followers_count), horizontal = TRUE )
```

distribution work

```
task.time <- c(rnorm(n=30, mean=30, sd =2.25)
               ,rnorm (n= 30, mean =25, sd = 1.5))
hist(task.time)
```




```
status <- c(rep("AMA", 30), rep("PRO", 30))
df <- data.frame(time = task.time, status= status)
```

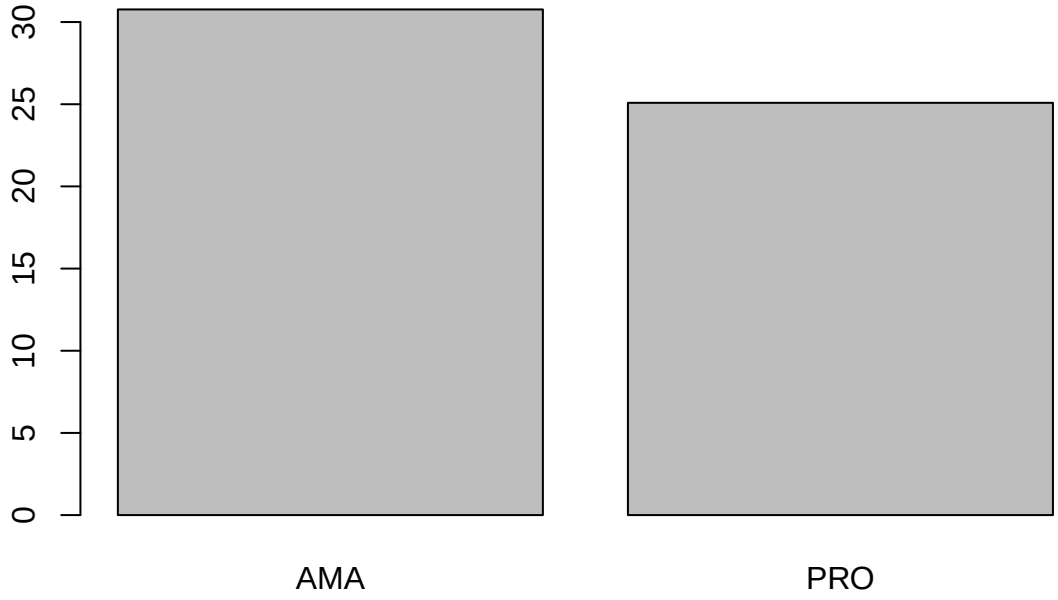
```
df.grouped <- aggregate(df$time, list(df$status), mean)
```

```
colnames(df.grouped) <- c("stat", "time")
```

```
df.grouped
```

```
##   stat     time
## 1  AMA 30.76752
## 2  PRO 25.08464
```

```
barplot(df.grouped$time, names.arg = df.grouped$stat)
```




```
M.grouped <- tapply(df$time, list(df$status), mean)
class(M.grouped)
```

```
## [1] "array"
```

returns a matrix or vector same as aggregate

```
tapply(df$time, list(df$status), range)
```

```
## $AMA
## [1] 26.66334 35.29661
##
## $PRO
## [1] 22.89067 28.05227
```

```
summary(task.time)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   22.89   25.29   27.38   27.93   30.61   35.30
```

```r
range(task.time)
```

```
## [1] 22.89067 35.29661
```

```r
aggregate(df$time, list(df$status), summary)
```

```
##   Group.1   x.Min. x.1st Qu. x.Median   x.Mean x.3rd Qu.   x.Max.
## 1     AMA 26.66334  29.11588 30.67346 30.76752  32.45261 35.29661
## 2     PRO 22.89067  24.19084 25.25836 25.08464  26.00617 28.05227
```

```r
table(df$status)
```

```
##
## AMA PRO
##  30  30
```

```r
table(round(df$time,2))
```

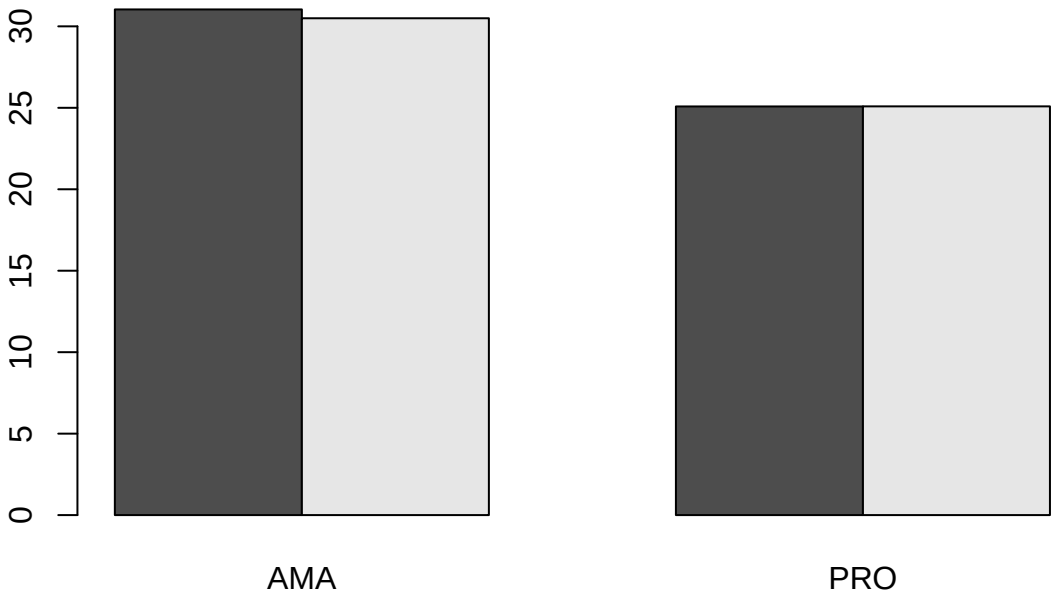
```
##
## 22.89 22.98 23.07 23.13 23.16 23.29  23.8 24.18 24.23 24.27  24.3 24.61  25.1
##     1     1     1     1     1     1     1     1     1     1     1     1     1
## 25.18  25.2 25.32 25.38  25.4 25.67 25.79 25.89 26.05 26.24 26.28 26.66 26.76
##     1     1     2     1     1     1     1     1     1     1     1     1     1
## 26.97    27 27.05 27.71 28.02 28.05 28.07 28.15 28.53 28.89 29.08 29.21 29.46
##     1     1     1     1     1     1     1     1     1     1     1     1     1
## 29.71 29.89 29.96 30.43 30.54  30.8 31.02 31.14 31.42 31.75 32.26 32.31  32.5
##     1     1     1     1     1     1     1     1     1     1     1     1     1
## 32.52 32.55 33.29 33.69    34 34.18  35.3
##     1     1     1     1     1     1     1
```

```r
df$sex <- sample(c("M","F"), 60, replace = T)
```

```r
aggregate(df$time, list(df$status, df$sex), mean)
```

```
##   Group.1 Group.2        x
## 1     AMA       F 31.03835
## 2     PRO       F 25.08262
## 3     AMA       M 30.49669
## 4     PRO       M 25.08812
```

```r
M <- tapply(df$time, list(df$sex, df$status), mean)
M <- tapply(df$time, list(df$sex, df$status), mean)
barplot(M, beside = TRUE)
```



reshaping data with tidyr

```
library(tidyr)
```

```
n <- 5
year <- 2001:(2000 + n)
q1 <- runif(n = n, min = 100, max = 120)
q2 <- runif(n=n, min = 103, max = 130)
q3 <- runif(n=n, min = 104, max = 140)
q4 <- runif(n=n, min = 108, max = 150)

df.wide<- data.frame(year, q1,q2,q3,q4)
gather(df.wide, qt, sales, q1:q4)
```

```
##    year qt    sales
## 1  2001 q1 104.6386
## 2  2002 q1 112.0746
## 3  2003 q1 110.8528
## 4  2004 q1 102.4208
## 5  2005 q1 100.6433
## 6  2001 q2 114.2042
## 7  2002 q2 111.5108
```

```
## 8  2003 q2 108.7388
## 9  2004 q2 118.9541
## 10 2005 q2 120.7866
## 11 2001 q3 105.5164
## 12 2002 q3 111.1778
## 13 2003 q3 132.1087
## 14 2004 q3 129.6848
## 15 2005 q3 117.6341
## 16 2001 q4 129.2879
## 17 2002 q4 125.3236
## 18 2003 q4 149.2614
## 19 2004 q4 131.4720
## 20 2005 q4 110.7867
```

```
df.wide %>% gather(qt, sales, q1:q4)
```

```
##    year qt    sales
## 1  2001 q1 104.6386
## 2  2002 q1 112.0746
## 3  2003 q1 110.8528
## 4  2004 q1 102.4208
## 5  2005 q1 100.6433
## 6  2001 q2 114.2042
## 7  2002 q2 111.5108
## 8  2003 q2 108.7388
## 9  2004 q2 118.9541
## 10 2005 q2 120.7866
## 11 2001 q3 105.5164
## 12 2002 q3 111.1778
## 13 2003 q3 132.1087
## 14 2004 q3 129.6848
## 15 2005 q3 117.6341
## 16 2001 q4 129.2879
## 17 2002 q4 125.3236
## 18 2003 q4 149.2614
## 19 2004 q4 131.4720
## 20 2005 q4 110.7867
```

```
df.long <- df.wide %>% gather(qt, sales, q1:q4)
o <- order(df.long$year, df.long$qt)
df.long <- df.long[o,]
```

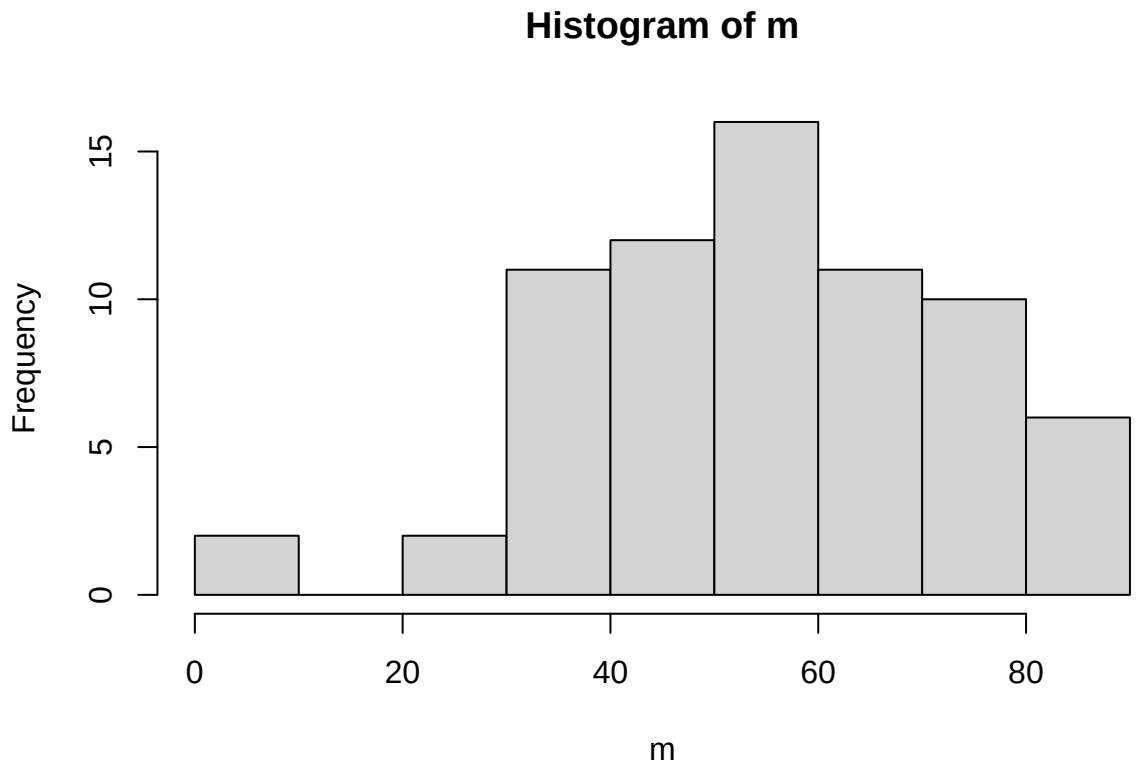
```
df <- data.frame (cat=rep(c("tap", "reg", "zed", "vum"),3)
                  , group = rep(letters[7:9], 4)
                  , x= 1:12)
```

```
spread(df,cat,x)
```

```
##   group reg tap vum zed
## 1     g  10   1   4   7
## 2     h   2   5   8  11
## 3     i   6   9  12   3
```

using rect function to build a custom plot

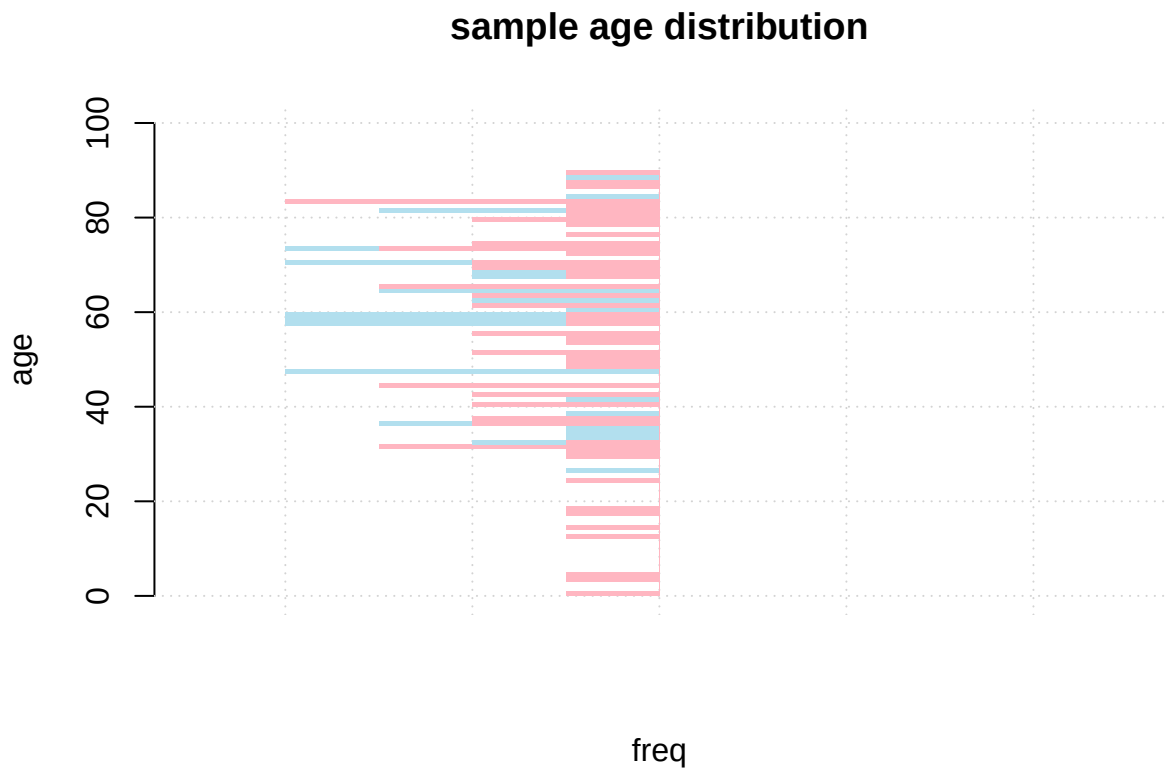
```
library(plotrix)
n <- 70
age.min <- 1
age.max <- 90
age.range <- c(age.min, age.max)
m <- round(rescale(rbeta(n,5,2.5), age.range), 0)
hist(m)
```




```
f <- round(rescale(rbeta(n,5,2.0), age.range), 0)
x <- age.min : age.max
f.y <- m.y <- rep (0, length(x))
```

```
m.tab <- table(m)
m.y[as.numeric((names(m.tab)))] <- as.numeric(m.tab)

f.tab <- table(f)
f.y[as.numeric((names(f.tab)))] <- as.numeric (f.tab)

age.freqs <- data.frame (ages = x, males =m.y, females =f.y)

max.x <- round(1.2 * max(age.freqs[,2:3]), 0)
plot(c(-max.x, max.x), c(0,100), type = "n", bty = "n", xaxt = "n"
```

```
     , ylab = "age", xlab = "freq", main = "sample age distribution")

grid()
last.y <- 0
for (i in 1:90) {
    rect(xleft = 0, ybottom = last.y, xright = -age.freqs$males[i]
       , ytop =age.freqs$ages[i], col = "lightblue2", border = NA)

    rect(xleft = 0, ybottom = last.y, xright = -age.freqs$females[i]
       , ytop =age.freqs$ages[i], col = "lightpink", border = NA)

    last.y <- age.freqs$ages[i]
}
```



‘‘‘