My R book

Introduction

In this book, I document some of the common uses of R as a data analysis tool, starting from basics.

Some resources I refer to:

General

- R Markdown: The Definitive Guide - probably the most important of all, for a start
- Cookbook for R
- My favorite RStudio tips and tricks
- Samuel Chan

Linear Algebra

- Personality Project
- R-bloggers

Statistics

- R-tutor
- SimplyR(College of Staten Island)
- Statsmethod

Data Science

- R for Data Science
- dplyr cheat sheet
- TSKam on RPubs
- Naimish Agarwal on RPubs

I have been experiencing compilation errors with Knitr when plots were involved. As such, I will cover visualization in a follow-up. As there is LaTeX support in R Markdown, I would also consider creating worksheets using RStudio in the future.

> "There is no end to education. It is not that you read a book, pass an examination, and finish with education. The whole of life, from the moment you are born to the moment you die, is a process of learning."
>
> —Jiddu Krishnamurti

Basic data structures

Vectors

Our first data structure is the vector, which is a collection of data of the same class. In R, there are two classes of numerical data (`integer` and `numeric`).

```r
# integer
a <- 3:8
a
```

```
## [1] 3 4 5 6 7 8
```

```r
class(a)
```

```
## [1] "integer"
```

```r
length(a)
```

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

```r
# numeric
b <- c(3.4,5.6,7.8)
class(b)
```

```
## [1] "numeric"
```

The `character` class stores strings.

```r
c <- c("KFC","MacDonald's","Popeyes","Gong Cha")
c[2:3]
```

```
## [1] "MacDonald's" "Popeyes"
```

Notice that c[-3] **drops** the third entry *Popeyes*, unlike in Python where the third entry would be indexed instead.

```r
c[-3]
```

```
## [1] "KFC"         "MacDonald's" "Gong Cha"
```

We can use logical operators to create boolean masks.

```r
a%%3!=0
```

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

And just like in Python, we can use boolean masks to index vectors.

```r
# remove multiples of 3
a[a%%3!=0]
```

```
## [1] 4 5 7 8
```

```r
rep(factor(LETTERS[3:6]), 5)
```

```
##  [1] C D E F C D E F C D E F C D E F C D E F
## Levels: C D E F
```

```r
# Stock chart look-alike
a <- cumsum(rnorm(1000))+80
# plot(a,type = "l")
# hist(a)
```

Vectors can be used in many ways.

Dataframes

Our first data structure is the vector, which is a collection of data of the same class. In R, there are two classes of numerical data (`integer` and `numeric`).

```r
# integer
a <- 3:8
a
```

```
## [1] 3 4 5 6 7 8
```

```r
class(a)
```

```
## [1] "integer"
```

```r
length(a)
```

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

```r
# numeric
b <- c(3.4,5.6,7.8)
class(b)
```

```
## [1] "numeric"
```