Crash Course in R

Matthew E. Aiello-Lammens

Use a script file

Let's all make a new directory to work in today.

Let's create a new script file in that directory.

Scripts make it easier to repeat your work. You can also add comments using the pound sign.

Short cut to execute commands and functions:

Mac: [Command] + return

Windows [Control] + return (RStudio) [Control] + r (R gui)

Setting your working directory

```
setwd("[your dir name here]")
```

R as a calculator

```
5 + 3
```

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

```
8^3
```

```
## [1] 512
```

```
6*3-1
```

```
## [1] 17
```

```
# Order of operations follows PEMDAS
6*(3-1)
```

```
## [1] 12
```

Variables

We can save things within our session as **variables**

```
pop_1 <- 1200
pop_2 <- 500

pop_total <- pop_1 + pop_2

pop_1 * 2
```

```
## [1] 2400
```

```
# Etc.
```

Challange

I chagne pop_1

```
pop_1 <- 2000
```

What is `pop_total` now?

Loading Data

Best to use the full path to your data, but could also change into the directory you're data is in, then call it in there.

```
bottlenose <- read.csv(file = "~/Dropbox/Projects/SCCC-Stats-Workshop/Symons_data.csv")
```

Let's have a look at these data

```
head(bottlenose)
```

```
##   order Idsample       Date    boat divetime age  sex logdivetime
## 1   408       25  11/4/2000 without       29 old male   3.3672958
## 2   409       25  11/4/2000 without       68 old male   4.2195077
## 3   410       25  11/4/2000 without       32 old male   3.4657359
## 4   411       25  11/4/2000 without       80 old male   4.3820266
## 5   412       25  11/4/2000 without        8 old male   2.0794415
## 6   761       35 14/04/2000 without        2 old male   0.6931472
##   simpleboat divetype bouts individual
## 1          0  surface     9         37
## 2          0   bottom     9         37
## 3          0  surface    10         37
## 4          0   bottom    10         37
## 5          0  surface    11         37
## 6          0  surface    12         11
```

```
tail(bottlenose)
```

```
##      order Idsample       Date    boat divetime age    sex logdivetime
## 1641 17329       941 16/02/2002 without        5 old female    1.609438
## 1642 17330       941 16/02/2002 without        5 old female    1.609438
## 1643 17331       941 16/02/2002 without        7 old female    1.945910
## 1644 17332       941 16/02/2002 without        8 old female    2.079442
## 1645 17333       941 16/02/2002 without        5 old female    1.609438
## 1646 17334       941 16/02/2002 without        4 old female    1.386294
##      simpleboat divetype bouts individual
## 1641          0  surface   862          3
## 1642          0  surface   862          3
```

```
## 1643          0  surface   862          3
## 1644          0  surface   862          3
## 1645          0  surface   862          3
## 1646          0  surface     1          3
```

```
summary(bottlenose)
```

```
##      order          Idsample            Date
##  Min.   :  408   Min.   : 25.0   22/04/2001: 168
##  1st Qu.: 5120   1st Qu.:257.5   9/7/2001  : 105
##  Median :11582   Median :633.0   9/4/2001  :  99
##  Mean   : 9821   Mean   :527.1   5/7/2001  :  87
##  3rd Qu.:12796   3rd Qu.:697.0   4/4/2001  :  85
##  Max.   :17334   Max.   :941.0   2/6/2000  :  82
##                                  (Other)   :1020
##                       boat         divetime           age
##  without                :1240   Min.   :  1.00   old    :1338
##  CP                     :  89   1st Qu.:  9.00   young  : 197
##  before                 :  85   Median : 13.00   younger: 111
##  after                  :  82   Mean   : 38.43
##  Naiad,Hostel,CP,Chimera:  35   3rd Qu.: 28.00
##  CP_Alert               :  27   Max.   :569.00
##  (Other)                :  88
##      sex       logdivetime      simpleboat        divetype
##  female:764   Min.   :0.000   Min.   :0.0000   bottom : 276
##  male  :882   1st Qu.:2.197   1st Qu.:0.0000   surface:1370
##               Median :2.565   Median :0.0000
##               Mean   :2.873   Mean   :0.1452
##               3rd Qu.:3.332   3rd Qu.:0.0000
##               Max.   :6.344   Max.   :1.0000
##
##      bouts         individual
##  Min.   :  1.0   Min.   : 1.00
##  1st Qu.:275.0   1st Qu.:10.00
##  Median :622.0   Median :20.00
##  Mean   :516.9   Mean   :18.88
##  3rd Qu.:695.0   3rd Qu.:28.00
##  Max.   :862.0   Max.   :37.00
##
```

```
names(bottlenose)
```

```
##  [1] "order"       "Idsample"    "Date"        "boat"        "divetime"
##  [6] "age"         "sex"         "logdivetime" "simpleboat"  "divetype"
## [11] "bouts"       "individual"
```

```
str(bottlenose)
```

```
## 'data.frame':    1646 obs. of  12 variables:
##  $ order      : int  408 409 410 411 412 761 762 763 764 765 ...
##  $ Idsample   : int  25 25 25 25 25 35 35 35 35 35 ...
##  $ Date       : Factor w/ 45 levels "1/4/2001","1/6/2000",..: 7 7 7 7 7 13 13 13 13 13 ...
```

```
##  $ boat       : Factor w/ 14 levels "4_kayak","4SMOUT",..: 14 14 14 14 14 14 14 14 14 14 ...
##  $ divetime   : int  29 68 32 80 8 2 10 15 22 6 ...
##  $ age        : Factor w/ 3 levels "old","young",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ sex        : Factor w/ 2 levels "female","male": 2 2 2 2 2 2 2 2 2 2 ...
##  $ logdivetime: num  3.37 4.22 3.47 4.38 2.08 ...
##  $ simpleboat : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ divetype   : Factor w/ 2 levels "bottom","surface": 2 1 2 1 2 2 2 2 2 2 ...
##  $ bouts      : int  9 9 10 10 11 12 12 12 12 12 ...
##  $ individual : int  37 37 37 37 37 11 11 11 11 11 ...
```

Subsetting data

Let's work with only a subset of this data, selecting from the `data.frame` by columns.

```
bottlenose_subset <- bottlenose[c("divetime", "logdivetime", "bouts")]
```

What if we wanted to subset by row/column number?

```
bottlenose_subset[1, 1]
```

```
## [1] 29
```

```
bottlenose_subset[5, 2]
```

```
## [1] 2.079442
```

Let's get a specific row.

```
bottlenose_subset[3, ]
```

```
##   divetime logdivetime bouts
## 3       32    3.465736    10
```

And a whole column

```
bottlenose_subset[ , 2 ]
```

Just part of the column

```
bottlenose_subset[1:10, 2 ]
```

```
##  [1] 3.3672958 4.2195077 3.4657359 4.3820266 2.0794415 0.6931472 2.3025851
##  [8] 2.7080502 3.0910425 1.7917595
```

Specific rows

```
bottlenose_subset[c(3, 5, 7), ]
```

```
##   divetime logdivetime bouts
## 3       32    3.465736    10
## 5        8    2.079442    11
## 7       10    2.302585    12
```

Simple calculations / built-in functions

Some statistics of note.

```
mean(bottlenose_subset$divetime)
```

```
## [1] 38.42831
```

```
max(bottlenose_subset$divetime)
```

```
## [1] 569
```

```
median(bottlenose_subset$divetime)
```

```
## [1] 13
```

What about getting this for all three columns at once?

```
apply(X = bottlenose_subset, MARGIN = 2, FUN = mean)
```

```
##    divetime logdivetime       bouts
##   38.428311    2.872534  516.927096
```

We could also do this by row.

```
apply(X = bottlenose_subset, MARGIN = 1, FUN = mean)
```

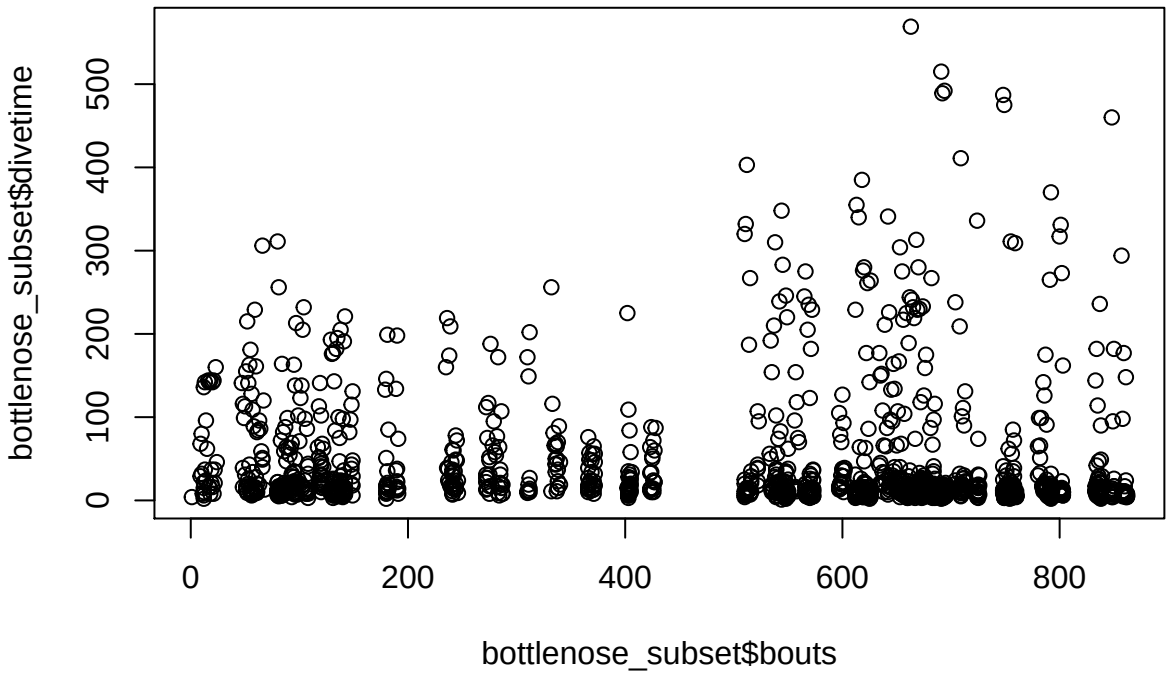
But we should be careful, because in this case, it doesn't make any sense to do this.

Challenge

Use indexing and the functions we just learned to determin the mean, min, and max of rows 21 to 45.

Simple plots

```
plot(x = bottlenose_subset$bouts, y = bottlenose_subset$divetime)
```



Challenge

Plot only the values from rows 21 to 45.