

IRC_Data_Analysis_Training_1

Rachel Kenny

1/11/2021

Introduction to R and Rstudio

What is “R”?

R is a coding language and environment for statistical analysis and visualization. When we refer to R, we may be referring to the language “R” or to the software itself.

What is “Rstudio” and why do I need it if I already have R installed?

Rstudio is a program that uses the R language, but provides a much more friendly user interface that allows for more intuitive computing and analyses. It is the most common program that people use to write R code.

What is an R script and an R markdown?

An R script is the more basic file version for writing code. An Rmarkdown is more advanced and gives you more organizational and formatting options. Think of an R script like a text file, and an R markdown like a word file. Rmarkdowns allow you to more easily create sections, organize your code chunks, and add/format text.

Let’s get started!

The first thing you do anytime you open a new R script or markdown document is to load our packages and data.

What is an R package?

An R package is a bundle of functions, code, and or data sets developed either by R staff or more commonly by the community. They give you additional functionality to what is already built into R, also known as the “base R functions”. To install a new packages, go to “packages” in the right hand panel, click “install”, then type in the name of the package and hit “install”. Once they are installed, packages can be loaded by running the function

Within a code chunk

- Use a `[#]` to tell R not to treat text as code when you are trying to make notes, or highlight the relevant text and hit `[ctrl + alt + C]` PC / `[command + alt + C]` MAC

- Using [`<-`] defines an object (dataframe, plot, etc.) within the code. For example your data file likely has a long name but might be easier to refer to as “veg_data” within code.
- To run code click “run” in the upper right hand corner and select an option for line, chunk, document etc., or hit [`ctrl + enter`] PC / [`cmd + enter`] MAC to run one line or selected code
- Functions always use parentheses (`()`). A common error is adding too many or too few

Load packages

```
library(tidyverse) # The tidyverse actually contains multiple packages, and is useful for data wrangling.

## Warning: package 'tidyverse' was built under R version 4.0.2
## -- Attaching packages ----- tidyverse 1.3.0 --
## v ggplot2 3.3.2      v purrr   0.3.4
## v tibble  3.0.3      v dplyr  1.0.1
## v tidyr   1.1.1      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.5.0
## Warning: package 'ggplot2' was built under R version 4.0.2
## Warning: package 'tibble' was built under R version 4.0.2
## Warning: package 'tidyr' was built under R version 4.0.2
## Warning: package 'dplyr' was built under R version 4.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
library(here) # The here library makes it easier to organize and load data so that our project folder d
## here() starts at /Users/rachelkenny/Documents/IRC/R Code/IRC_Data_Analysis_Training
library(readr) # This library contains functions to help you read data.
library(ggplot2) # The ggplot library is useful for visualizing data, creating plots and graphs.
library(rmarkdown) # The rmarkdown library contains code relevant to rmarkdown files, such as knitting
library(janitor) # This helps clean data. R does not like spaces and special characters in column names

##
## Attaching package: 'janitor'
## The following objects are masked from 'package:stats':
##
##   chisq.test, fisher.test
```

Load data

```
# When loading data, remember to include the full file name including extension (.csv or .xls etc.)
oak_data_raw <- read_csv(here("data", "Weir_Oak_Restoration_Data_winter19.csv"))

## Parsed with column specification:
## cols(
```

```
## `Short ID` = col_character(),
## Survival = col_logical(),
## Quantity = col_double(),
## `Height (cm)` = col_double(),
## `Open Closed` = col_character(),
## `Location UML` = col_character(),
## `Water Yes No` = col_character(),
## `Sampling Group` = col_character()
## )
```

```
View(oak_data_raw) # The "view" function allows you to view your data
```

```
# Now that I have my raw data loaded, I'm going to "clean" the column names so that there are no spaces
oak_data <- clean_names(oak_data_raw)
```

```
View(oak_data)
```

Common operators in R include the following

- `[*]` multiply
- `[/]` divide
- `[>]` greater than
- `[>=]` greater than or equal to
- `[<]` less than
- `[<=]` less than or equal to
- `[=]` equal to
- `[==]` find a match in the data (for example, when filtering)

Wrangle data

```
# To reveal column names, use the [names] function
names(oak_data)
```

```
## [1] "short_id"      "survival"      "quantity"      "height_cm"
## [5] "open_closed"   "location_uml"   "water_yes_no"   "sampling_group"
```

```
# To only retain certain columns, use the function [select]
oak_data_selection <- oak_data %>%
  select(survival, height_cm)
```

```
View(oak_data_selection)
```

```
# To filter data, use the function [filter]
oak_data_survived <- oak_data %>%
  filter(survival == "TRUE", height_cm >= 6)
```

```
View(oak_data_survived)
```

Visualize data

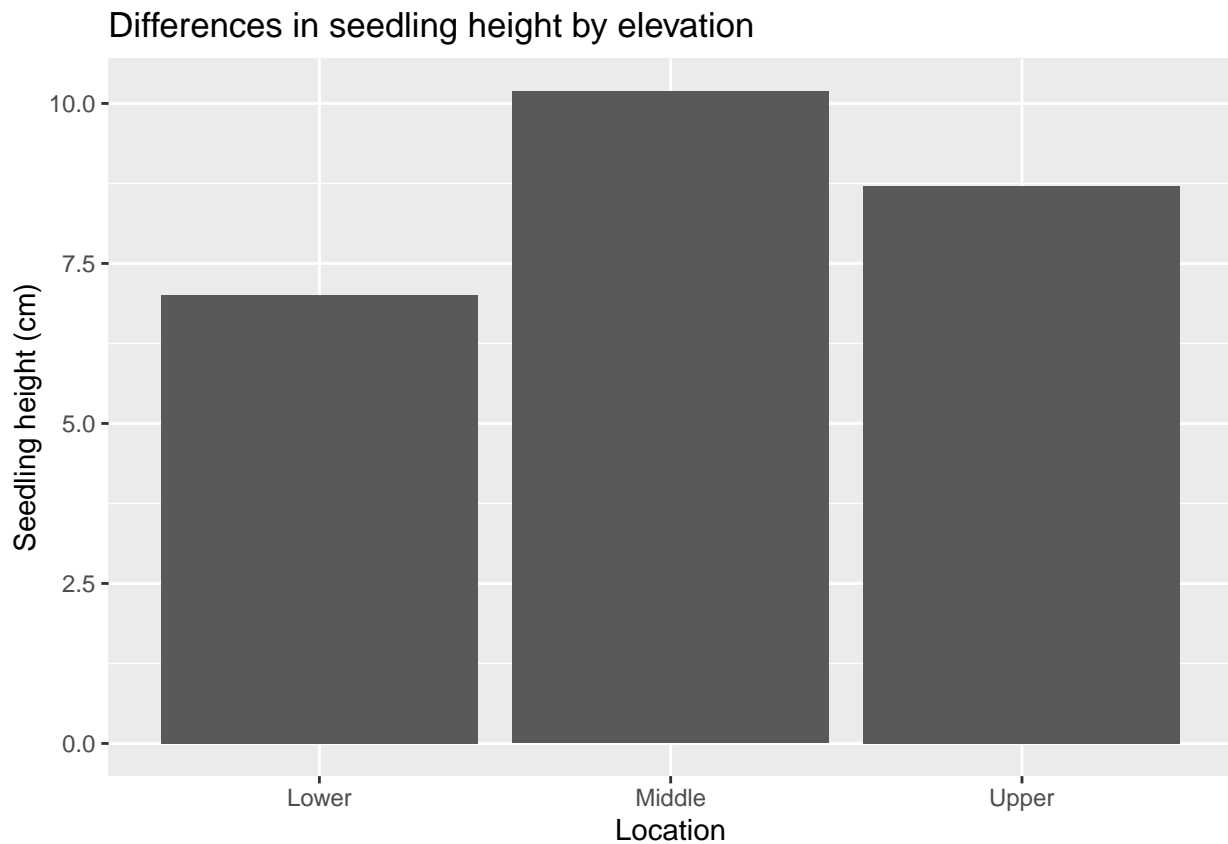
```
# GGplot allows us to visualize our data in a number of ways. Here's an example below.
```

```
# Differences in seedling height between acorns planted at three elevations: upper, middle, and lower
oak_plot <- ggplot(oak_data) +
  geom_bar(aes(x = location_uhl, y = height_cm),
           position = "dodge", stat = "summary", fun.y = "mean") +
  ggtitle("Differences in seedling height by elevation") +
  xlab("Location") +
  ylab("Seedling height (cm)")
```

```
## Warning: Ignoring unknown parameters: fun.y
```

```
oak_plot
```

```
## No summary function supplied, defaulting to `mean_se()`
```



```
# This plot shows a count of the number of seedlings that survived vs didn't
```

```
oak_plot2 <- ggplot(oak_data, aes(survival)) +
  geom_bar() +
  xlab("Survival") +
  ylab("Count") +
  ggtitle("Seedling Survival in Weir Canyon")
```

```
oak_plot2
```

