



# IF YOU CAN'T BEAT 'EM ...

## SOME COMMENTS ON USING R ALONG WITH OTHER LANGUAGES

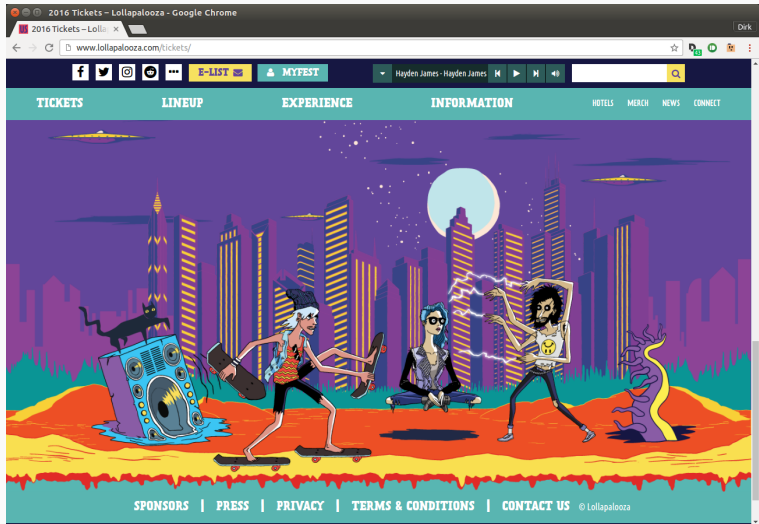
---

Dirk Eddebuettel

31 July 2016

Open Source Statistical Software for Data Science – Invited Papers  
Joint Statistical Meetings (JSM), Chicago, IL

# WELCOME TO LOLLAPALOOZA!



## Content

- Single- or Multi-Language ?
- Interlude
- Empirics
- Illustration
- Conclusion

# SINGLE- OR MULTI-LANGUAGE ?

---

### Better with more than one?

- No one language fits all
- Real-world projects are frequently multi-language
- See *e.g.* job ads which rarely ever list just one language

### Or better with just one?

- Mental switching cost between languages? Possibly
- Interop difficult and less portable? Maybe, but that is an argument against weak systems / OSs
- Easier / less to learn?
- “More hoops” to code?

## MENTAL SWITCHING COSTS?



**Dirk Eddelbuettel**

@eddelbuettel

Forcing us to alternate between comment characters %, # and // may have been the biggest trick ever pulled by the Devil.

RETWEETS

6

LIKES

5



10:22 AM - 10 Aug 2014



6



5



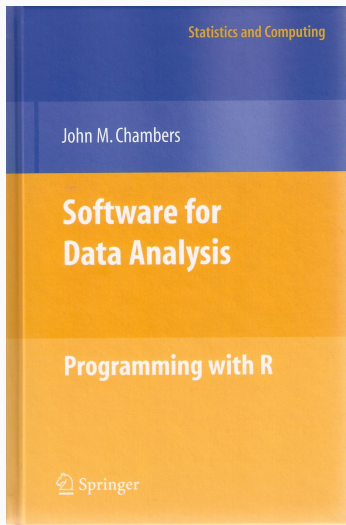
## Open Question

- Hard to measure or test: Any empirics on real world projects?
- Code competition / comparisons (*e.g.* Project Euler): Are they realistic?



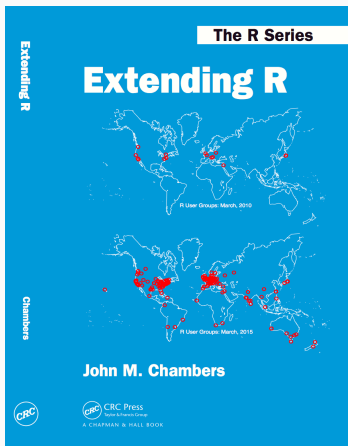
# INTERLUDE

---



## Chambers (2008) Software For Data Analysis

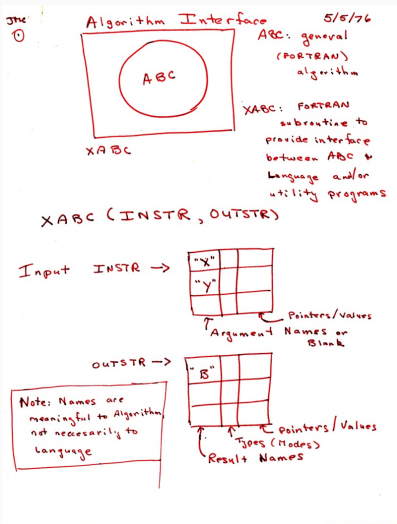
Chapters 10 and 11 devoted to *Interfaces I: C and Fortran* and *Interfaces II: Other Systems*.



Chambers (2016) Extending R  
An entire book about this with  
*concrete* Python, Julia and C++  
code and examples

## Chambers 2016, Chapter 4

*The fundamental lesson about programming in the large is that requires a correspondingly broad and flexible response. In particular, no single language or software system is likely to be ideal for all aspects. Interfacing multiple systems is the essence. Part IV explores the design of interfaces from R.*

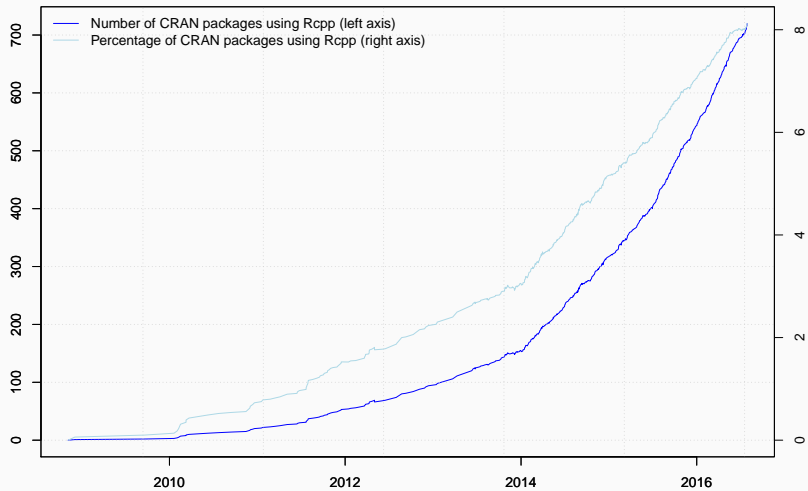


Thanks to John Chambers for a scanned copy of this historic sketch.

# EMPIRICS

---

## Growth of Rcpp usage on CRAN



```
library(pagerank) # github.com/andrie/pagerank
```

```
cran <- "http://cloud.r-project.org"
```

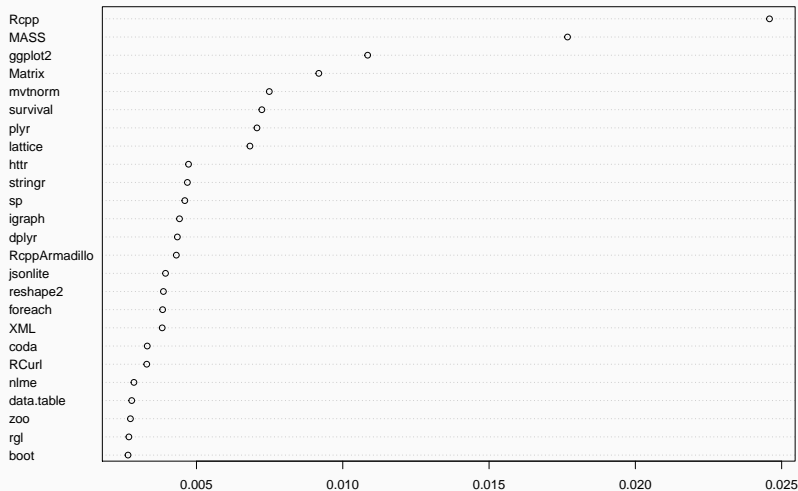
```
pr <- compute_pagerank(cran)
```

```
round(100*pr[1:5], 3)
```

```
##      Rcpp      MASS  ggplot2  Matrix  mvtnorm  
##  2.452  1.771  1.088  0.920  0.749
```



Top 25 of Page Rank as of July 2016





# ILLUSTRATION

---

## Setup

```
py_cflags <- system("python2.7-config --cflags", intern=TRUE)
se <- Sys.setenv; ge <- Sys.getenv # shorthands to typeset
se("PKG_CFLAGS"=sprintf("%s %s", ge("PKG_CFLAGS"), py_cflags))
se("PKG_CXXFLAGS"=sprintf("%s %s", ge("PKG_CXXFLAGS"), py_cflags))
py_ldflags <- system("python2.7-config --ldflags", intern=TRUE)
se("PKG_LIBS"=sprintf("%s %s %s", ge("PKG_CFLAGS"),
                      "-lboost_python-py27", py_ldflags))
```

## USING R TO C++ TO BOOST TO PYTHON, AND BACK

```
#include <Rcpp.h>
#include <Python.h>

// [[Rcpp::export]]
void initialize_python() {
    Py_SetProgramName(""); /* optional but recommended */
    Py_Initialize();
}

// [[Rcpp::export]]
void hello_python() {
    PyRun_SimpleString("from time import time,ctime\n"
                      "print 'Today is',ctime(time())\n");
}
```

Hello, World: Called from R

```
initialize_python()  
hello_python()
```

```
## Today is Sat Jul 30 13:38:01 2016
```

More at <http://gallery.rcpp.org/articles/rcpp-python/>

Disclaimer: *For illustration purposes.* Works as designed on Ubuntu. Not meant to be universally portable to all three OSs.

## CONCLUSION

---

## Mixing Languages

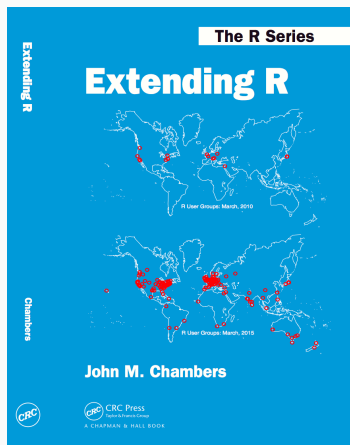
- Common
- Natural
- Unavoidable



## Consequences

- Must make it easier to interoperate
- Stop bickering among ourselves
- Build systems that are larger than the sum of their parts

## Just Do It



### Lars Wirzenius “Which license is the most free?”

Free software licences can be roughly grouped into permissive and copyleft ones. [...] A permissive licence lets you do things that a copyleft one forbids, so **clearly the permissive licence is more free**. A copyleft licence means software using it won't ever become non-free against the wills of the copyright holders, so **clearly a copyleft licence is more free than a permissive one**.

Both sides are both right and wrong, of course, which is why this argument will continue forever. [...]

If a discussion about the relative freedom of licence types becomes heated, step away. It's not worth participating anymore.

<http://yakking.branchable.com/posts/comparative-freeness/>