Potential risks in running R on the cloud

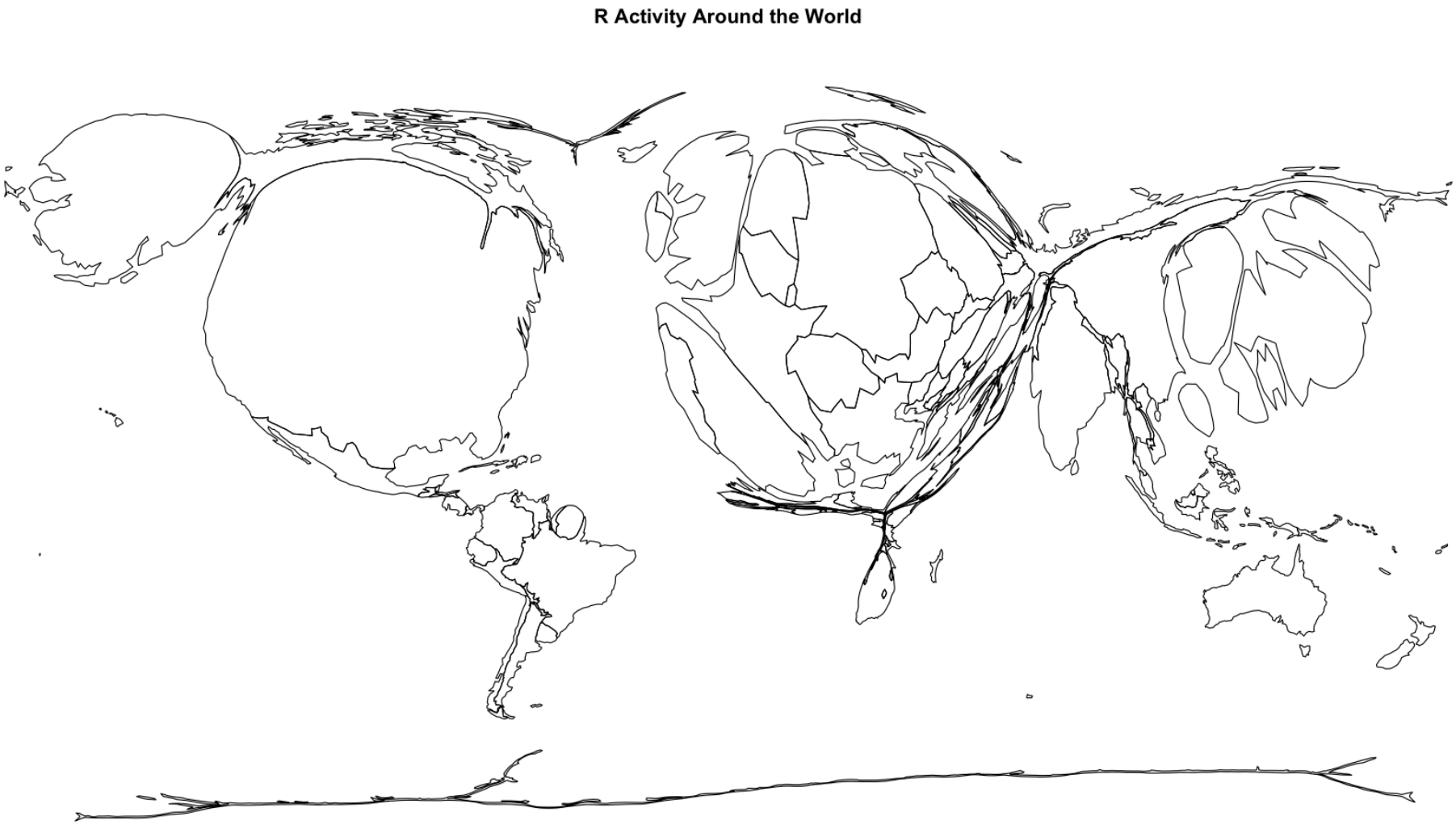
"This means it's the equivalent of a chainsaw: there are ways to use it safely." – rserve.js

Gergely Daróczi
daroczig@rapporter.net

13 January 2014

What is R?

The "lingua franca" of statistical analysis



R in the cloud

Full access to the servers right in your browser

```
> system('uname -a')
Linux nevermind 3.12.6-1-ck #1 SMP PREEMPT Fri Dec 20 14:27:18 EST 2

> readLines(pipe('whoami'))
[1] "daroczig"

> length(list.files('/etc'))
[1] 253

> cat('Hello, world!', file = '/tmp/foobar.sh')
> readLines('/tmp/foobar.sh')
[1] "Hello, world!"
Warning message:
In readLines("/tmp/foobar.sh") :
  incomplete final line found on '/tmp/foobar.sh'

> while (TRUE) fork()
```

Potential solutions

Workarounds, limitations, ongoing discussion

**"Are you sure our data is secure on the cloud?
I just saw my spreadsheet on the weather channel!"**

- chroot
- Renjin
- RAppArmor
- OpenCPU
- *sandboxR*
- custom solutions

Demo

```
R version 3.0.2 (2013-09-25) -- "Frisbee Sailing"
Copyright (C) 2013 The R Foundation for Statistical Computing
Platform: x86_64-unknown-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

>
```

sandboxR

Filtering "malicious" calls in R

```
> getParseData(parse(text = 'foo <- get("system")("whoami")'))
   line1 col1 line2 col2 id parent                token terminal     text
21     1    1     1   30 21      0                 expr    FALSE
1      1    1     1    3  1      3               SYMBOL     TRUE      foo
3      1    1     1    3  3     21                 expr    FALSE
2      1    5     1    6  2     21          LEFT_ASSIGN     TRUE       <-
19     1    8     1   30 19     21                 expr    FALSE
12     1    8     1   20 12     19                 expr    FALSE
4      1    8     1   10  4      6 SYMBOL_FUNCTION_CALL     TRUE      get
6      1    8     1   10  6     12                 expr    FALSE
5      1   11     1   11  5     12                  '('     TRUE        (
7      1   12     1   19  7      9           STR_CONST     TRUE "system"
9      1   12     1   19  9     12                 expr    FALSE
8      1   20     1   20  8     12                  ')'     TRUE        )
13     1   21     1   21 13     19                  '('     TRUE        (
14     1   22     1   29 14     16           STR_CONST     TRUE "whoami"
16     1   22     1   29 16     19                 expr    FALSE
15     1   30     1   30 15     19                  ')'     TRUE        )
```

sandboxR in action

http://hackme.rapporter.net

Summary

sandboxR + RAppArmor

	sandboxR	RAppArmor
Advantages	• R-specific rules • user-friendly msgs • can call RAppArmor	• easy to configure • kernel module • great for FS-rules
Disadvantages	• tedious to maintain • not 100% protection	• needs AppArmor • basic network-rules • should use profile instead of hats • eval.secure forks • no R-specific rules

Summary

sandboxR + RAppArmor

	sandboxR	RAppArmor
Advantages	• R-specific rules • user-friendly msgs • **can call RAppArmor**	• easy to configure • kernel module • great for FS-rules
Disadvantages	• tedious to maintain • not 100% protection	• needs AppArmor • basic network-rules • should use profile instead of hats • eval.secure forks • no R-specific rules

Questions?

daroczig@rapporter.net
http://hackme.rapporter.net