9/19/2018 Vectorization | Coursera

A Back to Week 1 X Lessons This Course: Advanced R Programming
Prev Next

One final aspect of this function that is worth noting is that as currently written it is not *vectorized*. This means that each argument must be a single value---a single package name and a single date. However, in R, it is a common paradigm for functions to take vector arguments and for those functions to return vector or list results. Often, users are bitten by unexpected behavior because a function is assumed to be vectorized when it is not.

One way to vectorize this function is to allow the pkgname argument to be a character vector of package names. This way we can get download statistics for multiple packages with a single function call. Luckily, this is fairly straightforward to do. The two things we need to do are

- 1. Adjust our call to filter() to grab rows of the data frame that fall within a vector of package names
- 2. Use a group_by() %>% summarize() combination to count the downloads for each package.

Now we can call the following

Note that the output of num_download() has changed. While it previously returned an integer vector, the vectorized function returns a data frame. If you are authoring a function that is used by many people, it is usually wise to give them some warning before changing the nature of the output.

Vectorizing the date argument is similarly possible, but it has the added complication that for each date you need to download another log file. We leave this as an exercise for the reader.

Mark as completed







9/19/2018 Vectorization | Coursera