How to Use strsplit Function

Yai

2/11/2019

`strsplit` Command and How to use it.

`strsplit` command is the main command we use to split the string variable based on a specific separator. However, it can be troublesome to use it effectively since the output from the command will always return list object and list object can be a little bit problematic to deal with, especially for those who never have experience in programming before. Here's take a look a the outputs of `strsplit`

```
data1 <- data.frame(name = c("Mary", "Michael", "Jane", "John"),
                    birthday = c("01/01/2000", "10/22/1988","09/18/1999","03/02/1995"),
                    stringsAsFactors = FALSE)
data1
```

```
##      name   birthday
## 1    Mary 01/01/2000
## 2 Michael 10/22/1988
## 3    Jane 09/18/1999
## 4    John 03/02/1995
```

We have a data frame that contain name and birthday. Now, our task will be to extract the birthday column into day, month, and year. We can see that the value of birthday is structured by usinf "/" as a separator. In this case, we can use "/" as a separator for `strsplit` command to divide the information.

```
birthday_split <- strsplit(data1$birthday, "/")
birthday_split
```

```
## [[1]]
## [1] "01"   "01"   "2000"
##
## [[2]]
## [1] "10"   "22"   "1988"
##
## [[3]]
## [1] "09"   "18"   "1999"
##
## [[4]]
## [1] "03"   "02"   "1995"
```

You can see the outputs from the command. It is in the `list` object and it can be troublesome to deal with.

```
birthday_split[1]
```

```
## [[1]]
## [1] "01"   "01"   "2000"
```

If you want to only copy the first element of the list which contains the value of month and we index by using `[1]`, or `[[1]]`, instead of getting the first columns, we will get the information from the first row.

In this case, we will use `sapply` command to apply the command (index command) to the whole list. We will talk about the `lapply`, and `sapply` command later in this semester but for `strsplit` you can copy and modify these commands to use in your project.

```
sapply(strsplit_output, `[`, 1)
```

You subtitue `strsplit_output` with your own outputs or you can use `strsplit` command inside this structure. Then, you replace the value 1 with the column index in the list that you want to extract. Here's example.

```
## save the ouputs to another object before append to the data frame
birthday_split <- strsplit(data1$birthday, "/")
data1$month <- sapply(birthday_split, `[`, 1)
data1
```

```
##      name   birthday month
## 1    Mary 01/01/2000    01
## 2 Michael 10/22/1988    10
## 3    Jane 09/18/1999    09
## 4    John 03/02/1995    03
```

```
## split the string and then append immediately
data1$day <- sapply(strsplit(data1$birthday, "/"), `[`, 2)
data1$year <- sapply(strsplit(data1$birthday, "/"), `[`, 3)
data1
```

```
##      name   birthday month day year
## 1    Mary 01/01/2000    01  01 2000
## 2 Michael 10/22/1988    10  22 1988
## 3    Jane 09/18/1999    09  18 1999
## 4    John 03/02/1995    03  02 1995
```