

RStudio

File Edit Code View Plots Session Build Debug Tools Help

Go to file/function Addins

Console

```
> summarize(cran, avg_bytes = mean(size))
# A tibble: 1
  avg_bytes
  <dbl>
1 844086.5
```

| Your dedication is inspiring!

| That's not particularly interesting. `summarize()` is most useful when working with data that has been grouped by the values of a particular variable.

... | 95%

| We'll look at grouped data in the next lesson, but the idea is that `summarize()` can give you the requested value FOR EACH group in your dataset.

... | 97%

| In this lesson, you learned how to manipulate data using dplyr's five main functions. In the next lesson, we'll look at how to take advantage of some other useful features of dplyr to make your life as a data analyst much easier.

... | 98%

| would you like to receive credit for completing this course on Coursera.org?

1: Yes  
2: NO

Selection:

Environment History

Global Environment

Data

Values

path2csv "C:/Users/EILEEN/Documents/R/win-library/3.3/..."

Files Plots Packages Help Viewer

R: Relational Operators

At least one of x and y must be an atomic vector, but if the other is a list it attempts to coerce it to the type of the atomic vector: this will succeed if the list is made up of elements of length one that can be coerced to the correct type.

If the two arguments are atomic vectors of different types, one is coerced to the type of the other, the (decreasing) order of precedence being character, complex, numeric, integer, logical and raw.

Missing values (NA) and NaN values are regarded as non-comparable even to themselves, so comparisons involving them will always result in NA. Missing values can also result when character strings are compared and one is not valid in the current collation locale.

Language objects such as symbols and calls are deparsed to character strings before comparison.

Value

A logical vector indicating the result of the element by element comparison. The elements of shorter vectors are recycled as necessary.

Objects such as arrays or time-series can be compared this way provided they are

Pregúntame cualquier cosa

ESP 09:00 p. m.  
ES 03/11/2016

RStudio

File Edit Code View Plots Session Build Debug Tools Help

Go to file/function Addins

summarize3.R x summarize4.R x result3 x chain1.R x chain2.R x chain3.R x chain4.R x

Run Source

12:8 (Top Level)

R Script

```
1 # arrange() the result by size_mb, in descending order.
2 # If you use the result in the console, you
3 # print the result. If you chain the
4 # result, you don't need to print it.
5
6 cran %>%
7   select(id, country, package, size) %>%
8   mutate(size_mb = size / 2^20) %>%
9   filter(size_mb <= 0.5) %>%
10  arrange(desc(size_mb)) %>%
11  print()
```

| Keep up the great work!

| In this lesson, you learned about grouping and chaining using dplyr. You combined some of the things you learned in the previous lesson with these more advanced ideas to produce concise, readable, and highly effective code. welcome to the wonderful world of dplyr!

... | 98%

| would you like to receive credit for completing this course on Coursera.org?

1: No  
2: Yes

Selection:

Environment History

Global Environment

cran 225468 obs. of 11 variables

chain1 225468 obs. of 11 variables

chain2 225468 obs. of 11 variables

chain3 6023 obs. of 5 variables

chain4 46 obs. of 5 variables

result1 46 obs. of 5 variables

result2 46 obs. of 5 variables

result3 46 obs. of 5 variables

Files Plots Packages Help Viewer

filter

Help on topic 'filter' was found in the following packages:

[Return rows with matching conditions.](#)  
(in package `dplyr` in library C:/Users/EILEEN/Documents/R/win-library/3.3)

[Linear Filtering on a Time Series](#)  
(in package `stats` in library C:/Program Files/R/R-3.3.1/library)

Pregúntame cualquier cosa

ESP 07:52 p. m.  
ES 05/11/2016

The screenshot shows the RStudio interface with a script editor, console, and environment pane. A large orange watermark "EILEEN FLORES ROMERO" is overlaid on the script editor.

**Script Editor:**

```

6 # 2. use mutate to add two new columns, whose values will be
7 # automatically computed group-by-group:
8 #
9 # * total = sum(count)
10 # * prop = count / total
11 #
12 sat %>%
13   select(count = tot, prop = prop) %>%
14   gather(part_sex, count, tot, prop) %>%
15   separate(part_sex, c("part", "sex")) %>%
16   group_by(part, sex) %>%
17   mutate(total = sum(count), prop = count / total) %>%
18   print
19
20 # ... with 26 more rows

```

**Console:**

```

9 500-590 read fem 259553 883955 0.29362694
10 400-490 read fem 296793 883955 0.33575578
# ... with 26 more rows

|=====| 98%

| In this lesson, you learned how to tidy data with tidyr and dplyr. These tools will help
| you spend less time and energy getting your data ready to analyze and more time actually
| analyzing it.

...

|=====| 100%

| would you like to receive credit for completing this course on Coursera.org?
1: Yes
2: No

```

**Environment Pane:**

Variable	Obs.	Vars.
failed	6	4
failed	6	4
auto	1	6
ck	12	6
passed	4	4
res	20	3
result1	46	5

**Files Pane:**

R: group a tbl by one or more variables. Find in Topic

group\_by (dplyr) R Documentation

**Description**

Most data operations are useful done on groups defined by variables in the dataset. The `group_by` function takes an existing tbl and converts it into a grouped tbl where operations are performed "by group".

**Usage**

```
group_by(.data, ..., add = FALSE)
```

**Arguments**

```
group_by(.data, ..., .dots, add = FALSE)
```

The screenshot shows the RStudio interface. The top pane displays a message from Coursera congratulating the user on completing the course. The bottom pane shows the R console output, which includes a large 'Great job!' message, a progress bar at 95%, and a confirmation of the user's selection (No). The R console also shows the execution of `stopwatch()` and the resulting time taken to complete the course.