

HW6_Hye Soo Choi

Hye Soo Choi

July 23, 2015

Motivation

Most R packages live in the Comprehensive R Archive Network better known as CRAN. Many different versions of those packages have been submitted to CRAN. Through this report, we hope to analyze how often the packages that we has used most frequently have been updated and how the size of each packages has changed.

Source the functions (assuming the file is in working directory):

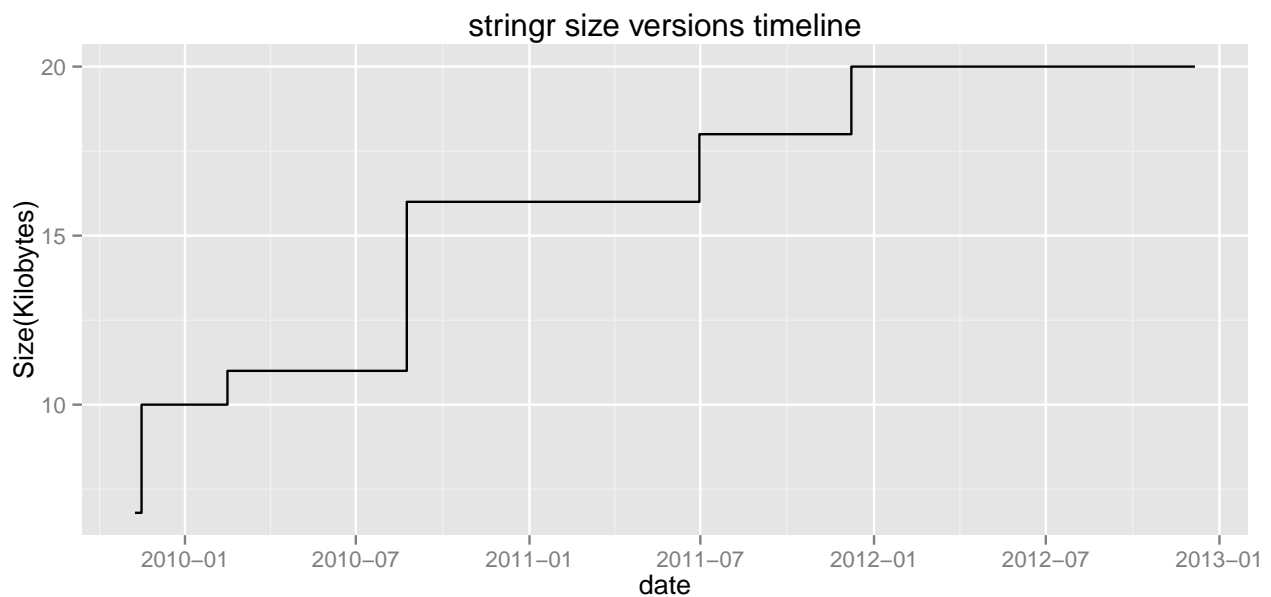
```
source('./functions.R')
```

```
## Loading required package: bitops
```

Stringr package

Test code with the package 'stringr'; call `get_archive_table()` and `ggstep()`. This shows a graph of package size and update dates of package 'stringr' since it first appeared.

```
stringr_df <- get_archive_table('stringr')  
ggstep(stringr_df)
```



Other packages : XML, RCurl, ggplot2, boot

Looking at various packages:

Besides getting a graph about “stringr”, get the CRAN archive tables for the packages “XML”, “RCurl”, “ggplot2” and “boot”. Combine all the data tables in one single data frame.

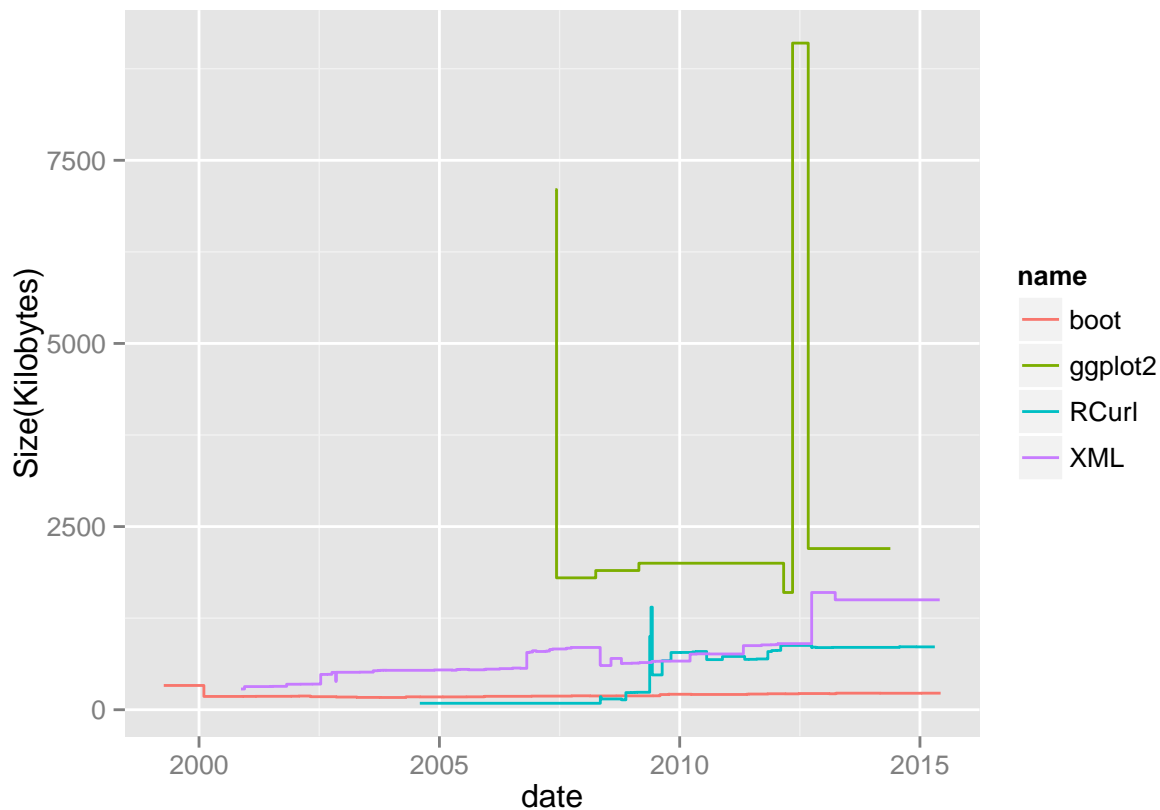
```
# get archive tables
XML_df <- get_archive_table('XML')
RCurl_df <- get_archive_table('RCurl')
ggplot2_df <- get_archive_table('ggplot2')
boot_df <- get_archive_table('boot')

# combine all the data frames
all_df <- rbind(boot_df, ggplot2_df, RCurl_df, XML_df)
```

Graph

Plot A: all packages in one single frame These are graphs of sizes and update dates of 4 packages in one single frame.

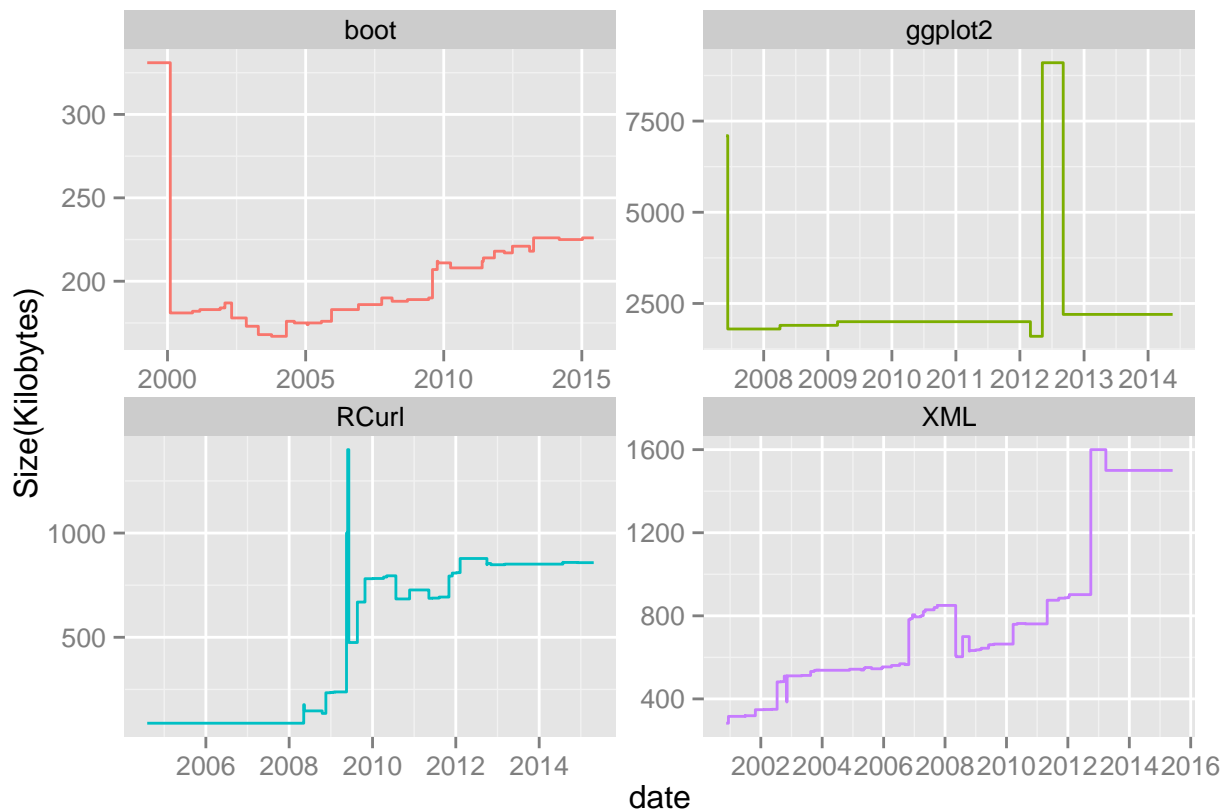
```
ggplot(all_df, aes( x= date, y = size )) +
  geom_step(aes(col = name)) +
  ylab ('Size(Kilobytes)')
```



Plot B: one package per facet

These are graphs of sizes and update dates of 4 packages, one package per facet with different scale.

```
ggplot(all_df, aes( x= date, y = size )) +  
  geom_step(aes(col = name)) +  
  facet_wrap( ~ name, scales = "free") +  
  ylab ('Size(Kilobytes)') +  
  theme(legend.position="none")
```



Analysis

It is Boot that was first invented out of these 4 packages in 1999. Boot almost has kept the smallest package size, within the range 0-350 kilobytes, and although since 2001 the size itself has maintained increasing tendency over time, the size variation has been ignorably small compared to other packages. In contrast, ggplot2 was last invented in 2009 and has kept the largest size, within the range 1250-9500 kilobytes, since its birth. Unprecedented drastic increase of around 7000 kilobytes in the package size has occurred in 2012, but in few months a newer version came out. RCurl was third invented with the smallest initial size in 2004 and also had two unprecedented drastic increases around 2009 and then a newer version of smaller size came out in a few months. The range of RCurl package size is in between 0 and 1500 kilobytes. XML was second invented in 2000 and it is observed that only minute size changes has made before 2012. In 2012 and late 2006, the size drastically increased by 700 and 200 kilobytes respectively. After the first drastic increase, a newer version with smaller size came out in a short time, while after the second drastic increase the increased size has been still persistent up to now. The range of XML package size is in between 200 and 1600 kilobytes.