

# Select function

*Pema Lama*

*April 3, 2017*

In this session, we will explore the `select()` function of `dplyr` package.

How do I load a R-package using `library()` function?

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##   filter, lag
##
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

How do I set a current working directory using `setwd()` function?

```
setwd("C:/Users/Owner/Desktop/dplyr")
```

How do I load a data set using `read.csv()` function?

```
titanic <- read.csv('titanic.csv', stringsAsFactors = TRUE, header = TRUE)
```

How do I see the structure of a data set with `str()` function?

```
str(titanic)
```

```
## 'data.frame':   1310 obs. of  14 variables:
## $ pclass   : int   1 1 1 1 1 1 1 1 1 1 ...
## $ survived : int   1 1 0 0 0 1 1 0 1 0 ...
## $ name      : Factor w/ 1308 levels "", "Abbing, Mr. Anthony",...: 23 25 26 27 28 32 47 48 52 56 ...
## $ sex       : Factor w/ 3 levels "", "female", "male": 2 3 2 3 2 3 2 3 2 3 ...
## $ age       : num   29 0.917 2 30 25 ...
## $ sibsp     : int   0 1 1 1 1 0 1 0 2 0 ...
## $ parch     : int   0 2 2 2 2 0 0 0 0 0 ...
## $ ticket    : Factor w/ 930 levels "", "110152", "110413",...: 189 51 51 51 51 126 94 17 78 827 ...
## $ fare      : num   211 152 152 152 152 ...
## $ cabin     : Factor w/ 187 levels "", "A10", "A11",...: 45 81 81 81 81 151 147 17 63 1 ...
## $ embarked : Factor w/ 4 levels "", "C", "Q", "S": 4 4 4 4 4 4 4 4 2 ...
## $ boat      : Factor w/ 28 levels "", "1", "10", "11",...: 13 4 1 1 1 14 3 1 28 1 ...
## $ body      : int   NA NA NA 135 NA NA NA NA NA 22 ...
## $ home.dest : Factor w/ 370 levels "", "?Havana, Cuba",...: 310 232 232 232 232 238 163 25 23 230 ...
```

How do I see the column names of the data set?

We can use `names()` function to see the column names of a data set.

```
names(titanic)
```

```
## [1] "pclass" "survived" "name" "sex" "age"
## [6] "sibsp" "parch" "ticket" "fare" "cabin"
```

```
## [11] "embarked" "boat" "body" "home.dest"
```

How do I slice out some of the columns from a data set?

We use select() function to slice out subsets of columns from a data set.

```
subtitanic <- select(titanic, pclass, survived, sex, sibsp)
```

How do I see the column names of the slice data set?

```
names(subtitanic)
```

```
## [1] "pclass" "survived" "sex" "sibsp"
```

How do I change the order of the columns of a sliced data set?

We can change the order of the columns of a data set by selecting column names order which I want.

```
subtitanic = select(titanic, name, age, pclass, body)
```

How do I see column names of the slice data set?

```
names(subtitanic)
```

```
## [1] "name" "age" "pclass" "body"
```

How do I drop columns in a data set?

We drop columns simply adding minus sign front of variables in the select() function.

```
subtitanic <- select(titanic, -name, -body)
```

```
names(subtitanic)
```

```
## [1] "pclass" "survived" "sex" "age" "sibsp"
## [6] "parch" "ticket" "fare" "cabin" "embarked"
## [11] "boat" "home.dest"
```

How do I drop a group of columns from a data set?

We drop group of columns simply adding minus sign front of vector, -c(), in the select() function.

```
subtitanic <- select(titanic, -c(name, age, sibsp, body, parch))
```

```
names(subtitanic)
```

```
## [1] "pclass" "survived" "sex" "ticket" "fare" "cabin"
## [7] "embarked" "boat" "home.dest"
```

How do I slice out subsets of columns from a data set using (:)?

```
subtitanic <- select(titanic, pclass:age)
```

```
names(subtitanic)
```

```
## [1] "pclass" "survived" "name" "sex" "age"
```

How do I drop multiple columns from a data set?

```
subtitanic <- select(titanic, -c(pclass:age))
```

```
names(subtitanic)
```

```
## [1] "sibsp" "parch" "ticket" "fare" "cabin" "embarked"
## [7] "boat" "body" "home.dest"
```

How do I select columns with index values?

```
subtitanic <- select(titanic, 1, 2, 3, 4)
```

```
names(subtitanic)
```

```
## [1] "pclass" "survived" "name" "sex"
```

How do I drop columns using index values?

```
subtitanic <- select(titanic, -1, -2, -3, -4)
names(subtitanic)
```

```
## [1] "age"      "sibsp"    "parch"    "ticket"    "fare"
## [6] "cabin"    "embarked" "boat"     "body"      "home.dest"
```

How do I drop a group columns using index values?

```
subtitanic <- select(titanic, -c(1, 2, 3, 4))
names(subtitanic)
```

```
## [1] "age"      "sibsp"    "parch"    "ticket"    "fare"
## [6] "cabin"    "embarked" "boat"     "body"      "home.dest"
```

How do I change a column names of data set?

```
subtitanic <- select(titanic, 'gender' = sex, 1:10)
names(subtitanic)
```

```
## [1] "gender"   "pclass"   "survived" "name"      "age"      "sibsp"
## [7] "parch"    "ticket"   "fare"     "cabin"
```

How do I see the first few rows of a data set?

```
head(subtitanic)
```

```
##   gender pclass survived          name
## 1 female     1         1      Allen, Miss. Elisabeth Walton
## 2  male     1         1      Allison, Master. Hudson Trevor
## 3 female     1         0      Allison, Miss. Helen Loraine
## 4  male     1         0      Allison, Mr. Hudson Joshua Creighton
## 5 female     1         0 Allison, Mrs. Hudson J C (Bessie Waldo Daniels)
## 6  male     1         1      Anderson, Mr. Harry

##   age sibsp parch ticket   fare  cabin
## 1 29.0000    0     0  24160 211.3375    B5
## 2  0.9167    1     2  113781 151.5500 C22 C26
## 3  2.0000    1     2  113781 151.5500 C22 C26
## 4 30.0000    1     2  113781 151.5500 C22 C26
## 5 25.0000    1     2  113781 151.5500 C22 C26
## 6 48.0000    0     0  19952  26.5500   E12
```