E2M2: R basics

Fidisoa Rasambainarivo
Mahaliana Labs
fidy@Mahaliana.org
www.mahaliana.org
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 immediately

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

# Violin Plots
library(ggplot2)
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")

Violin Plots of Miles Per Gallon
**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

- 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

<table>
<thead>
<tr>
<th>Tree species</th>
<th>DBH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maple</td>
<td>41</td>
</tr>
<tr>
<td>Birch</td>
<td>7</td>
</tr>
<tr>
<td>Beech</td>
<td>4</td>
</tr>
<tr>
<td>Maple</td>
<td>10</td>
</tr>
<tr>
<td>Maple</td>
<td>2</td>
</tr>
<tr>
<td>Maple</td>
<td>2</td>
</tr>
<tr>
<td>Beech</td>
<td>5</td>
</tr>
<tr>
<td>Beech</td>
<td>14</td>
</tr>
</tbody>
</table>
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
Record your data: general rules

• Avoid spaces: use period “.” or underscore “_”.
• Keep column names short, simple and unique.
• Be very careful of typos.
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
Import data in R

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

Make it an object!

Function

Location of the file on your computer, in quotes