

Manipulating Data in R

John Muschelli

January 4, 2016

Overview

You can read data into R using *read.csv*. In this module, we will show you how to:

1. Select specific elements of an object by an index or logical condition
2. Subset rows of a `data.frame`
3. Subset columns of a `data.frame`
4. Add new columns to a `data.frame`
5. Order the rows of a `data.frame`

Setup

We will show you how to do each operation in base R then show you how to use the `dplyr` package to do the same operation (if applicable).

Many resources on how to use `dplyr` exist and are straightforward:

- ▶ <https://cran.rstudio.com/web/packages/dplyr/vignettes/>
- ▶ https://stat545-ubc.github.io/block009_dplyr-intro.html
- ▶ <https://www.datacamp.com/courses/dplyr-data-manipulation-r-tutorial>

Select specific elements using an index

Often you only want to look at subsets of a data set at any given time. As a review, elements of an R object are selected using the brackets ([and]).

For example, `x` is a vector of numbers and we can select the second element of `x` using the brackets and an index (2):

```
x = c(1, 4, 2, 8, 10)
x[2]
```

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

Select specific elements using an index

We can select the fifth or second AND fifth elements below:

```
x = c(1, 2, 4, 8, 10)  
x[5]
```

```
## [1] 10
```

```
x[c(2,5)]
```

```
## [1] 2 10
```

Subsetting by deletion of entries

You can put a minus (-) before integers inside brackets to remove these indices from the data.

```
x[-2] # all but the second
```

```
## [1] 1 4 8 10
```

Note that you have to be careful with this syntax when dropping more than 1 element:

```
x[-c(1,2,3)] # drop first 3
```

```
## [1] 8 10
```

```
# x[-1:3] # shorthand. R sees as -1 to 3  
x[-(1:3)] # needs parentheses
```

```
## [1] 8 10
```

Select specific elements using logical operators

What about selecting rows based on the values of two variables?
We use logical statements. Here we select only elements of `x` greater than 2:

```
x
```

```
## [1] 1 2 4 8 10
```

```
x > 2
```

```
## [1] FALSE FALSE TRUE TRUE TRUE
```

```
x[ x > 2 ]
```

```
## [1] 4 8 10
```

Select specific elements using logical operators

You can have multiple logical conditions using the following:

- ▶ `&` : AND
- ▶ `|` : OR

```
x[ x > 2 & x < 5 ]
```

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

```
x[ x > 5 | x == 2 ]
```

```
## [1] 2 8 10
```