Introduction to R for Plant Breeding Applications

September 5, 2019
Bahir Dar, Ethiopia

Hannah De Jong

Downloading R and Rstudio

- **R** is the statistical software

https://cran.r-project.org/mirrors.html

- **Rstudio** is an environment that will make it easier for you to develop R code

https://www.rstudio.com/products/rstudio/download

Suggested reference books

- O'Reilly *Learning R*
  - Best for an overview of the different components of R, when learning for the first time

- O'Reilly *R Cookbook*
  - Have a specific problem? Look here!

I'll reference specific chapters here on each slide if you want to learn more

Why do we use R?

...instead of Microsoft Excel or JMP?
– Ability to perform repeated statistical tests without lots of clicking around!
– Much prettier graphs
– It's free!

...instead of another programming language (Python, Perl, Java, etc?)
– R is specifically designed for data manipulation and statistics

RStudio Layout

Overview

- Part I: Basics
  - R data structures
  - R operations
  - Reading and writing to files
  - Plotting data

- Part II: Application- Analyzing NVT data

Comments are important!!

```
#This is a comment
code
```

- Comments will help you and others understand what your code does
- Add them as you are writing! Don't wait until later

Data Structures

Scalar	Vector	Matrix
x = 3	v = c(1,2,3)	m = matrix(c(1,2,3,4,5,6), nrow = 2)
x [3]	v [1] [2] [3]	m [1 2 3] [4 5 6]

Data Structures

Scalar	Vector	Matrix
x = "hello"	v = c("hello", "greetings", "good morning")	m = matrix(c("a", "e", "i" ,"o", "u", "y"), nrow = 2)

x

"hello"

v

"hello"
"greetings"
"good morning"

m

"a"	"e"	"i"
"o"	"u"	"y"

Basic operations on scalars

x + y

x − y

x * y

x / y

x % y (modulo)

sqrt(x)

x^ y

Exercise 1

See "Exercises.R" code

Solutions can be found in
"Exercises_solutions.R"

Basic operations on vectors

mean(v)

median(v)

max(v)

min(v)

length(v)   # Gives number of elements in vector

v + 2   # Adds 2 to every value in the vector

c(v1, v2)   # Concatenates vectors

Basic operations on matrices

```
m + 1       #Add 1 to every element of matrix m
mean(m)  #Calculate mean of all elements in m


cbind(m1, m2)       #Concatenate by columns
rbind(m1, m2)       #Concatenate by rows


dim(m)                    #Gives dimensions of m
```

Element operations on matrices

Find a certain row or column:

    m[3,] # get the third row

    m[,5] # get the fifth column

Change one element:

    m[4,7] <- 8 # change the element in
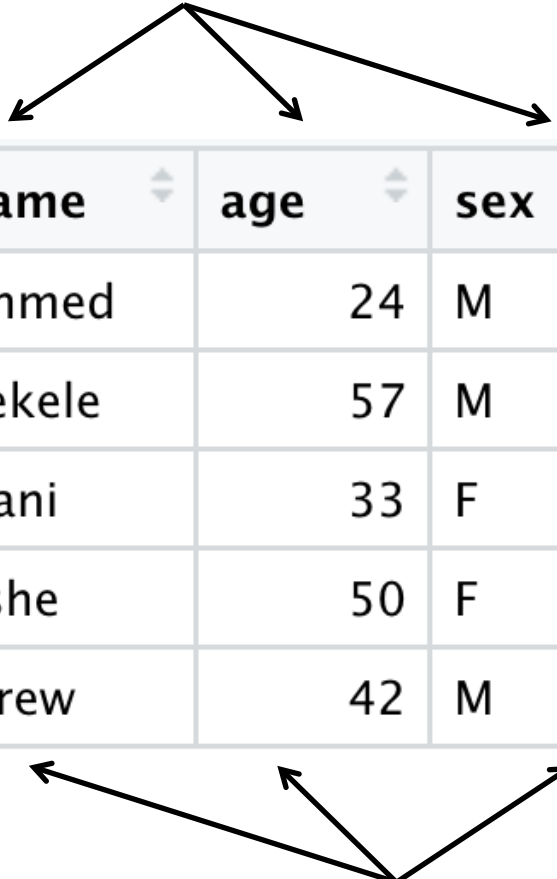
                 # row 4, col 7 to the value 8

Exercise 2

See "Exercises.R" code

Solutions can be found in "Exercises_solutions.R"

R's special data structure: data frames

Named columns

	name	age	sex
1	Ahmed	24	M
2	Bekele	57	M
3	Dani	33	F
4	Eshe	50	F
5	Firew	42	M

Different data types

Learning R Ch. 5

Data frames: $ notation

- Data frames allow you to name each column (e.g. variety, yield, color)

- You can then access that column with the column name as shown:

```
> df$variety
[1] "Lehigh"  "Eva"    "YukonGold"
"Burbank" "Belete"
```

Exercise 3

See "Exercises.R" code

Solutions can be found in
"Exercises_solutions.R"

Working directory

- R always needs to know where to read files from and print new files

- To get your current working directory:

getwd()

- To set a new working directory:

setwd(*pathname*)

e.g. setwd("C:/users/Hannah/Rfolder")

Reading and writing to files

- Many times, we don't want to type out all the data we have into the R code

- Instead, we can import data from a pre-existing file

- Most common formats are "tab-delimited text" (tdt or txt) and "comma-separated values" (csv)

- We also want to store our analyzed data in new files

Formatting a file for import to R

- Make sure your data is "rectangular"- all rows and columns are the same length

- Fill in missing values with "NA"

- No funny symbols ($, %, #, etc.) in data- periods and underscores OK

- Better if no spaces

- Save Excel file as csv or tdt

Code to read in a file

```
data <- read.table(
                filename,
                header = T,
                sep = "\t"
                )
```

Code to write to a file

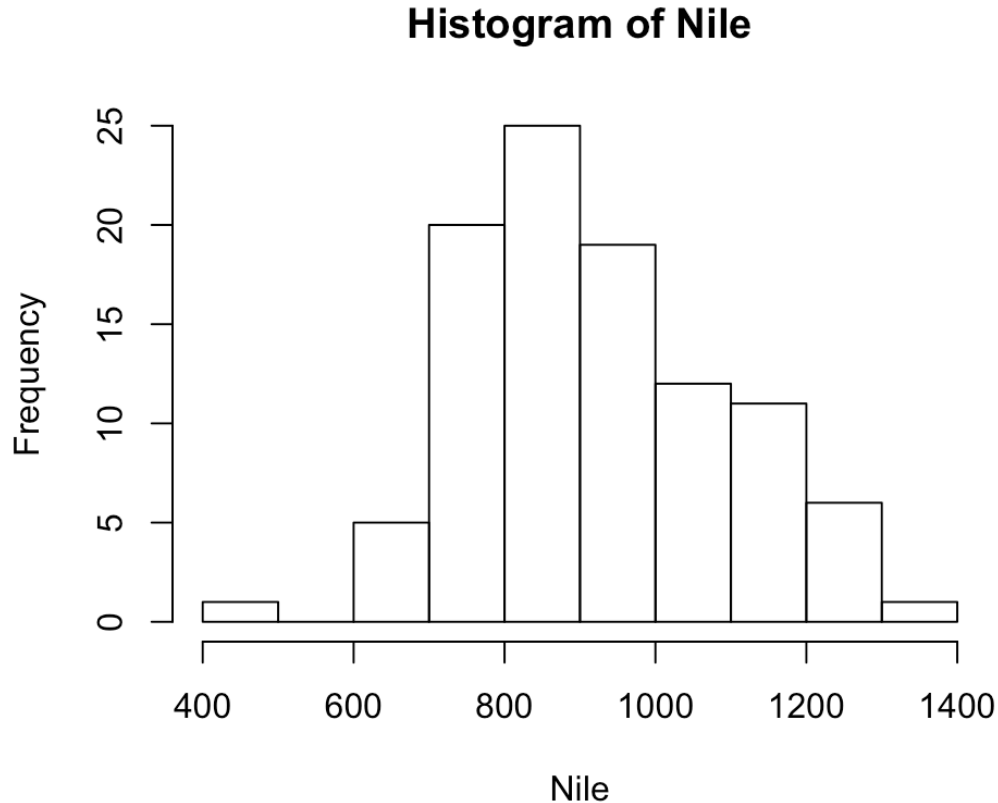```
write.table(
        data,
        filename = "datafile.txt",
        sep = ","
)
```

# Exercise 4

See "Exercises.R" code

Solutions can be found in
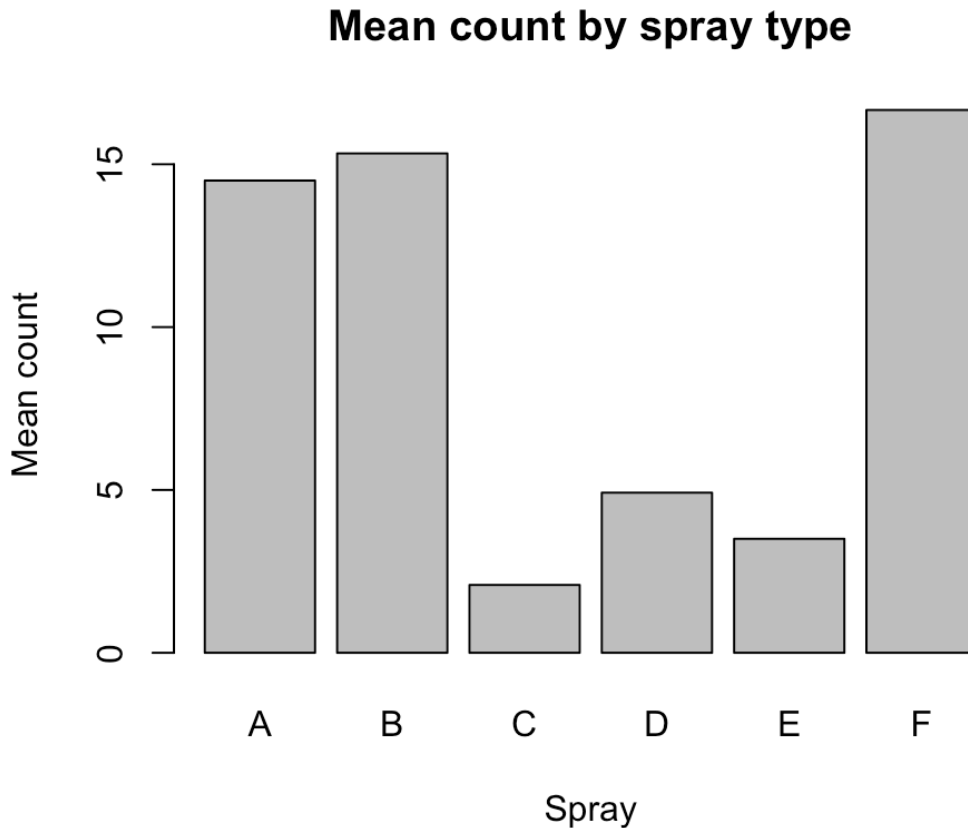"Exercises_solutions.R"

# Plotting: histograms
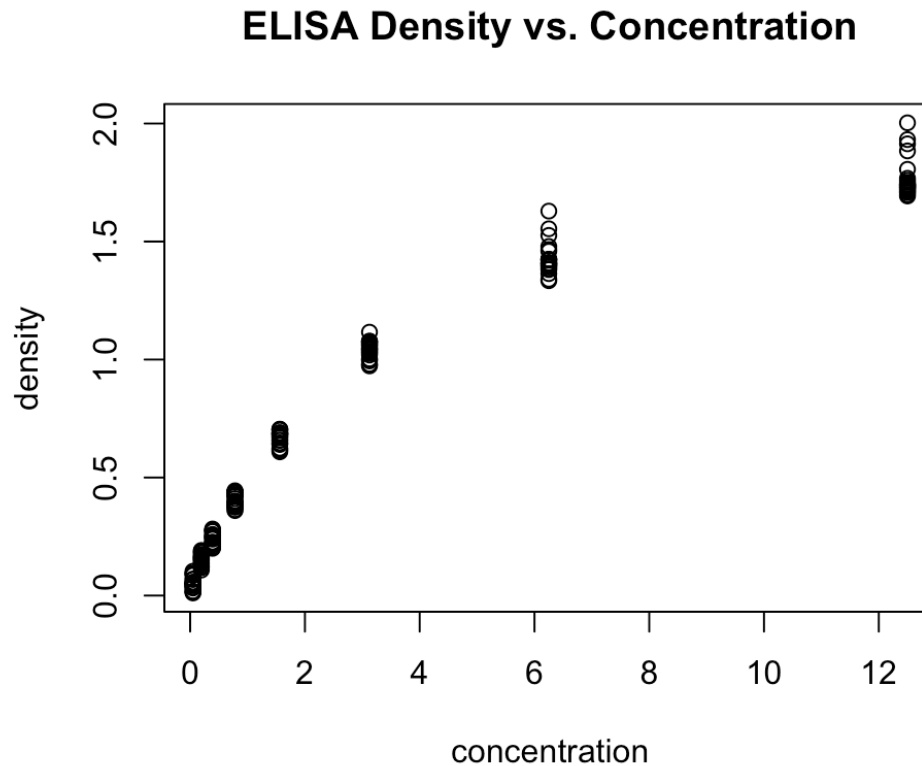
## histogram(data, [other parameters])



**Histogram of Nile**

# Plotting: barplots

## barplot(data, [other parameters])



Mean count by spray type

# Plotting: scatterplots

## plot(x, y, [other parameters])



ELISA Density vs. Concentration

# Exercise 5

See "Exercises.R" code

Solutions can be found in
"Exercises_solutions.R"

# Analyzing Potato Variety Trial Data

See "PotatoVarietyTrialAnalysis.R" code

This code provides some example analysis for potato variety trial data. It can be modified for your own data sets.