

# E2M2: R basics



**Fidisoa Rasambainarivo**

Mahaliana Labs

[fidy@Mahaliana.org](mailto:fidy@Mahaliana.org)

[www.mahaliana.org](http://www.mahaliana.org)



*Mahaliana*  
IT ALWAYS STARTS WITH A QUESTION



# Objective

- To teach the basic knowledge necessary to use R.
  - What is R?
  - Why use R?
  - How R works?
  - Your environment in R and R studio
- Experience R



# ***1. Introduction***



# What is R?

- R is a language and environment for statistical computing and graphics. It is used for
  - Data management
  - Statistical analysis
  - Scientific programming and simulation
  - Interfacing with other programs (GIS...)

# What is R?

- R is a language and environment for statistical computing and graphics. It is used for
  - Data management
  - Statistical analysis
  - Scientific programming and simulation
  - Interfacing with other programs (GIS...)
- **Language** because it allows you to communicate flexibly with your computer.

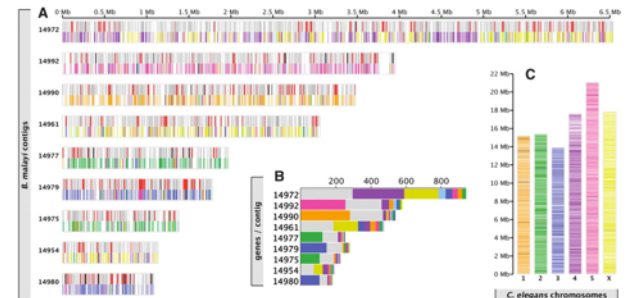
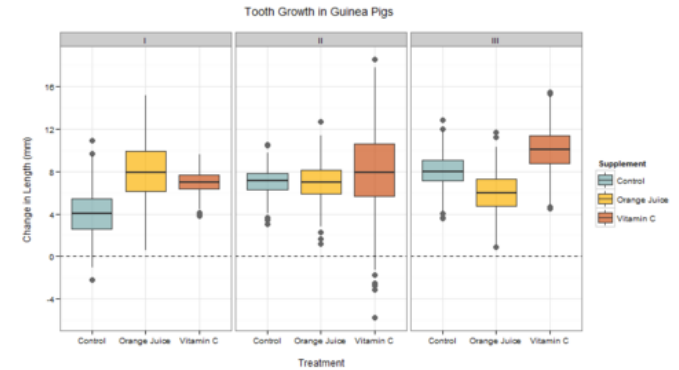
*Like any other language:*

- Learning R will be easier for some than for others **AND it is okay!!!**
- Learning R takes work and practice

# Why use R?

## 1. R is free!!!!

1. SPSS \$99/month
2. SAS \$2,500/year



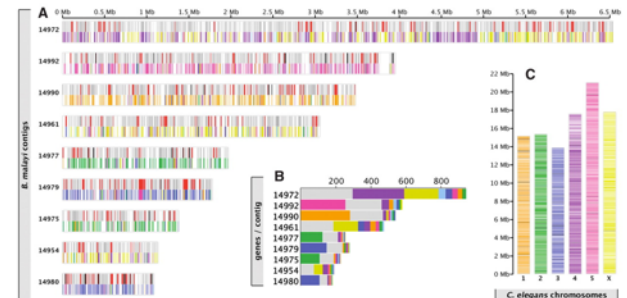
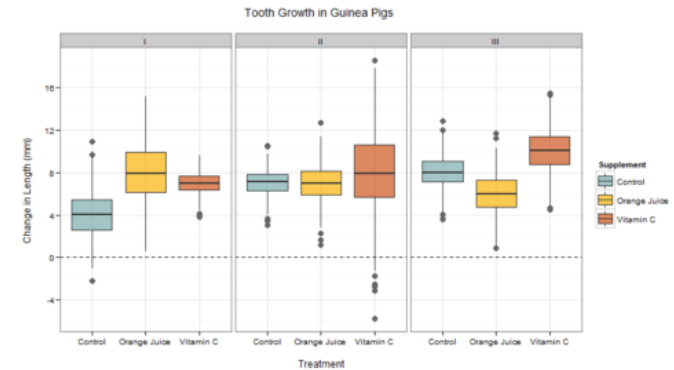
# Why use R?

## 1. R is free!!!!

1. SPSS \$99/month
2. SAS \$2,500/year

## 2. Excellent at making figures

## 3. Thousands of tools for statistical analysis (packages).

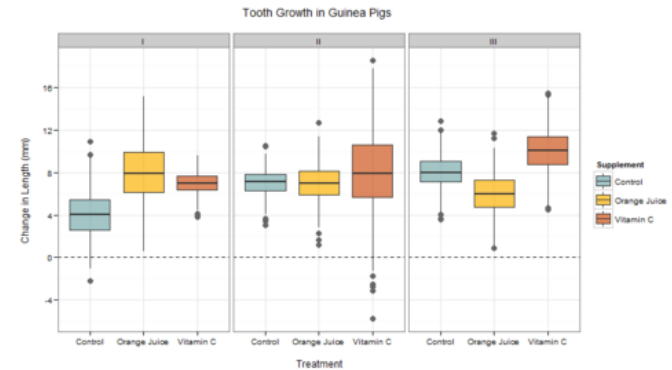


# Why use R?

## 1. R is free!!!!

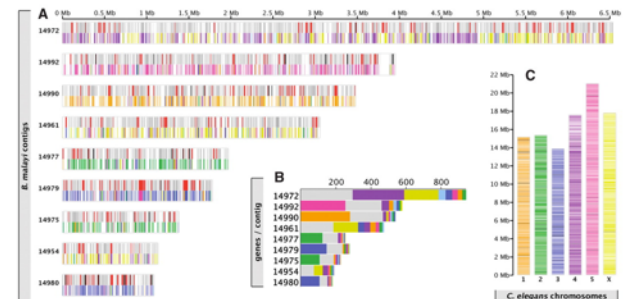
1. SPSS \$99/month
2. SAS \$2,500/year

## 2. Excellent at making figures



## 3. Thousands of tools for statistical analysis (packages).

## 4. Many recently developed tools available immediatly

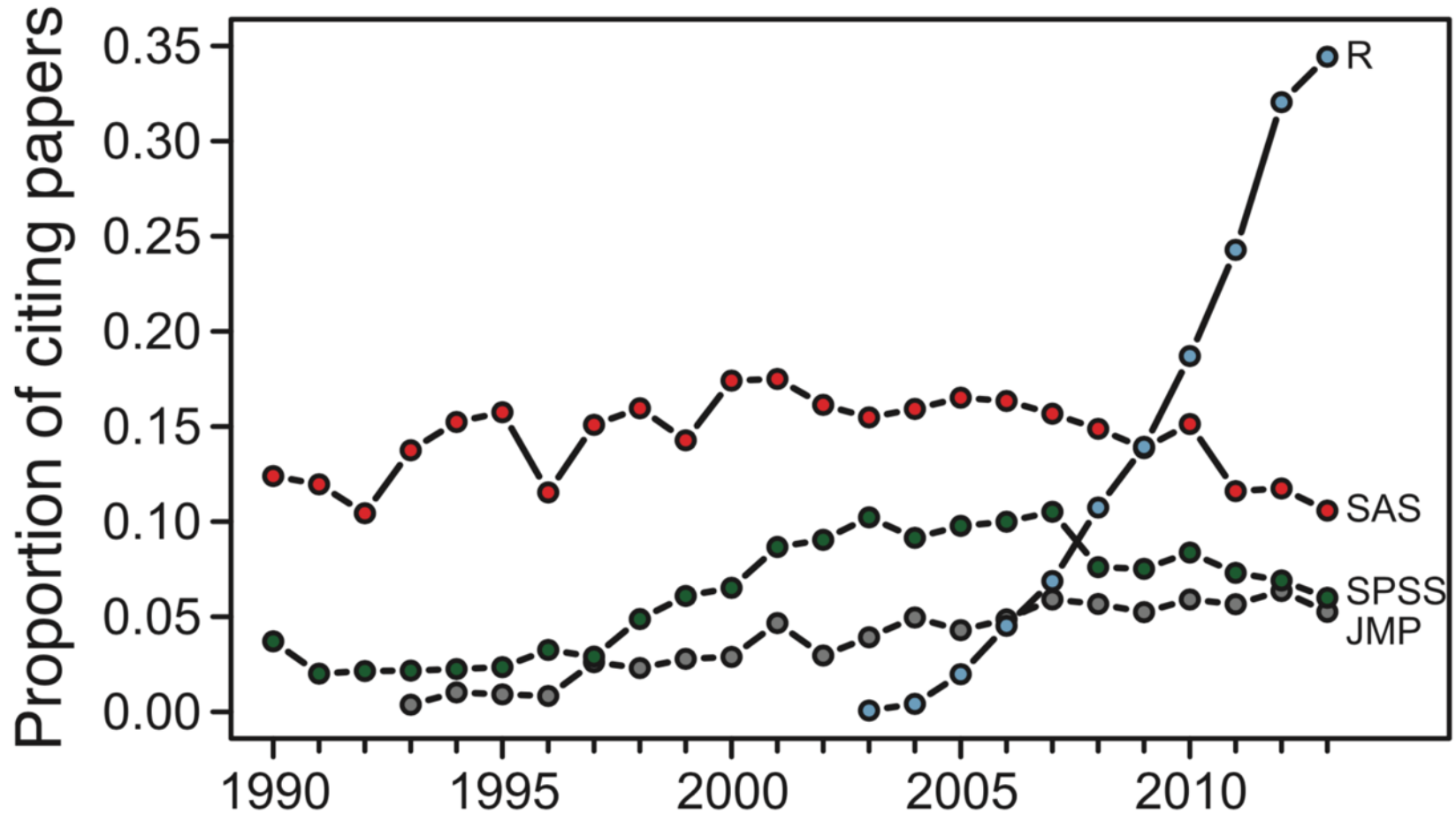


## 5. Freedom to develop your own tools



# Why use R?

## 1. Software of reference in ecology





# Why use R and how does it work?

The base program is very small (~65 mb)

- Designed to have task-specific packages downloaded and added to it. **There is probably a package that is designed to do the analysis that you want to do**
- A package is a collections of functions, data, and help files generally centered around certain themes of analyses.
- 10,000+ packages are currently available to download (you will never need most of these)

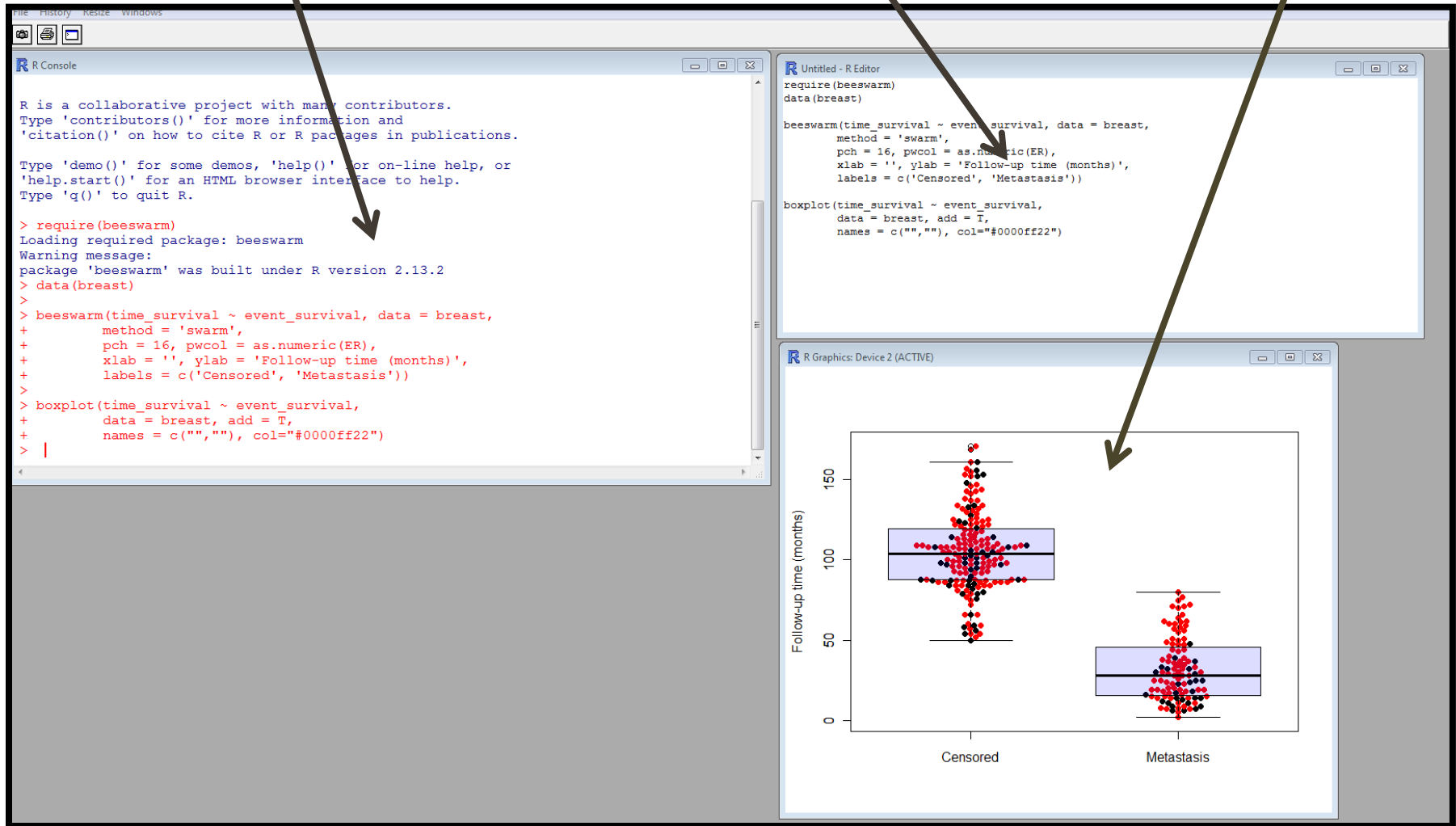


# Your environment in R

Console

Editor

Graphics

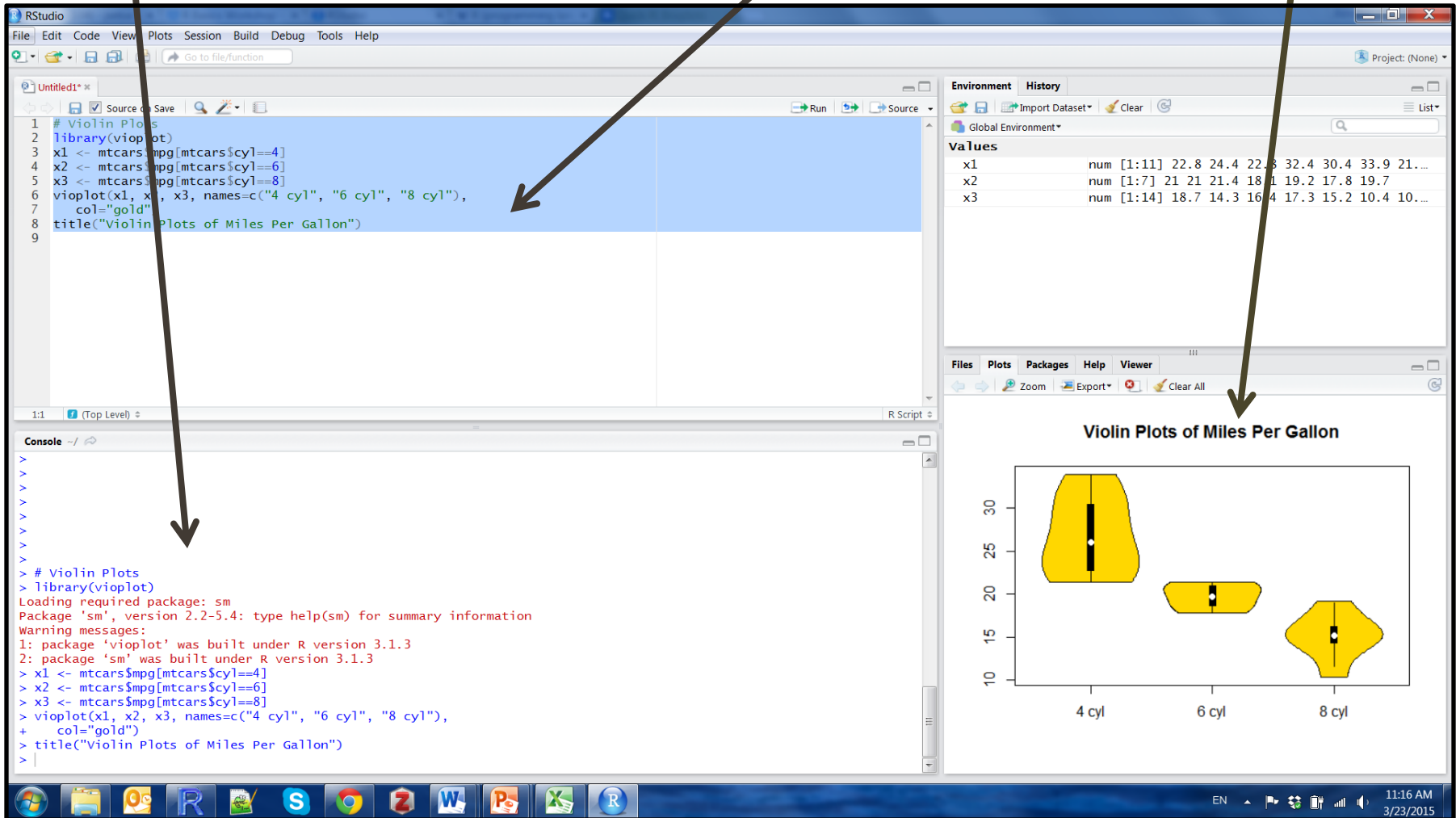


# Main windows in R Studio

Console

Editor

Graphics



The screenshot shows the R Studio interface with three main windows highlighted by arrows:

- Editor:** The top-left window containing R code for creating violin plots.
- Console:** The bottom-left window showing the execution output of the code.
- Graphics:** The bottom-right window displaying the resulting violin plot.

**Editor Code:**

```
1 # Violin Plots
2 library(vioplot)
3 x1 <- mtcars$mpg[mtcars$cyl==4]
4 x2 <- mtcars$mpg[mtcars$cyl==6]
5 x3 <- mtcars$mpg[mtcars$cyl==8]
6 vioplot(x1, x2, x3, names=c("4 cyl", "6 cyl", "8 cyl"),
7         col="gold")
8 title("Violin Plots of Miles Per Gallon")
9
```

**Console Output:**

```
> # Violin Plots
> library(vioplot)
Loading required package: sm
Package 'sm', version 2.2-5.4: type help(sm) for summary information
Warning messages:
1: package 'vioplot' was built under R version 3.1.3
2: package 'sm' was built under R version 3.1.3
> x1 <- mtcars$mpg[mtcars$cyl==4]
> x2 <- mtcars$mpg[mtcars$cyl==6]
> x3 <- mtcars$mpg[mtcars$cyl==8]
> vioplot(x1, x2, x3, names=c("4 cyl", "6 cyl", "8 cyl"),
+         col="gold")
> title("Violin Plots of Miles Per Gallon")
>
```

**Graphics Window:** The plot titled "Violin Plots of Miles Per Gallon" shows the distribution of miles per gallon (mpg) for three car categories: 4 cyl, 6 cyl, and 8 cyl. The y-axis represents mpg values from 10 to 30. The violins are colored gold.

Category	Approximate mpg Range (Violin)
4 cyl	22 - 30
6 cyl	18 - 22
8 cyl	10 - 19

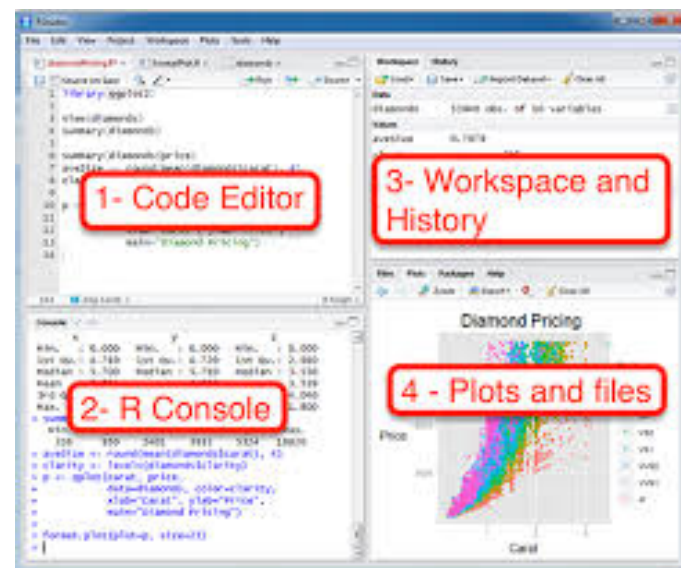
# Working in R/R Studio

- **Always** use a text editor to save your work
  - Allows for repeatability when you save your code.
  - Allows you to add comments to scripts to remember what you have done.
  - Use # to make comments that won't be executed
  - Makes it easy to share code with collaborators
- When you type things into the console and execute them, **they are run but they are not saved.**

- To execute commands:

Mac: ⌘↵, PC: CTRL-R

Can highlight multiple lines of code and run at once





## Exercise 1: a first session in R

- **Objective:** experiencing R/R studio



## ***2. Enter and Import your data***



# Objective

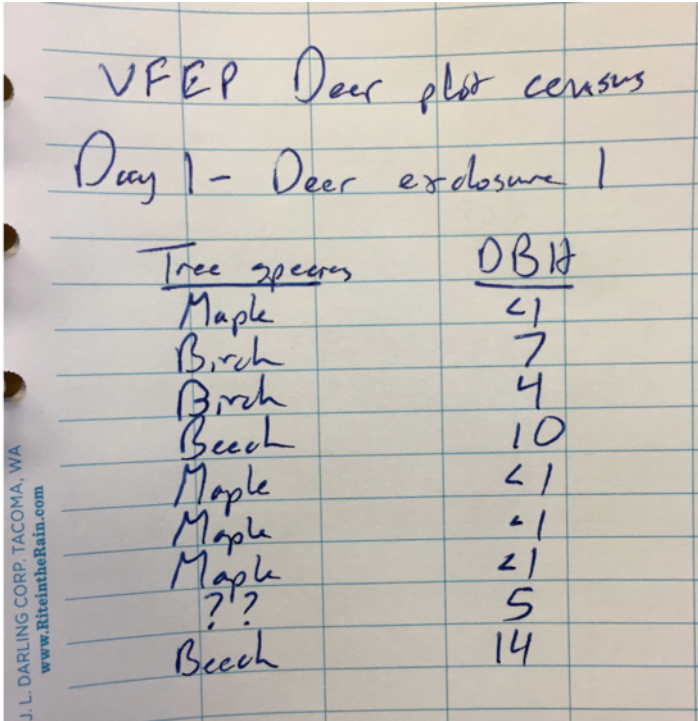
- To teach the basic knowledge necessary to use R.
  - How to record your data?
  - How do you import them into R?
- Experience R: live coding



# Record your data

*Hypothetical data on sizes of trees in deer exclosures*

- Most of the time have a data book where you write down your data, observations, etc.
- Most people use MS Excel to enter and store data from the notebook on the computer.
- **But... BEWARE of how data is recorded on excel**



A photograph of a handwritten data table on lined paper. The title is 'VFEP Deer plot census' and the subtitle is 'Day 1 - Deer exclosure 1'. The table has two columns: 'Tree species' and 'DBH'. The data is as follows:

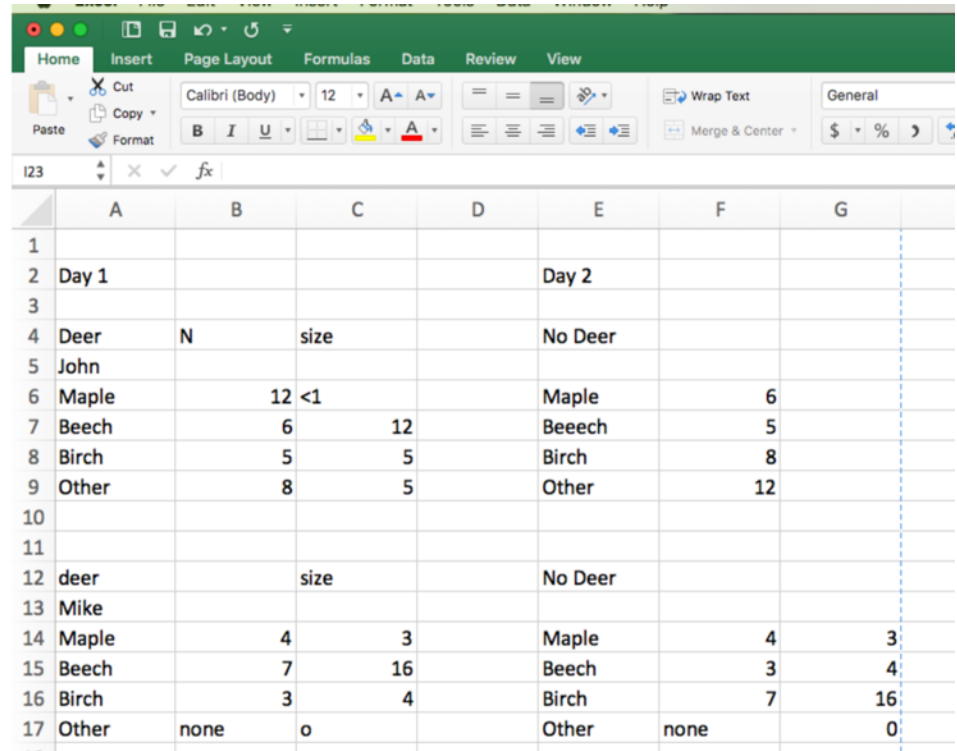
Tree species	DBH
Maple	41
Birch	7
Birch	4
Beech	10
Maple	41
Maple	41
Maple	21
??	5
Beech	14

On the left margin of the paper, there is vertical text: 'J.L. DARLING CORP. TACOMA, WA' and 'www.RiteintheRain.com'.

# Record your data

- Most of the time have a data book where you write down your data, observations, etc.
- Most people use MS Excel to enter and store data from the notebook on the computer.
- **But... BEWARE of how data is recorded on excel**

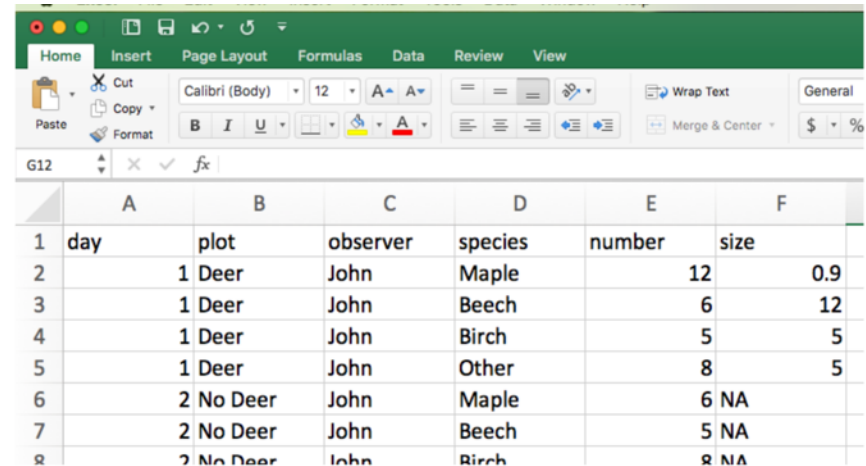
*Hypothetical data on sizes of trees in deer exclosures*



	A	B	C	D	E	F	G
1							
2	Day 1				Day 2		
3							
4	Deer	N	size		No Deer		
5	John						
6	Maple	12	<1		Maple	6	
7	Beech	6	12		Beech	5	
8	Birch	5	5		Birch	8	
9	Other	8	5		Other	12	
10							
11	deer		size		No Deer		
12	Mike						
13	Maple	4	3		Maple	4	3
14	Beech	7	16		Beech	3	4
15	Birch	3	4		Birch	7	16
16	Other	none	o		Other	none	0

# Record your data: general rules

- Avoid spaces: use period “.” or underscore “\_”.
- Keep column names short, simple and unique.
- **Be very careful of typos.**

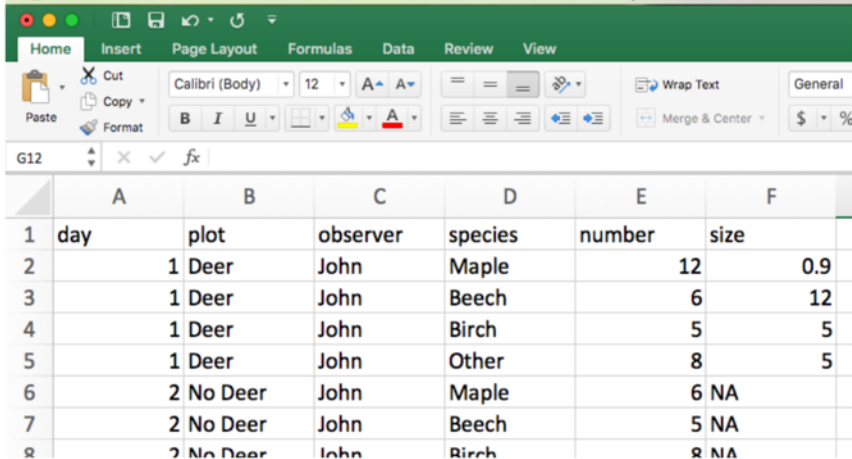


The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D	E	F
1	day	plot	observer	species	number	size
2		1 Deer	John	Maple	12	0.9
3		1 Deer	John	Beech	6	12
4		1 Deer	John	Birch	5	5
5		1 Deer	John	Other	8	5
6		2 No Deer	John	Maple	6	NA
7		2 No Deer	John	Beech	5	NA
8		2 No Deer	John	Birch	8	NA

# Record your data: general rules

- Avoid spaces: use period “.” or underscore “\_”.
- Keep column names short, simple and unique.
- **Be very careful of typos.**
- **One variable per column (no merged column, no more than one).**
- Consistent unit throughout observations
- One observation per cell.
- Save as csv file



The screenshot shows a Microsoft Excel spreadsheet with a green ribbon at the top. The ribbon includes tabs for Home, Insert, Page Layout, Formulas, Data, Review, and View. The Home tab is active, showing options for Paste, Cut, Copy, Format, and various text and alignment tools. The spreadsheet itself has a grid with columns labeled A through F and rows numbered 1 through 8. The data is organized as follows:

	A	B	C	D	E	F
1	day	plot	observer	species	number	size
2		1 Deer	John	Maple	12	0.9
3		1 Deer	John	Beech	6	12
4		1 Deer	John	Birch	5	5
5		1 Deer	John	Other	8	5
6		2 No Deer	John	Maple	6	NA
7		2 No Deer	John	Beech	5	NA
8		2 No Deer	John	Birch	8	NA

# Import data in R

```
> data1<-read.csv("file_location/file_name.csv")
```

Make it  
an object!

Function

Location of the file on your  
computer, **in quotes**