

Crash Course in R

Matthew E. Aiello-Lammens

Contents

Use a script file	1
R as a calculator	1
Variables	2
Loading Data	2
Navigating data	3
Fixing or Cleaning Data	4
Subset the data	4
Simple calculations / built-in functions	5
Simple plots	5

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.
```

Challenge

I change 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 your data is in, then call it in there.

```
fral_pres <- read.csv(file = "https://www.dropbox.com/s/x7s7fpu4bepj7xf/F_alnus_CompiledPres.csv?dl=1")
```

Let's have a look at these data

```
head(fral_pres)
```

```
##           SPEC      LONG      LAT UNCERTAIN      PRIM_SOURCE
## 1 Frangula_alnus -71.17625 44.15129      NA WMNF Invasive Survey
## 2 Frangula_alnus -71.22623 44.17989      NA WMNF Invasive Survey
## 3 Frangula_alnus -71.19045 44.05796      NA WMNF Invasive Survey
## 4 Frangula_alnus -71.88835 43.84155      NA WMNF Invasive Survey
## 5 Frangula_alnus -71.18697 44.14703      NA WMNF Invasive Survey
## 6 Frangula_alnus -71.10204 44.15750      NA WMNF Invasive Survey
##   FIELD_HERB YEAR
## 1      Field 2002
## 2      Field 2001
## 3      Field 2006
## 4      Field 2002
## 5      Field 2002
## 6      Field 2001
```

```
tail(fral_pres)
```

```
##           SPEC      LONG      LAT UNCERTAIN PRIM_SOURCE FIELD_HERB
## 2350 Frangula_alnus -90.63560 44.19780      NA      GLIFWC      Unknown
## 2351 Frangula_alnus -89.57310 45.80220      NA      GLIFWC      Unknown
## 2352 Frangula_alnus -80.08867 40.54090      NA      GLIFWC      Unknown
## 2353 Frangula_alnus -88.22000 42.57000      NA      GLIFWC      Unknown
## 2354 Frangula_alnus -86.94662 45.87737      NA      GLIFWC      Unknown
```

```
## 2355 Frangula_alnus -87.65442 41.85320      NA      GLIFWC      Unknown
##      YEAR
## 2350 2012
## 2351 2012
## 2352 2012
## 2353 2012
## 2354 2012
## 2355 2012
```

```
summary(fral_pres)
```

```
##              SPEC              LONG              LAT              UNCERTAIN
## Frangula_alnus:2355  Min.   :-96.61  Min.   :38.60  Min.   :   10
##                   1st Qu.: -89.54  1st Qu.:42.46  1st Qu.:   10
##                   Median : -77.01  Median :43.68  Median :   10
##                   Mean    : -80.98  Mean    :43.89  Mean    : 2414
##                   3rd Qu.: -71.56  3rd Qu.:45.82  3rd Qu.: 1000
##                   Max.    : -63.00  Max.    :47.82  Max.    :40000
##                                NA's    :1506
##  PRIM_SOURCE      FIELD_HERB      YEAR
## GLIFWC :827      Field    :1631  Min.    :1879
## IPANE   :553      Herbarium: 643  1st Qu.:2001
## NY_iMAP:308      Unknown   : 81   Median :2004
## WIS      : 85                                Mean    :1998
## CONN     : 84                                3rd Qu.:2008
## CM       : 68                                Max.    :2012
## (Other):430                                NA's    :3
```

```
names(fral_pres)
```

```
## [1] "SPEC"      "LONG"      "LAT"      "UNCERTAIN" "PRIM_SOURCE"
## [6] "FIELD_HERB" "YEAR"
```

```
str(fral_pres)
```

```
## 'data.frame':   2355 obs. of  7 variables:
## $ SPEC      : Factor w/ 1 level "Frangula_alnus": 1 1 1 1 1 1 1 1 1 ...
## $ LONG      : num  -71.2 -71.2 -71.2 -71.9 -71.2 ...
## $ LAT       : num   44.2 44.2 44.1 43.8 44.1 ...
## $ UNCERTAIN  : int   NA NA NA NA NA NA NA NA NA NA ...
## $ PRIM_SOURCE: Factor w/ 34 levels "A","ACAD","B",...: 33 33 33 33 33 33 33 33 33 ...
## $ FIELD_HERB : Factor w/ 3 levels "Field","Herbarium",...: 1 1 1 1 1 1 1 1 1 ...
## $ YEAR       : int   2002 2001 2006 2002 2002 2001 2006 2007 2005 2002 ...
```

Navigating data

What if we wanted to look at specific row/column entries?

```
fral_pres[1, 1]
```

```
## [1] Frangula_alnus
## Levels: Frangula_alnus
```

```
fral_pres[1, 2]
```

```
## [1] -71.17625
```

Let's get a specific row.

```
fral_pres[3, ]
```

```
##           SPEC      LONG      LAT UNCERTAIN      PRIM_SOURCE
## 3 Frangula_alnus -71.19045 44.05796      NA WMNF Invasive Survey
##   FIELD_HERB YEAR
## 3      Field 2006
```

And a whole column

```
fral_pres[, 2 ]
```

Just part of the column

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

```
## [1] -71.17625 -71.22623 -71.19045 -71.88835 -71.18697 -71.10204 -71.22834
## [8] -71.13196 -71.10707 -71.15898
```

Specific rows

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

```
##           SPEC      LONG      LAT UNCERTAIN      PRIM_SOURCE
## 3 Frangula_alnus -71.19045 44.05796      NA WMNF Invasive Survey
## 5 Frangula_alnus -71.18697 44.14703      NA WMNF Invasive Survey
## 7 Frangula_alnus -71.22834 44.18053      NA WMNF Invasive Survey
##   FIELD_HERB YEAR
## 3      Field 2006
## 5      Field 2002
## 7      Field 2006
```

How about rows that meet certain criteria?

```
subset(fral_pres, FIELD_HERB == "Herbarium")
```

Fixing or Cleaning Data

Let's say we realized that we had a mistake in our data. For example, one of `UNCERTAIN` values was recorded incorrectly. How can we change this?

```
fral_pres_fixed <- fral_pres
fral_pres_fixed$UNCERTAIN[1]
```

```
## [1] NA
```

```
fral_pres_fixed$UNCERTAIN[1] <- 20
fral_pres_fixed$UNCERTAIN[1]
```

```
## [1] 20
```

Subset the data

Let's get only a subset of these data, selecting from the `data.frame` by columns.

```
fral_pres_subset <- fral_pres[c("SPEC", "LONG", "LAT")]
```

Next, let's rename our columns so they are in the format used in Wallace.

```
names(fral_pres_subset)
```

```
## [1] "SPEC" "LONG" "LAT"
```

```
names(fral_pres_subset) <- c("name", "longitude", "latitude")
```

Let's now make a new file with the fixed data.

```
write.csv(x = fral_pres_subset, file = "~/Dropbox/SCCS-Workshop/fral_pres.csv", row.names = FALSE)
```

Simple calculations / built-in functions

Some statistics of note.

```
mean(fral_pres$LAT)
```

```
## [1] 43.89405
```

```
max(fral_pres$LAT)
```

```
## [1] 47.81744
```

```
min(fral_pres$LAT)
```

```
## [1] 38.6
```

```
median(fral_pres$LAT)
```

```
## [1] 43.67897
```

Challenge

Use indexing and the functions we just learned to determine the mean, min, and max latitude and longitude of all of the Herbarium specimens vs. the Field specimens.

Simple plots

```
plot(x = fral_pres$LONG, y = fral_pres$LAT)
```

