R fundamentals

Jeikosd

02 September, 2017

Introduction

Installing R and R-Studio

- Base R **https://cran.r-project.org/mirrors.html**
- RStudio **https://www.rstudio.com/products/RStudio/**

What is R?

- it's a statistical software

What is R?

- it's a statistical software
- it's a object base

What is R?

- it's a statistical software
- it's a object base
  - Types of objects (scalar, vector, matrices, arrays and lists)

What is R?

- it's a statistical software
- it's a object base
  - Types of objects (scalar, vector, matrices, arrays and lists)
  - Assignment of objects

why use R?

- Taken from Hadley Wickham "Fundamentally learning about the world through data is really, really good"
- it's open source

R as calculator

```
2+4
```

```
## [1] 6
```

```
sqrt(16)
```

```
## [1] 4
```

```
3*(2+4)
```

```
## [1] 18
```

More examples

Table 1: Operation Symbols

symbol	Meaning
+	Addition
-	Subtraction
*	Multiplication
/	Division
%%	Modulo (estimates remainder in a division)
^	Exponential

- **See http://www.statmethods.net/management/operators.html**

First Steps in R

Objects in R

- Objects in R obtain values by assignment.
- This is achieved by the gets arrow, <-, and not the equal sign, =.
- Objects can be of different kinds.

Types

- Primitives (numeric, integer, character, logical, factor)
- Data Frames
- Lists
- Tables
- Arrays
- Environments
- Others (functions, closures, promises..)

Simple Types - Vectors

The basic type unit in R is a vector

```
x <- c(1,2,3)
x
## [1] 1 2 3
x <- 1:3
x[1]
## [1] 1
x[0]
## integer(0)
x[-1]
## [1] 2 3
```

Generating Vectors

R provides lots of convenience functions for data generation:

```
rep(0, 5)
## [1] 0 0 0 0 0
seq(1,10)
##  [1]  1  2  3  4  5  6  7  8  9 10
seq(1,2,.1)
##  [1] 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2.0
seq(1,2,length.out=6)
## [1] 1.0 1.2 1.4 1.6 1.8 2.0
```

Indexing

```
x <- c(1, 3, 4, 10, 15, 20, 50, 1, 6)
x > 10
## [1] FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE FALSE FALS
which(x > 10)
## [1] 5 6 7
x[x>10]
## [1] 15 20 50
x[!x>10]
## [1]  1  3  4 10  1  6
x[x<=10]
## [1]  1  3  4 10  1  6
x[x>10 & x<30]
## [1] 15 20
```

Functions

```
square <- function(x) x^2
square(2)
## [1] 4

pow <- function(x, p=2) x^p
pow(10)
## [1] 100
pow(10,3)
## [1] 1000
pow(p=3,10)
## [1] 1000
```

Data Frames

- Data frames are the fundamental structure used in data analysis
- Similar to a database table in spirit (named columns, distinct types)

```
d <- data.frame(x=1:6, y="AUDUSD", z=c("one","two"))
d
```

```
##   x      y   z
## 1 1 AUDUSD one
## 2 2 AUDUSD two
## 3 3 AUDUSD one
## 4 4 AUDUSD two
## 5 5 AUDUSD one
## 6 6 AUDUSD two
```

Lists

```
d <- data.frame(x=1:6, y="AUDUSD", z=c("one","two"))
e <- data.frame(x=1:4, y="Center", z=c("one","two"))
f <- c(1, 2, 3)

g <- list(d, e,f)
f[[3]]
```

```
## [1] 3
```

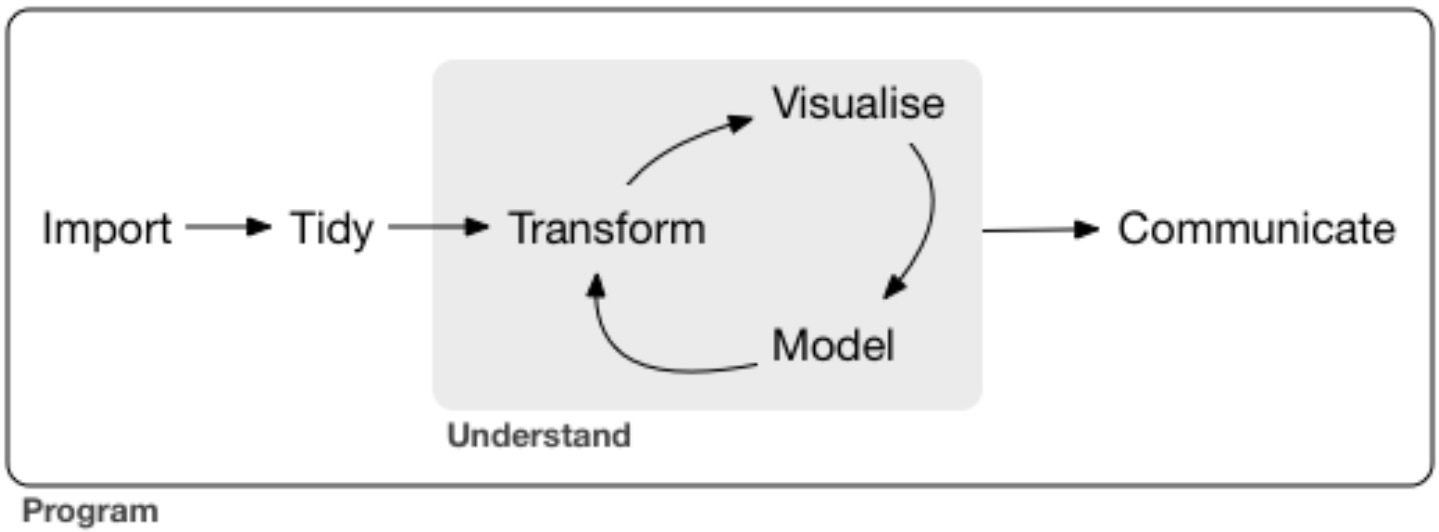
Installing Packages

There are some functions to make easier the management the information into R or to make a particular statistical method

```
install.packages('name')
```

World of Tidyverse

Why use tidyverse package



Untidy Data

See the paper Tidy Data by Hadley Wickham in Journal of Statistical Software (2014)

- Each variable forms a column

Untidy Data

See the paper Tidy Data by Hadley Wickham in Journal of Statistical Software (2014)

- Each variable forms a column
- Each observation forms a row

Untidy Data

See the paper Tidy Data by Hadley Wickham in Journal of Statistical Software (2014)

- Each variable forms a column
- Each observation forms a row
- Each type of observational unit forms a table

Tidy Data

Table 2: Example of common untidy data

Station	Tmax.2014	Tmax.2015	Tmin.2014	Tmin.2015	Prec.2014	Prec.2015
1	32	33	25	26	0	200
2	28	26	19	20	164	0
3	19	18	12	14	0	10

Tidy Data

Table 3: Resulting tidy data set

Station	variable	year	Value
1	Tmax	2014	32
2	Tmax	2014	28
3	Tmax	2014	19
1	Tmax	2015	33
2	Tmax	2015	26
3	Tmax	2015	18
1	Tmin	2014	25
2	Tmin	2014	19
3	Tmin	2014	12
1	Tmin	2015	26
2	Tmin	2015	20
3	Tmin	2015	14
1	Prec	2014	0

Installing Tidyverse

```
install.packages('tidyverse')
```

Loading Packages

```
library('tidyverse')
```

Working with Tidyverse

```
library('tidyverse')

x <- read_csv(file = 'data/weather.csv')
select(x, origin, temp)
select(x, origin, humid)
select(x, year, month, day, temp)
```

Working with Tidyverse

```
filter(x, year == 2013)
filter(x, origin == 'EWR')
filter(x, origin == 'JFK')
filter(x, origin == 'JFK', temp >= 38, humid < 55)
```

Slide with R Output

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

Slide with Plot