

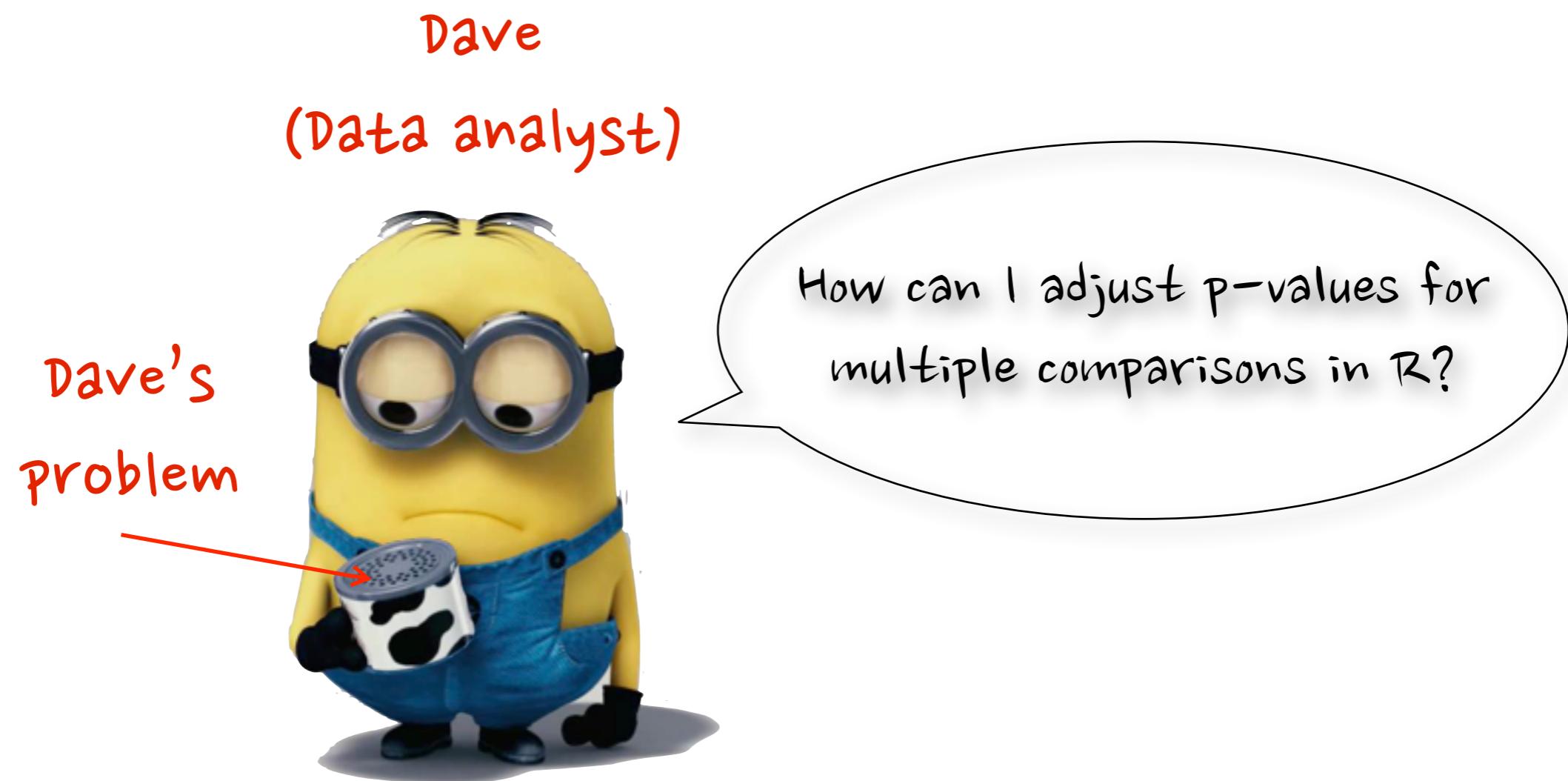
Stack**Exchange**

Bogdan Vasilescu, Alexander Serebrenik  
Eindhoven University of Technology  
@b\_vasilescu, @aserebrenik

Prem Devanbu, Vladimir Filkov  
University of California, Davis  
@devanbu



# User support over the years



@b\_vasilescu



# User support over the years

Dave



Kevin



Ask a colleague



@b\_vasilescu



# User support over the years



Ask a colleague



# User support over the years

**R-bloggers**  
R news and tutorials contributed by (452) R bloggers

[Home](#) | [About](#) | [add your blog!](#) | [Contact us](#) | [RSS](#) | [R Jobs](#)

**WELCOME!**

**About**  
December 1, 2009  
By Tal Galili

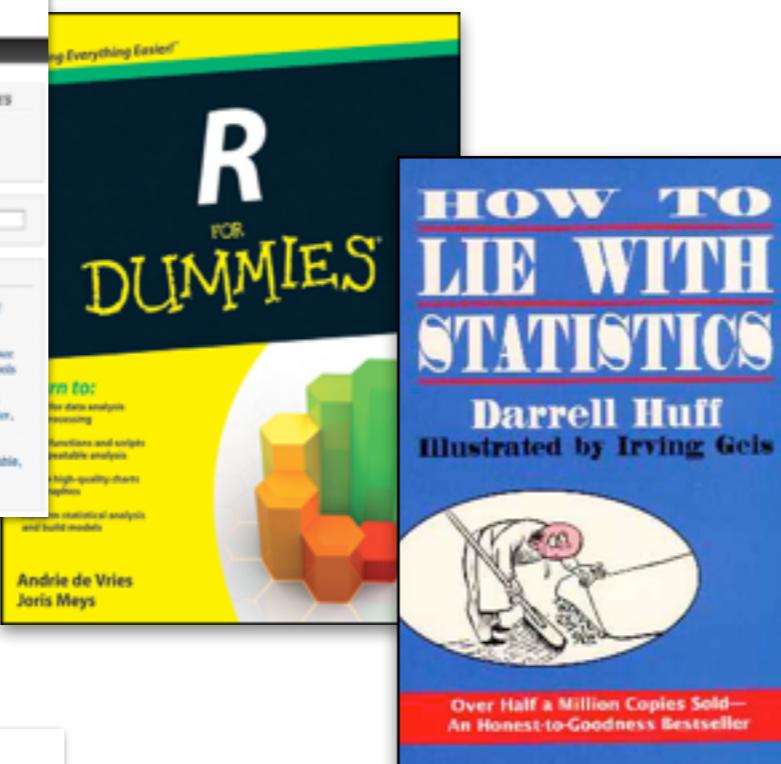
If this is your first time encountering "R": The R language (and open-source software) is the de-facto standard among statisticians for the development of statistical software, and is widely used for statistical software development and data analysis (for more details about R you can read the post "What is R?")

**What is R-Bloggers.com?**  
R-Bloggers.com is a central hub (i.e.: A blog aggregator) of content collected from bloggers who write about R (in English). The site will help R-bloggers and users to connect and follow the "R-bloggerosphere" (you can view a 7 minute talk, from useR2011, for more information about the R-bloggerosphere).

**How does R-Bloggers operate?**  
This site will aggregate feeds (only with permission!) from participating R blogs. The beginnings of each participating blog's posts will automatically be displayed on the main page; inside every post there is a link to the original blog and links to other related articles. All participating blogs will have links in the "Contributors" section of our sidebar.

**R Bloggers**  
16,099 people like R Bloggers.  
 And get updates to your Facebook!

## Blogs



## Books



## R help archive by thread

2674 messages Starting Sat 31 Dec 2011 - 12:34:58 GMT, Ending Tue 31 Jan 2012 - 12:19:21 GMT

This period: [Most recent messages](#)  
sort by: [ [thread](#) ] [ [author](#) ] [ [date](#) ] [ [subject](#) ] [ [attachment](#) ]  
Other periods: [ [Next](#) ] [ [Thread view](#) ] [ [List of Folders](#) ]  
Nearby: [ [About this archive](#) ] [ [Other mail archives](#) ]

- [R] Cross-validation error with tune and with rpart Israel Saeta Pérez (Sat 31 Dec 2011 - 12:34:58 GMT)
  - Re: [R] Cross-validation error with tune and with rpart Prof Brian Ripley (Sat 31 Dec 2011 - 14:13:12 GMT)
  - Re: [R] Is it possible to "right align" text in R graphics? Duncan Mackay (Sun 01 Jan 2012 - 03:34:26 GMT)
    - Re: [R] Is it possible to "right-align" text in R graphics? Tal Galili (Sun 01 Jan 2012 - 11:11:44 GMT)
    - Re: [R] Is it possible to "right-align" text in R graphics? Jeff Newmiller (Sun 01 Jan 2012 - 19:35:01 GMT)
    - Re: [R] Is it possible to "right-align" text in R graphics? Richard M. Heiberger (Sun 01 Jan 2012 - 20:03:20 GMT)
  - Re: [R] Is it possible to "right-align" text in R graphics? Majid Einiain (Tue 03 Jan 2012 - 06:14:45 GMT)
    - Re: [R] Is it possible to "right-align" text in R graphics? Tal Galili (Tue 03 Jan 2012 - 17:39:25 GMT)
    - Re: [R] Is it possible to "right-align" text in R graphics? Majid Einiain (Sun 15 Jan 2012 - 12:55:33 GMT)
  - Re: [R] Is it possible to "right-align" text in R graphics? Tal Galili (Tue 03 Jan 2012 - 22:30:16 GMT)
    - Re: [R] Is it possible to "right-align" text in R graphics? Hadley Wickham (Tue 03 Jan 2012 - 22:51:50 GMT)
    - Re: [R] Is it possible to "right-align" text in R graphics? Michael Weylandt (Wed 04 Jan 2012 - 07:49:43 GMT)
- Re: [R] RGtk2: How to overlay a gtkDrawingArea with a button or any other widget? Michael Lawrence (Mon 16 Jan 2012 - 15:04:02 GMT)
  - [R] question re\_package playwith not able to run command getting error message that I'm attempting to use non function Farhat Maha (Thu 19 Jan 2012 - 22:28:49 GMT)
    - Re: [R] question re\_package playwith not able to run command getting error message that I'm attempting to use non function R. Michael Weylandt (Fri 20 Jan 2012 - 08:05:56 GMT)
    - Re: [R] question re\_package playwith not able to run command getting error message that I'm attempting to use non function Farhat Maha (Fri 20 Jan 2012 - 17:56:39 GMT)
    - Re: [R] question re\_package playwith not able to run command getting error message that I'm attempting to use non function R. Michael Weylandt (Sat 21 Jan 2012 - 06:10:14 GMT)
- Re: [R] Bivariate Partial Dependence Plots in Random Forests Liaw, Andy (Tue 31 Jan 2012 - 13:25:33 GMT)
  - [R] indexing by empty string (was RE: Error in predict.randomForest ... subscript out of bounds with NULL.name in X) Liaw, Andy (Tue 31 Jan 2012 - 13:44:13 GMT)

## Mailing lists

**stackoverflow**

**Tagged Questions**

R is a free, open source graphics. It is advised to use stats.stackexchange learn more... | improve

**CrossValidated**

**Tagged Questions**

4,707 questions tagged

**Community Bulletin**

**Related Tags**

**Stack Exchange Q&A sites**

**R**

R is a language and environment for statistical computing and graphics. It is a GNU project which is similar to the S language and environment which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues. R-Project home is [here](#).

[New Topic](#) [Sub-Forums](#) [People](#) [Options](#)

Topics (92392)	Replies	Last Post	Views	Sub Forum
<a href="#">match values in dependence of ID and Date</a> by Mat	0	<a href="#">1:36pm</a> by Mat	2	<a href="#">R help</a>
<a href="#">How would I sum the number of NAs in multiple vectors</a> by mattbjju2013	1	<a href="#">12:13pm</a> by Carl Witthoft	13	<a href="#">R help</a>
<a href="#">Extract a predictors form constparty object (CHAID output) in R</a> by christiaan pauw-2	1	<a href="#">12:56pm</a> by christiaan pauw-2	3	<a href="#">R help</a>
<a href="#">extract column's from different dataframe</a> by catalin roibu	1	<a href="#">12:55pm</a> by Jim Lemon	3	<a href="#">R help</a>
<a href="#">Constraint on regression parameters</a> by Robert U	1	<a href="#">12:22pm</a> by S Ellison-2	2	<a href="#">R help</a>
<a href="#">map with inset</a> by markw	4	<a href="#">11:47am</a> by markw	17	<a href="#">R help</a>
<a href="#">saveXML() prefix argument</a> by Earl Brown	1	<a href="#">11:36am</a> by Milan Bouchet-Valat	2	<a href="#">R help</a>
<a href="#">Plot time series data irregularly-spaced</a> by Charles Novae de Sa...	4	<a href="#">10:56am</a> by Charles Novae de Sa...	12	<a href="#">R help</a>
<a href="#">Problem with Random Forest predict</a> by Michael Conklin	3	<a href="#">9:29am</a> by warner121	1056	<a href="#">R help</a>

## Forums

# Mailing lists vs. Stack Exchange

## December 2003 Archives by thread

- Messages sorted by: [\[ subject \]](#) [\[ author \]](#) [\[ date \]](#)
- [More info on this list...](#)

Starting: Mon Dec 1 00:27:40 CEST 2003

Ending: Wed Dec 31 21:39:13 CEST 2003

Messages: 1159

- [\[R\] fitting a theoretical distribution with truncated tails](#) Piyush Sharma
  - [\[R\] fitting a theoretical distribution with truncated tails](#) Spencer Graves
- [\[R\] Discovering methods](#) Gabor Grothendieck
  - [\[R\] Discovering methods](#) Duncan Murdoch
  - [\[R\] Discovering methods](#) Thomas Lumley
    - [\[R\] Discovering methods](#) Duncan Murdoch
    - [\[R\] Discovering methods](#) Thomas Lumley
- [\[R\] hdf library for windows](#) Toby.Patterson at csiro.au
- [\[R\] hdf library for windows](#) Mulholland, Tom
- [\[R\] strptime Usage](#) Ko-Kang Kevin Wang
  - [\[R\] strptime Usage](#) Prof Brian Ripley
- [\[R\] Discovering methods](#) Prof Brian Ripley
- [\[R\] significance in difference of proportions: What problem](#) Torsten Hothorn
- [\[R\] Indexing ANOVA table](#) Pascal A. Niklaus
- [\[R\] wilcoxon-pratt signed rank test in R - drug-efficacy](#) Niels Steen Krogh
  - [\[R\] wilcoxon-pratt signed rank test in R - drug-efficacy](#) Peter Dalgaard
  - [\[R\] wilcoxon-pratt signed rank test in R - drug-efficacy](#) Knut M. Witkowski
- [\[R\] Changing distance scale in plclust\(\)](#) Mike White
- [\[R\] Rd Files?](#) Wolski
  - [Historically, at entelnet.bo](#) Wolski
  - [\[R\] Rd Files?](#) Wolski
  - [\[R\] Rd Files?](#) Prof Brian D Ripley
  - [\[R\] Rd Files?](#) Wolski
  - [\[R\] Rd Files?](#) Prof Brian Ripley
  - [\[R\] Rd Files?](#) Wolski
  - [\[R\] Rd Files?](#) Prof Brian Ripley
  - [\[R\] Rd Files?](#) Wolski
  - [\[R\] Rd Files?](#) Prof Brian Ripley
  - [\[R\] Rd Files?](#) Peter Dalgaard
  - [\[R\] Rd Files?](#) Duncan Murdoch

**Historically,**  
coordinating dev't and user support  
activities

- [Guzzi] [\[R\] Rd Files?](#) Prof Brian Ripley
- [\[R\] Rd Files?](#) Wolski
- [\[R\] Rd Files?](#) Prof Brian Ripley
- [\[R\] Rd Files?](#) Wolski
- [\[R\] Rd Files?](#) Prof Brian Ripley
- [\[R\] Rd Files?](#) Wolski
- [\[R\] Rd Files?](#) Prof Brian Ripley
- [\[R\] Rd Files?](#) Peter Dalgaard
- [\[R\] Rd Files?](#) Duncan Murdoch

The screenshot shows the Stack Overflow homepage with the search bar set to 'Tagged Questions' and the tag 'r'. The results list several questions about R:

- How to make a great R reproducible example?**  
431 votes, 11 answers, 24k views. This question discusses creating reproducible examples for R, mentioning mailing lists and SO.
- How can we make xkcd style graphs in R?**  
275 votes, 8 answers, 25k views. This question asks how to create xkcd-style graphs in R, mentioning ggplot2.
- Changing**  
collaborate, learn and communicate among themselves and with their users [Bege1]  
25 votes, 1 answer, 15k views. This question discusses communication and collaboration in R development.
- Fastest Q&A site**  
by column(s) in R  
230 votes, 11 answers, 156k views. This question asks how to sort a data frame by multiple columns in R.
- Good technical solutions**  
I want to sort a data frame by multiple columns in R. For example, with the data.frame below I would like to sort by column z (descending) then by column b (ascending): dd <- data.frame(b = ...  
11 votes, 1 answer, 156k views. This question asks for good technical solutions for sorting data frames in R.

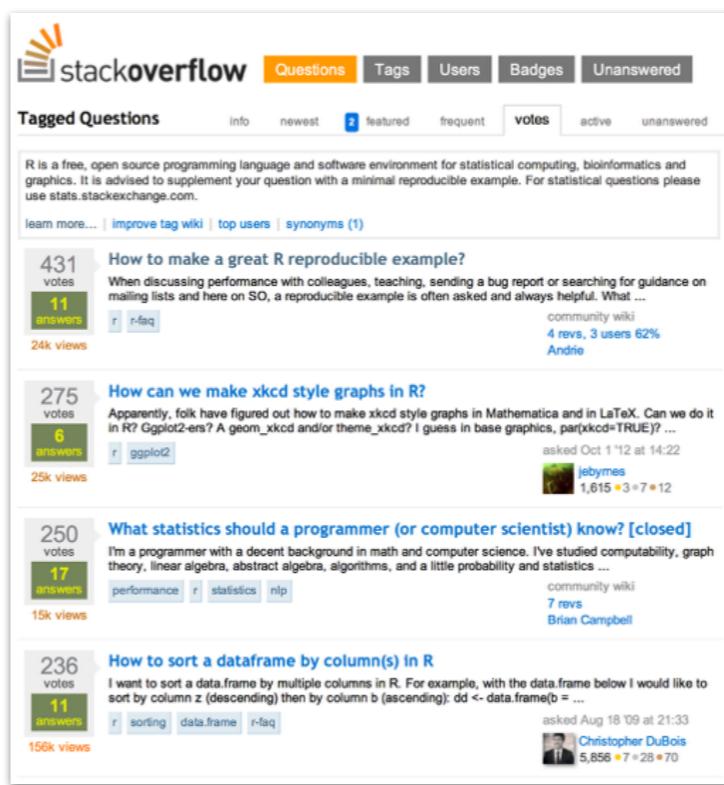
# Research goal

How did knowledge sharing in the  community change with the emergence of StackExchange  Q&A sites?



## December 2003 Archives by thread

- Messages sorted by: [\[subject\]](#) [\[author\]](#) [\[date\]](#)
  - [More info on this list...](#)
- Starting: Mon Dec 1 00:27:40 CEST 2003  
Ending: Wed Dec 31 21:39:13 CEST 2003  
Messages: 1159
- [\[R\] fitting a theoretical distribution with truncated tails](#), Piyush Sharma
    - [\[R\] fitting a theoretical distribution with truncated tails](#), Spencer Graves
  - [\[R\] Discovering methods](#), Gabor Grothendieck
    - [\[R\] Discovering methods](#), Duncan Murdoch
    - [\[R\] Discovering methods](#), Thomas Lumley
      - [\[R\] Discovering methods](#), Duncan Murdoch
      - [\[R\] Discovering methods](#), Thomas Lumley
    - [\[R\] hdf library for windows](#), Toby Patterson at csiro.au
    - [\[R\] hdf library for windows](#), Mulholland, Tom
  - [\[R\] strftime Usage](#), Ko-Kang Kevin Wang
    - [\[R\] strftime Usage](#), Prof Brian Ripley
  - [\[R\] Discovering methods](#), Prof Brian Ripley
  - [\[R\] significance in difference of proportions: What problema](#), Torsten Hothorn
  - [\[R\] Indexing ANOVA table](#), Pascal A. Niklaus
  - [\[R\] wilcoxon-pratt signed rank test in R - drug-efficacy](#), Niels Steen Krogh
    - [\[R\] wilcoxon-pratt signed rank test in R - drug-efficacy](#), Peter Dalgaard
    - [\[R\] wilcoxon-pratt signed rank test in R - drug-efficacy](#), Knut M. Wittkowski
  - [\[R\] Changing distance scale in plclust\(\)](#), Mike White
  - [\[R\] Rd Files?](#), Wolski
    - [\[R\] Rd Files?](#), Kjetil at entelnet.bo
    - [\[R\] Rd Files?](#), Wolski
      - [\[R\] Rd Files?](#), Prof Brian Ripley
      - [\[R\] Rd Files?](#), Wolski
      - [\[R\] Rd Files?](#), Prof Brian D Ripley
      - [\[R\] Rd Files?](#), Wolski
      - [\[R\] Rd Files?](#), Prof Brian Ripley
      - [\[R\] Rd Files?](#), Wolski
      - [\[R\] Rd Files?](#), Prof Brian Ripley
      - [\[R\] Rd Files?](#), Wolski
      - [\[R\] Rd Files?](#), Prof Brian Ripley
      - [\[R\] Rd Files?](#), Peter Dalgaard
      - [\[R\] Rd Files?](#), Duncan Murdoch



The screenshot shows the Stack Overflow homepage with several R-related questions listed:

- How to make a great R reproducible example?**  
When discussing performance with colleagues, teaching, sending a bug report or searching for guidance on mailing lists and here on SO, a reproducible example is often asked and always helpful. What ...  
431 votes, 11 answers, 24k views.
- How can we make xkcd style graphs in R?**  
Apparently, folk have figured out how to make xkcd style graphs in Mathematica and in LaTeX. Can we do it in R? Ggplot2-ers? A geom\_xkcd and/or theme\_xkcd? I guess in base graphics, par(xkcd=TRUE)? ...  
275 votes, 6 answers, 25k views.
- What statistics should a programmer (or computer scientist) know? [closed]**  
I'm a programmer with a decent background in math and computer science. I've studied computability, graph theory, linear algebra, abstract algebra, algorithms, and a little probability and statistics ...  
250 votes, 17 answers, 15k views.
- How to sort a data frame by column(s) in R**  
I want to sort a data.frame by multiple columns in R. For example, with the data.frame below I would like to sort by column z (descending) then by column b (ascending): dd <- data.frame(b = ...  
236 votes, 11 answers, 150k views.



@b\_vasilescu

# Mailing lists vs. Stack Exchange

## [R] Merge dataframes

jdanieldn [jdanieldn@gmail.com](mailto:jdanieldn@gmail.com)

Fri Oct 7 15:34:33 CEST 2011

- Previous message: [\[R\] "r squared" and anova for linear mixed-effects model](#)
- Next message: [\[R\] Merge dataframes](#)
- **Messages sorted by:** [\[ date \]](#) [\[ thread \]](#) [\[ subject \]](#) [\[ author \]](#)

Hello,

I am having some problems to use the 'merge' function. I'm not sure if I got its working right.

What I want to do is:

1) Suppose I have a dataframe like:

	height	width
1	1.1	2.3
2	2.1	2.5
3	1.8	1.9
4	1.6	2.1
5	1.8	2.4

[...]

4) So, I want to merge those dataframes, so that the new variable, color, is binded to the first dataframe. Of course some cases won't have value for it, since I generated this variable in a smaller dataframe. In those cases I want the value to be NA. The result dataframe should be:

	height	width	color
1	1.1	2.3	red
2	2.1	2.5	NA
3	1.8	1.9	red
4	1.6	2.1	NA
5	1.8	2.4	blue

I have written some codes, but they're not working properly. The new variable has its values mixed up, and they do not correspond to its row.names.

```
# Generate the first dataframe
data1 <- data.frame(height=rnorm(20,3,0.2),width=rnorm(20,2,0.5))
# Sample a smaller dataframe from data1
data2 <- data1[sample(1:20,15,replace=F),]
# Generate the new variable
color <- sample(c("red","blue"),15,replace=T)
# Bind the new variable to data2
data2 <- cbind(data2, color)
# Merge the data1 and data2$color by row.names, and force it to have the same
# values that data1. Next it generates a new dataframe where column 1 is the
# row.name, and then sort it by the row.name from data1.
data.frame(merge(data1,data2$color, by=0,
all.x=T),row.names=1)[row.names(data1),]
```

I'm not sure what am I doing wrong. Can anyone see where the mistake is?



## Loops in R - Need to use index, anyway to avoid 'for'?

11  
4  
I know it's not the best practice in R to use the `for` loop because it doesn't have an enhanced performance. For almost all cases there is a function of the family `*apply` that solves our problems.

However I'm facing a situation where I don't see a workaround.

I need to calculate percent variation for consecutive values:

```
pv[1] <- 0
for(i in 2:length(x)) {
  pv[i] <- (x[i] - x[i-1])/x[i-1]
}
```

So, as you can see, I have to use both the `x[i]` element, but also the `x[i-1]` element. By using the `*apply` functions, I just see how to use the `x[i]`. Is there anyway I can avoid the `for` loops?

r loops for-loop

asked May 6 '12 at 1:17

João Daniel  
1,372 3 29

add comment

start a bounty

## 3 Answers

15  
What you offered would be the fractional variation, but if you multiplied by 100 you get the "percent variation":

```
pv<- vector("numeric",length(x))
pv[1] <- 0
pv[-1] <- 100* ( x[-1] - x[-length(x)] )/ x[-length(x)]
```

Vectorized solution. ( And you should note that for-loops are going to be just as slow as `*apply` solutions ... just not as pretty. Always look for a vectorized approach.)

To explain a bit more: The `x[-length(x)]` is the vector, `x[1:(length(x)-1)]`, and the `x[-1]` is the vector, `x[2:length(x)]`, and the vector operations in R are doing the same operations as in your for-loop body, although not using an explicit loop. R first constructs the differences in those shifted vectors, `x[-length(x)] - x[-1]`, and then divides by `x[1:(length(x)-1)]`.

edited May 6 '12 at 3:31

answered May 6 '12 at 1:28

DWin  
82.9k 3 39 102

Nice response DWin. I didn't actually know what the poster was accomplishing but I'm in 100% agreement on the vectorization. +1 – [Tyler Rinker](#) May 6 '12 at 1:37

Very nice answer! I didn't know that the vectorized approach was the fastest, I thought `lapply` was. But in the last line of code, shouldn't be `x[-1] - x[-length(x)]`? – [João Daniel](#) May 6 '12 at 2:14

@JoãoDaniel: Yes, it should. Edit applied. – [DWin](#) May 6 '12 at 3:31 ↗

# Mailing lists vs. Stack Exchange

## [R] Merge dataframes

jdanieldn jdanieldn at gmail.com

Fri Oct 7 15:34:33 CEST 2011

- Previous message: [R] "r squared" and anova for linear mixed-effects model
- Next message: [R] Merge dataframes
- Messages sorted by: [ date ] [ thread ] [ subject ] [ author ]

Hello,

I am having some problems to use the 'merge' function. I'm not sure if I got its working right.

What I want to do is:

1) Suppose I have a dataframe like:

	height	width
1	1.1	2.3
2	2.1	2.5
3	1.8	1.9
4	1.6	2.1
5	1.8	2.4

[...]

4) So, I want to merge those dataframes, so that the new variable, color, is binded to the first dataframe. Of course some cases won't have value for it, since I generated this variable in a smaller dataframe. In those cases I want the value to be NA. The result dataframe should be:

	height	width	color
1	1.1	2.3	red
2	2.1	2.5	NA
3	1.8	1.9	red
4	1.6	2.1	NA
5	1.8	2.4	blue

I have written some codes, but they're not working properly. The new variable has its values mixed up, and they do not correspond to its row.names.

```
# Generate the first dataframe
data1 <- data.frame(height=rnorm(20,3,0.2),width=rnorm(20,2,0.5))
# Sample a smaller dataframe from data1
data2 <- data1[sample(1:20,15,replace=F),]
# Generate the new variable
color <- sample(c("red","blue"),15,replace=T)
# Bind the new variable to data2
data2 <- cbind(data2, color)
# Merge the data1 and data2$color by row.names, and force it to has the same
values that data1. Next it generates a new dataframe where column 1 is the
row.name, and then sort it by the row.name from data1.
data.frame(merge(data1,data2$color, by=0,
all.x=T),row.names=1)[row.names<-data1[,]
```

I'm not sure what am I doing wrong. Can anyone see where the mistake is?

Tags

code

highlighting



## Loops in R - Need to use index, anyway to avoid 'for'?

I know it's not the best practice in R to use the `for` loop because it doesn't have an enhanced performance. For almost all cases there is a function of the family `*apply` that solves our problems.

However I'm facing a situation where I don't see a workaround.

I need to calculate percent variation for consecutive values:

```
pv[1] <- 0
for(i in 2:length(x)) {
  pv[i] <- (x[i] - x[i-1])/x[i-1]
}
```

So, as you can see, I have to use both the `x[i]` element, but also the `x[i-1]` element. By using the `*apply` functions, I just see how to use the `x[i]`. Is there anyway I can avoid the `for` loops?

`r loops for-loop`

asked May 6 '12 at 1:17

João Daniel  
1,372 ● 3 ● 29

add comment

start a bounty

## 3 Answers

What you offered would be the fractional variation, but if you multiplied by 100 you get the "percent variation":

```
pv<- vector("numeric",length(x))
pv[1] <- 0
pv[-1] <- 100* ( x[-1] - x[-length(x)] ) / x[-length(x)]
```

Vectorized solution. ( And you should note that for-loops are going to be just as slow as `*apply` solutions ... just not as pretty. Always look for a vectorized approach.)

To explain a bit more: The `x[-length(x)]` is the vector, `x[1:(length(x)-1)]`, and the `x[-1]` is the vector, `x[2:length(x)]`, and the vector operations in R are doing the same operations as in your for-loop body, although not using an explicit loop. R first constructs the differences in those shifted vectors, `x[-length(x)] - x[-1]`, and then divides by `x[1:(length(x)-1)]`.

edited May 6 '12 at 3:31

answered May 6 '12 at 1:28

DWin  
82.9k ● 3 ● 39 ● 102

Nice response DWin. I didn't actually know what the poster was accomplishing but I'm in 100% agreement on the vectorization. +1 – Tyler Rinker May 6 '12 at 1:37

Very nice answer! I didn't know that the vectorized approach was the fastest, I thought `lapply` was. But in the last line of code, shouldn't be `x[-1] - x[-length(x)]`? – João Daniel May 6 '12 at 2:14

@JoãoDaniel Yes, it should. Edit applied – DWin May 6 '12 at 2:31

Modern (fancy) user interface



# Mailing lists vs. Stack Exchange

## [R] Merge dataframes

jdanieldn jdanieldn at gmail.com

Fri Oct 7 15:34:33 CEST 2011

- Previous message: [R] "r squared" and anova for linear mixed-effects model
- Next message: [R] Merge dataframes
- Messages sorted by: [ date ] [ thread ] [ subject ] [ author ]

Hello,

I am having some problems to use the 'merge' function. I'm not sure if I got its working right.

What I want to do is:

1) Suppose I have a dataframe like:

```
height      width
1   1.1       2.3
2   2.1       2.5
3   1.8       1.9
4   1.6       2.1
5   1.8       2.4
```

[...]

4) So, I want to merge those dataframes, so that the new variable, color, is binded to the first dataframe. Of course some cases won't have value for it, since I generated this variable in a smaller dataframe. In those cases I want the value to be NA. The result dataframe should be:

```
height      width      color
1   1.1       2.3       red
2   2.1       2.5       NA
3   1.8       1.9       red
4   1.6       2.1       NA
5   1.8       2.4       blue
```

I have written some codes, but they're not working properly. The new variable has its values mixed up, and they do not correspond to its row.names.

```
# Generate the first dataframe
data1 <- data.frame(height=rnorm(20,3,0.2),width=rnorm(20,2,0.5))
# Sample a smaller dataframe from data1
data2 <- data1[sample(1:20,15,replace=F),]
# Generate the new variable
color <- sample(c("red","blue"),15,replace=T)
# Bind the new variable to data2
data2 <- cbind(data2, color)
# Merge the data1 and data2$color by row.names, and force it to has the same
values that data1. Next it generates a new dataframe where column 1 is the
row.name, and then sort it by the row.name from data1.
data.frame(merge(data1,data2$color, by=0,
all.x=T),row.names=1)[row.names(data1),]
```

I'm not sure what am I doing wrong. Can anyone see where the mistake is?

vote  
count



## Loops in R - Need to use index, anyway to avoid 'for'?

know it's not the best practice in R to use the `for` loop because it doesn't have an enhanced performance. For almost all cases there is a function of the family `*apply` that solves our problems.

However I'm facing a situation where I don't see a workaround.

I need to calculate percent variation for consecutive values:

```
pv[1] <- 0
for(i in 2:length(x)) {
  pv[i] <- (x[i] - x[i-1])/x[i-1]
}
```

So, as you can see, I have to use both the `x[i]` element, but also the `x[i-1]` element. By using the `*apply` functions, I just see how to use the `x[i]`. Is there anyway I can avoid the `for` loops?

r loops for-loop

asked May 6 '12 at 1:17

João Daniel  
1,372 3 29

add comment

start a bounty

## 3 Answers

What you offered would be the fractional variation, but if you multiplied by 100 you get the "percent variation":

```
pv<- vector("numeric",length(x))
pv[1] <- 0
pv[-1] <- 100* ( x[-1] - x[-length(x)] ) / x[-length(x)]
```

Vectorized solution. ( And you should note that for-loops are going to be just as slow as `*apply` solutions ... just not as pretty. Always look for a vectorized approach.)

To explain a bit more: The `x[-length(x)]` is the vector, `x[1:(length(x)-1)]`, and the `x[-1]` is the vector, `x[2:length(x)]`, and the vector operations in R are doing the same operations as in your for-loop body, although not using an explicit loop. R first constructs the differences in those shifted vectors, `x[-length(x)] - x[-1]`, and then divides by `x[1:(length(x)-1)]`.

edited May 6 '12 at 3:31

answered May 6 '12 at 1:28

DWin  
82.9k 3 39 102

Nice response DWin. I didn't actually know what the poster was accomplishing but I'm in 100% agreement on the vectorization. +1 – Tyler Rinker May 6 '12 at 1:37

Very nice answer! I didn't know that the vectorized approach was the fastest, I thought `lapply` was. But in the last line of code, shouldn't be `x[-1] - x[-length(x)]`? – João Daniel May 6 '12 at 2:14

@JoãoDaniel: Yes, it should. Edit applied – DWin May 6 '12 at 2:31

Gamification features



# Mailing lists vs. Stack Exchange

## [R] Merge dataframes

jdanielnd jdanielnd at gmail.com

Fri Oct 7 15:34:33 CEST 2011

- Previous message: [R] "r squared" and anova for linear mixed-effects model
- Next message: [R] Merge dataframes
- Messages sorted by: [ date ] [ thread ] [ subject ] [ author ]

He: jdanielnd jdanielnd at gmail.com

I: Fri Oct 7 15:34:33 CEST 2011

sure if I got

What I want to do is:

1) Suppose I have a dataframe like:

	height	width
1	1.1	2.3
2	2.1	2.5
3	1.8	1.9
4	1.6	2.1
5	1.8	2.4

Author

[...]

4) So, I want to merge those dataframes, so that the new variable, color, is binded to the first dataframe. Of course some cases won't have value for it, since I generated this variable in a smaller dataframe. In those cases I want the value to be NA. The result dataframe should be:

	height	width	color
1	1.1	2.3	red
2	2.1	2.5	NA
3	1.8	1.9	red
4	1.6	2.1	NA
5	1.8	2.4	blue

I have written some codes, but they're not working properly. The new variable has its values mixed up, and they do not correspond to its row.names.

```
# Generate the first dataframe
data1 <- data.frame(height=rnorm(20,3,0.2),width=rnorm(20,2,0.5))
# Sample a smaller dataframe from data1
data2 <- data1[sample(1:20,15,replace=F),]
# Generate the new variable
color <- sample(c("red","blue"),15,replace=T)
# Bind the new variable to data2
data2 <- cbind(data2, color)
# Merge the data1 and data2$color by row.names, and force it to has the same
values that data1. Next it generates a new dataframe where column 1 is the
row.name, and then sort it by the row.name from data1.
data.frame(merge(data1,data2$color, by=0,
all.x=T),row.names=1)[row.names(data1),]
```

I'm not sure what am I doing wrong. Can anyone see where the mistake is?



## Loops in R - Need to use index, anyway to avoid 'for'?

I know it's not the best practice in R to use the `for` loop because it doesn't have an enhanced performance. For almost all cases there is a function of the family `*apply` that solves our problems.

However I'm facing a situation where I don't see a workaround.

I need to calculate percent variation for consecutive values:

```
pv[1] <- 0
for(i in 2:length(x)) {
  pv[i] <- (x[i] - x[i-1])/x[i-1]
}
```

So, as you can see, I have to use both the `x[i]` element, but also the `x[i-1]` element. By using the `*apply` functions, I just see how to use the `x[i]`. Is there anyway I can avoid the `for` loops?

r loops for-loop

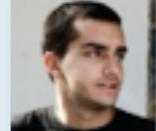
asked May 6 '12 at 1:17

 João Daniel  
1,372 3 29

add comment

start a bounty

asked May 6 '12 at 1:17

 João Daniel  
1,372 3 29

## 3 Answers

What you offered would be the fractional variation, but if you multiplied by 100 you get the "percent variation":

```
pv<- vector("numeric",length(x))
pv[1] <- 0
pv[-1] <- 100* ( x[-1] - x[-length(x)] ) / x[-length(x)]
```

Vectorized solution. ( And you should note that for-loops are going to be just as slow as `*apply` solutions ... just not as pretty. Always look for a vectorized approach.)

To explain a bit more: The `x[-length(x)]` is the vector, `x[1:(length(x)-1)]`, and the `x[-1]` is the vector, `x[2:length(x)]`, and the vector operations in R are doing the same operations as in your for-loop body, although not using an explicit loop. R first constructs the differences in those shifted vectors, `x[-length(x)] - x[-1]`, and then divides by `x[1:(length(x)-1)]`.

edited May 6 '12 at 3:31

answered May 6 '12 at 1:28  
 DWin  
82.9k 3 39 102

Nice response DWin. I didn't actually know what the poster was accomplishing but I'm in 100% agreement on the vectorization. +1 – Tyler Rinker May 6 '12 at 1:37

Very nice answer! I didn't know that the vectorized approach was the fastest, I thought `lapply` was. But in the last line of code, shouldn't be `x[-1] - x[-length(x)]`? – João Daniel May 6 '12 at 2:14

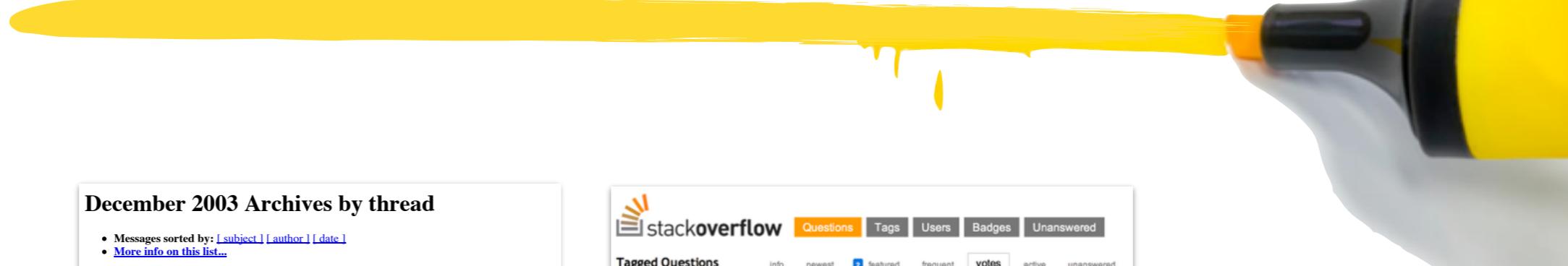
@JoãoDaniel: Yes, it should. Edit applied – DWin May 6 '12 at 2:31

Recognition



# Research goal

How did knowledge sharing in the  community change with the emergence of StackExchange  Q&A sites?



**December 2003 Archives by thread**

- Messages sorted by: [\[subject\]](#) [\[author\]](#) [\[date\]](#)
- [More info on this list...](#)

Starting: Mon Dec 1 00:27:40 CEST 2003  
Ending: Wed Dec 31 21:39:13 CEST 2003  
Messages: 1159

- [\[R\] fitting a theoretical distribution with truncated tails](#), Piyush Sharma
  - [\[R\] fitting a theoretical distribution with truncated tails](#), Spencer Graves
- [\[R\] Discovering methods](#), Gabor Grothendieck
  - [\[R\] Discovering methods](#), Duncan Murdoch
  - [\[R\] Discovering methods](#), Thomas Lumley
    - [\[R\] Discovering methods](#), Duncan Murdoch
    - [\[R\] Discovering methods](#), Thomas Lumley
- [\[R\] hdf library for windows](#), Toby.Patterson at csiro.au
- [\[R\] hdf library for windows](#), Mulholland, Tom
- [\[R\] strptime Usage](#), Ko-Kang Kevin Wang
  - [\[R\] strptime Usage](#), Prof Brian Ripley
- [\[R\] Discovering methods](#), Prof Brian Ripley
- [\[R\] significance in difference of proportions: What problema](#), Torsten Hothorn
- [\[R\] Indexing ANOVA table](#), Pascal A. Niklaus
- [\[R\] wilcoxon-pratt signed rank test in R - drug-efficacy](#), Niels Steen Krogh
  - [\[R\] wilcoxon-pratt signed rank test in R - drug-efficacy](#), Peter Dalgaard
  - [\[R\] wilcoxon-pratt signed rank test in R - drug-efficacy](#), Knut M. Wittkowski
- [\[R\] Changing distance scale in plclust\(\)](#), Mike White
- [\[R\] Rd Files?](#), Wolski
  - [\[R\] Rd Files?](#), kjetil at entelnet.bo
    - [\[R\] Rd Files?](#), Prof Brian Ripley
    - [\[R\] Rd Files?](#), Wolski
    - [\[R\] Rd Files?](#), Prof Brian D Ripley
    - [\[R\] Rd Files?](#), Wolski
    - [\[R\] Rd Files?](#), Prof Brian Ripley
    - [\[R\] Rd Files?](#), Wolski
    - [\[R\] Rd Files?](#), Prof Brian Ripley
    - [\[R\] Rd Files?](#), Peter Dalgaard
    - [\[R\] Rd Files?](#), Duncan Murdoch

 stackoverflow Questions Tags Users Badges Unanswered

Tagged Questions info newest featured frequent votes active unanswered

R is a free, open source programming language and software environment for statistical computing, bioinformatics and graphics. It is advised to supplement your question with a minimal reproducible example. For statistical questions please use stats.stackexchange.com.

learn more... | [improve tag wiki](#) | [top users](#) | [synonyms \(1\)](#)

**How to make a great R reproducible example?**  
When discussing performance with colleagues, teaching, sending a bug report or searching for guidance on mailing lists and here on SO, a reproducible example is often asked and always helpful. What ...  
[r](#) [r-faq](#)

**How can we make xkcd style graphs in R?**  
Apparently, folk have figured out how to make xkcd style graphs in Mathematica and in LaTeX. Can we do it in R? Ggplot2-ers? A geom\_xkcd and/or theme\_xkcd? I guess in base graphics, par(xkcd=TRUE)? ...  
[r](#) [ggplot2](#)

**What statistics should a programmer (or computer scientist) know? [closed]**  
I'm a programmer with a decent background in math and computer science. I've studied computability, graph theory, linear algebra, abstract algebra, algorithms, and a little probability and statistics ...  
[performance](#) [r](#) [statistics](#) [nlp](#)

**How to sort a data frame by column(s) in R**  
I want to sort a data.frame by multiple columns in R. For example, with the data.frame below I would like to sort by column z (descending) then by column b (ascending): dd <- data.frame(b = ...  
[r](#) [sorting](#) [data.frame](#) [r-faq](#)



@b\_vasilescu

# Research goal

How did knowledge sharing in the  community change with the emergence of StackExchange  Q&A sites?



Transition to StackExchange  ?

“Young experts don’t want to have to monitor email all day to be part of the discussion. Their answers belong on a website with a normal content management system, with good search functions and user interactions. Go [to Stack Exchange and] sign up.”

[2010 blog entry]

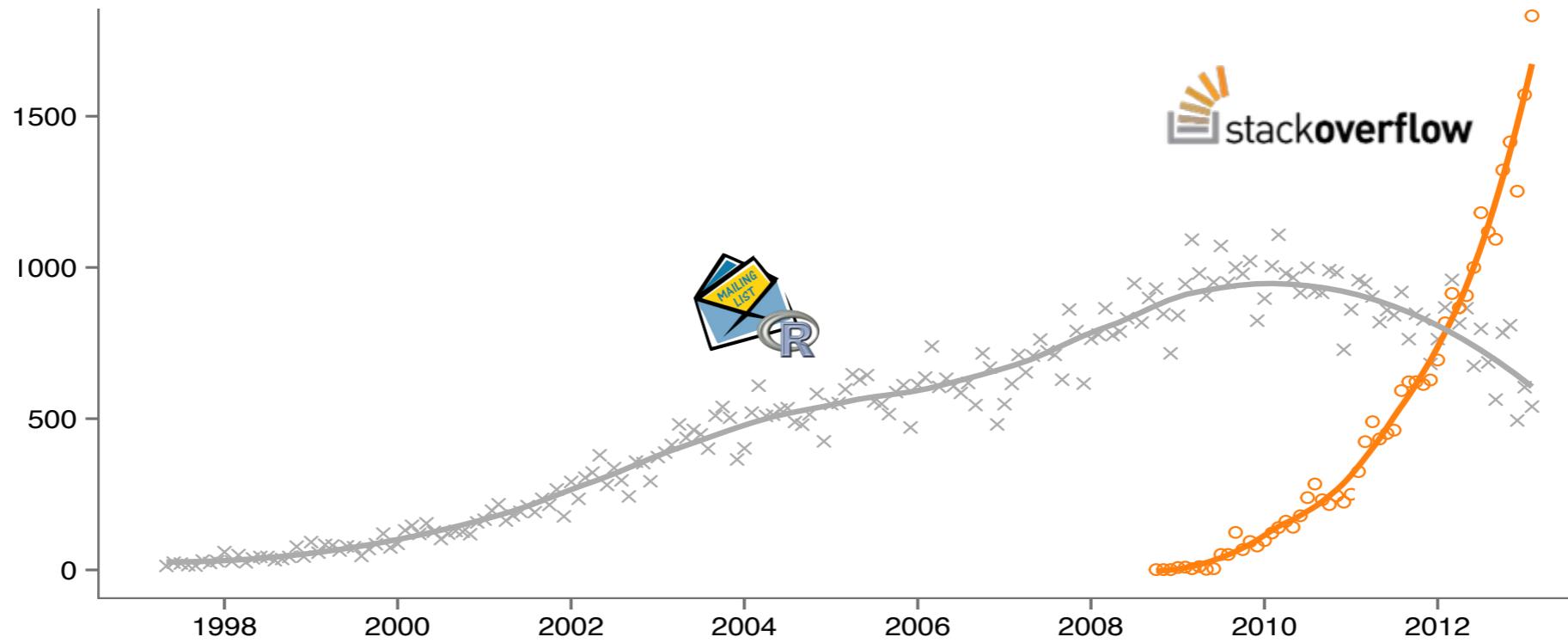


@b\_vasilescu

# Research goal

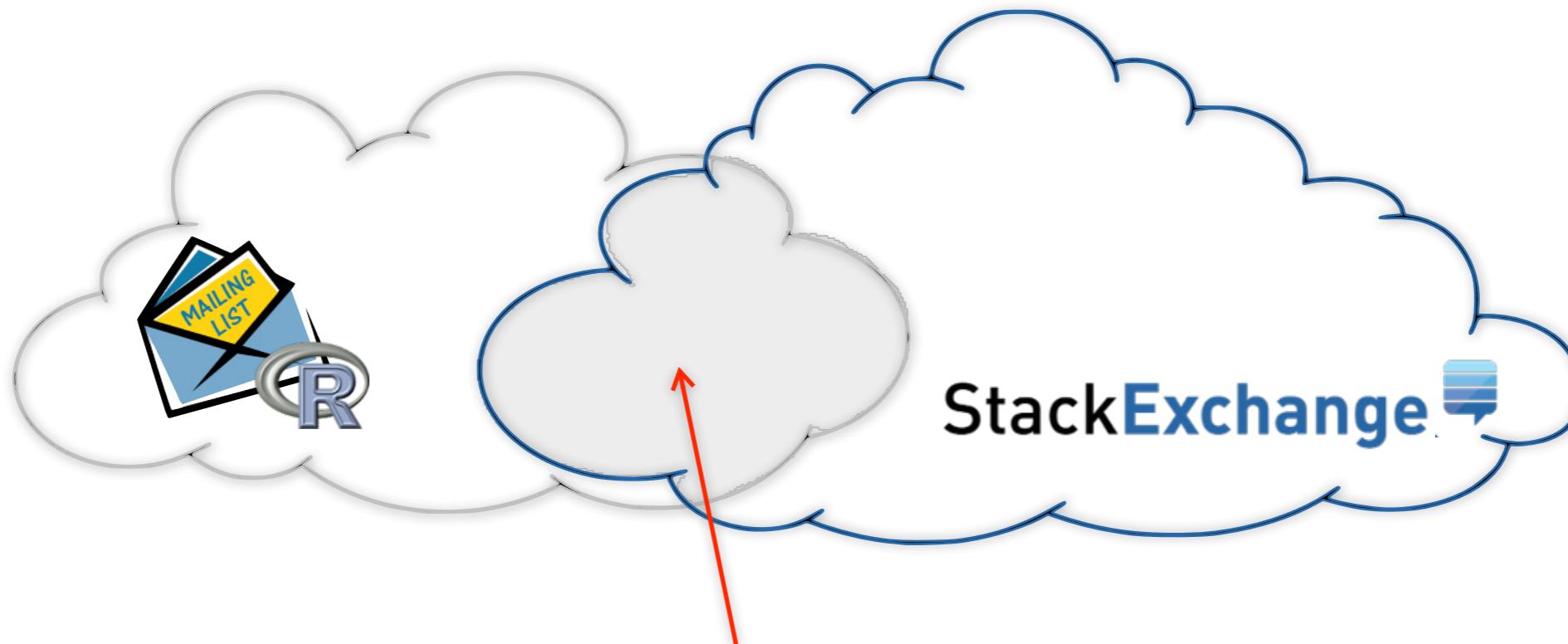
How did knowledge sharing in the  community change with the emergence of StackExchange  Q&A sites?

#Questions asked  
each month



@b\_vasilescu

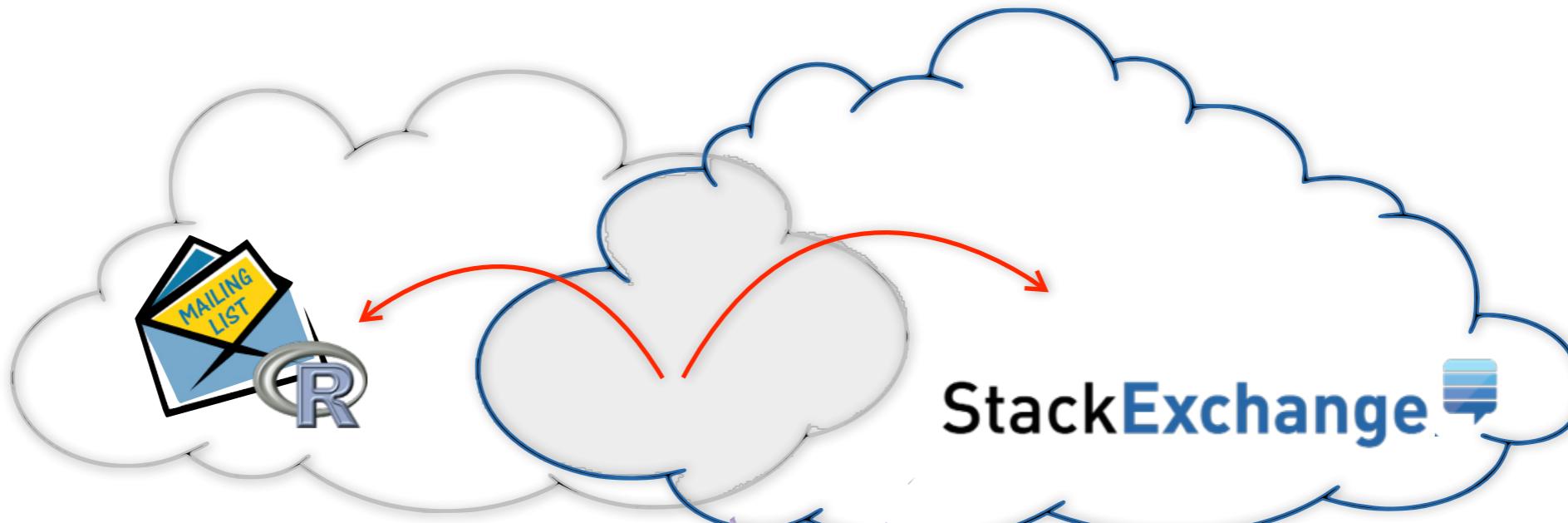
# Today



Who are they?



# Today



Who are they?



Do they behave  
any differently?



@b\_vasilescu

# Today



Who are they?

Mining repositories



+



Do they behave  
any differently?



User survey (112 resp.)

# Mining challenges



Laurent Gautier

Laurent Gautier - [laurent@cbs.dtu.dk](mailto:laurent@cbs.dtu.dk)

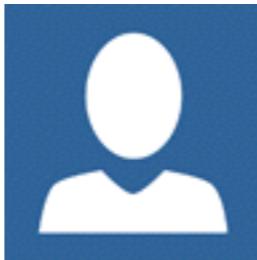
# Mining challenges



Laurent Gautier

Laurent Gautier - [laurent@cbs.dtu.dk](mailto:laurent@cbs.dtu.dk)

# Mining challenges

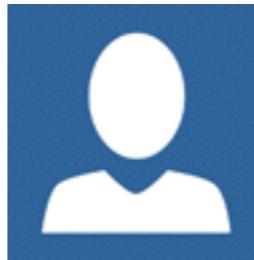


Laurent Gautier

Laurent Gautier - [laurent@cbs.dtu.dk](mailto:laurent@cbs.dtu.dk)

Laurent Gautier - [s010592@student.dtu.dk](mailto:s010592@student.dtu.dk)

# Mining challenges

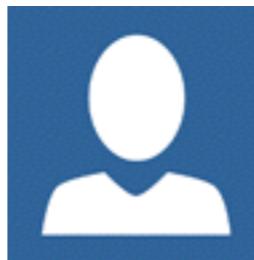


Laurent Gautier

Laurent Gautier - [laurent@cbs.dtu.dk](mailto:laurent@cbs.dtu.dk)

Laurent **Gautier** - [s010592@student.dtu.dk](mailto:s010592@student.dtu.dk)

# Mining challenges



**Laurent Gautier**

Laurent Gautier - [laurent@cbs.dtu.dk](mailto:laurent@cbs.dtu.dk)

Laurent Gautier - [s010592@student.dtu.dk](mailto:s010592@student.dtu.dk)

Laurent - [lgautier@gmail.com](mailto:lgautier@gmail.com)

# Mining challenges



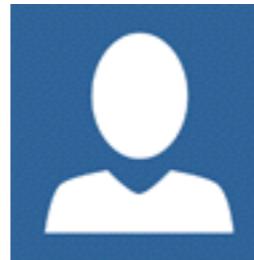
**Laurent Gautier**

Laurent Gautier - [laurent@cbs.dtu.dk](mailto:laurent@cbs.dtu.dk)

Laurent Gautier - [s010592@student.dtu.dk](mailto:s010592@student.dtu.dk)

Laurent - [lgautier@gmail.com](mailto:lgautier@gmail.com)

# Mining challenges



Laurent Gautier

Laurent Gautier - [laurent@cbs.dtu.dk](mailto:laurent@cbs.dtu.dk)

Laurent Gautier - [s010592@student.dtu.dk](mailto:s010592@student.dtu.dk)

Laurent - [lgautier@gmail.com](mailto:lgautier@gmail.com)

- [lgautier@altern.org](mailto:lgautier@altern.org)

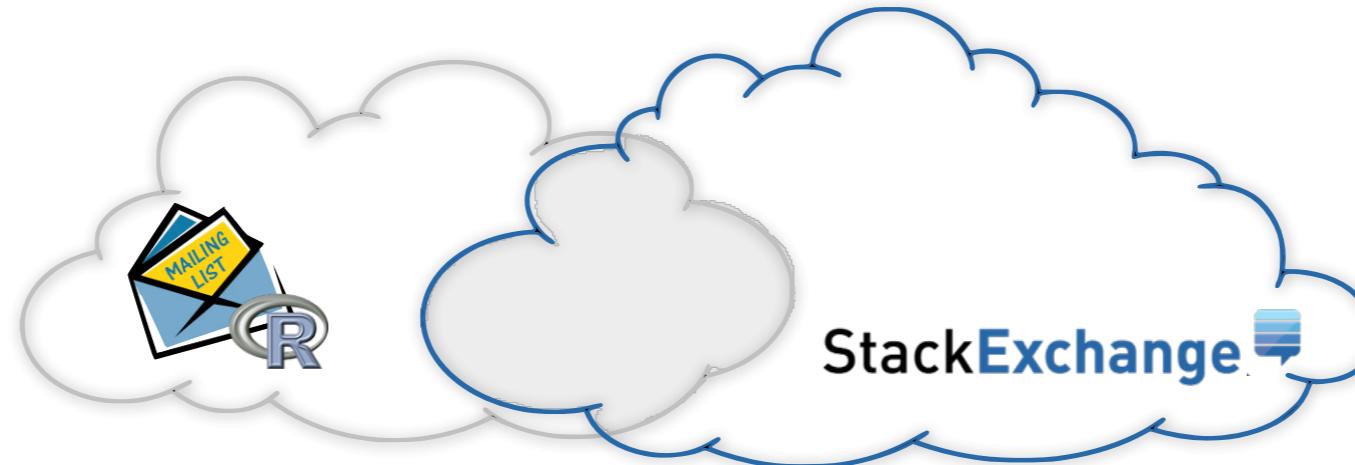
...

9 email addresses, 3 names



Identity merging is far from trivial!

# Mining challenges



Laurent Gautier - [laurent@cbs.dtu.dk](mailto:laurent@cbs.dtu.dk)

Laurent Gautier - [s010592@student.dtu.dk](mailto:s010592@student.dtu.dk)

Laurent  
- [lgautier@gmail.com](mailto:lgautier@gmail.com)  
- [lgautier@altern.org](mailto:lgautier@altern.org)

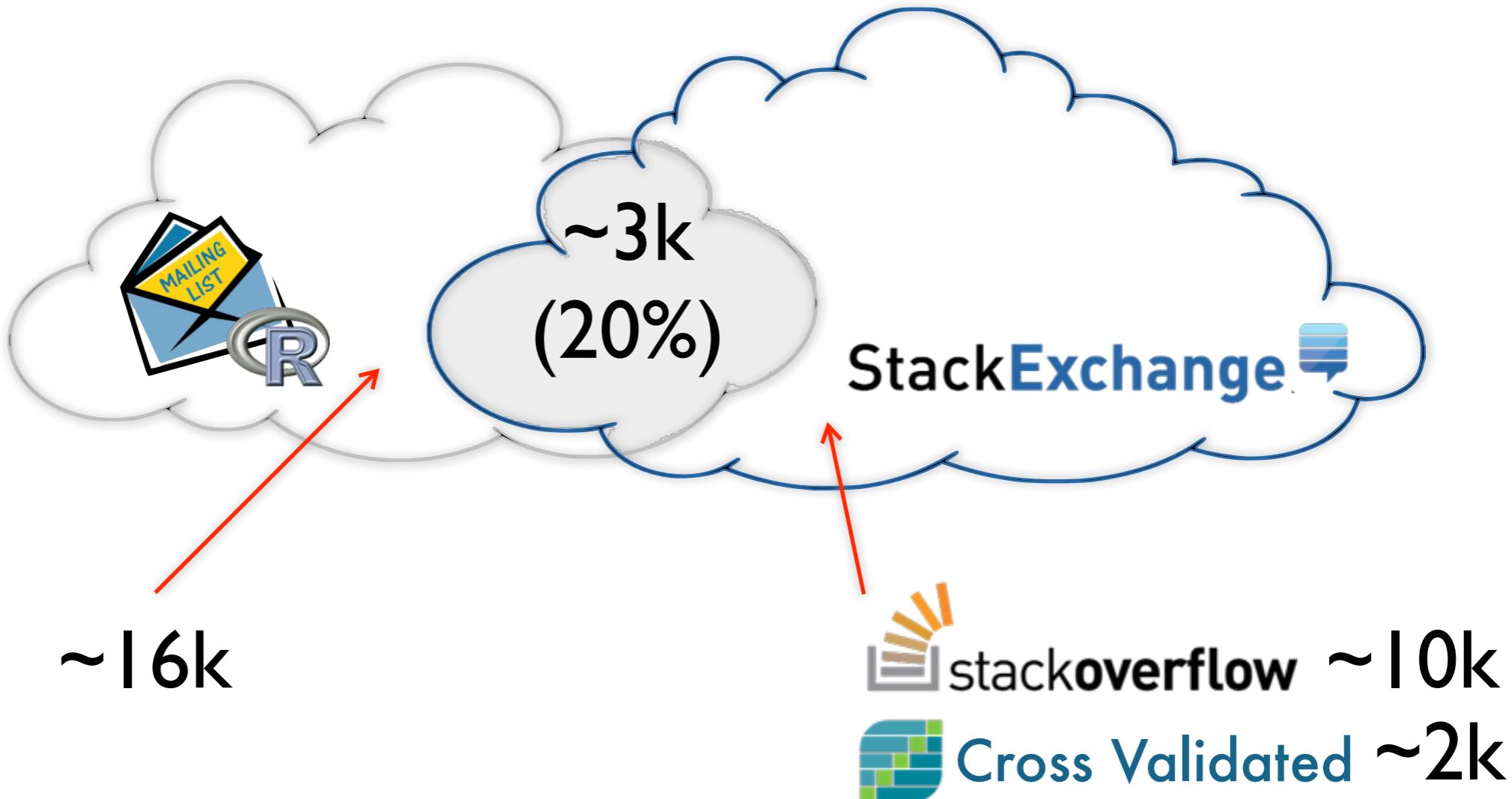
MD5([lgautier@gmail.com](mailto:lgautier@gmail.com))  
= 8e5422a173711c086b685140dbc498fe

...



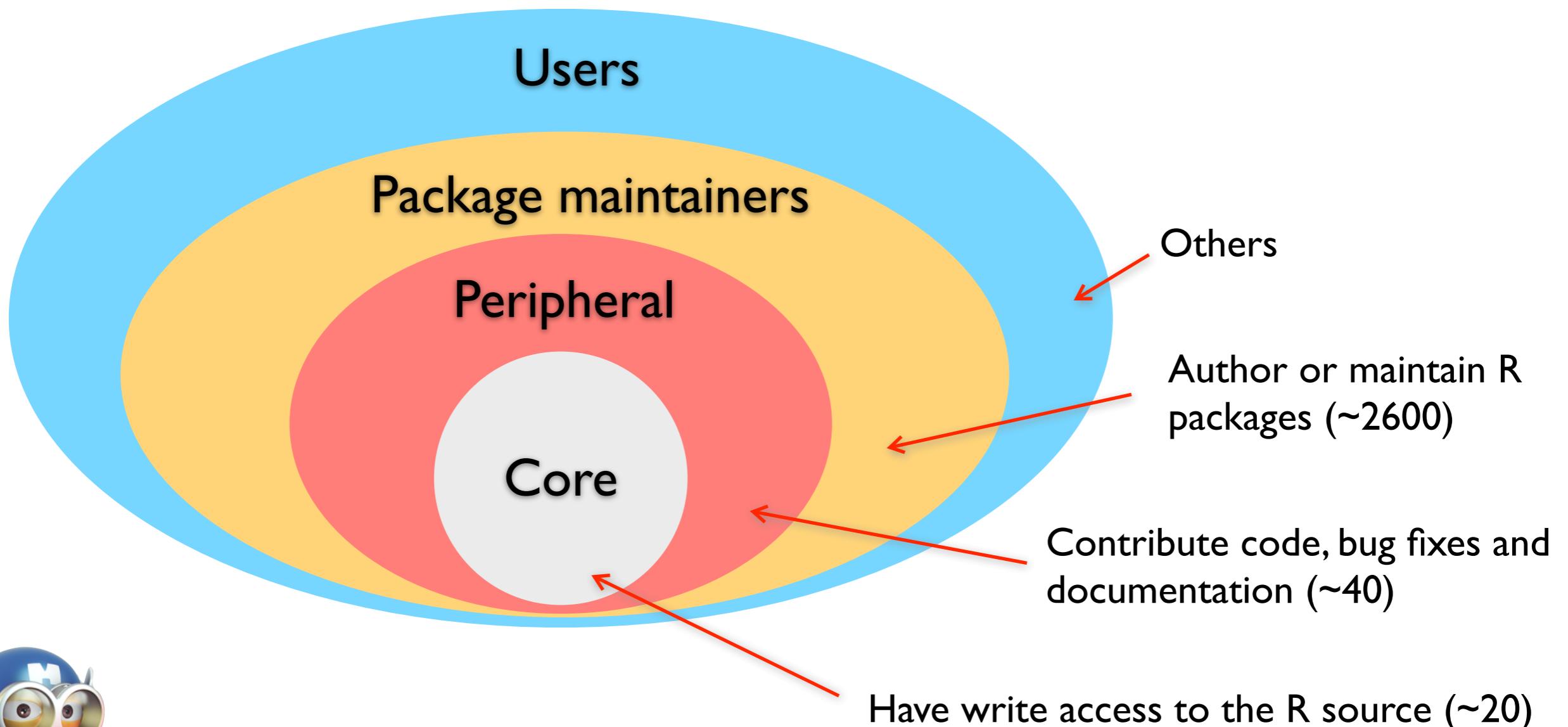
# The R community

> September 2008



Who are they?

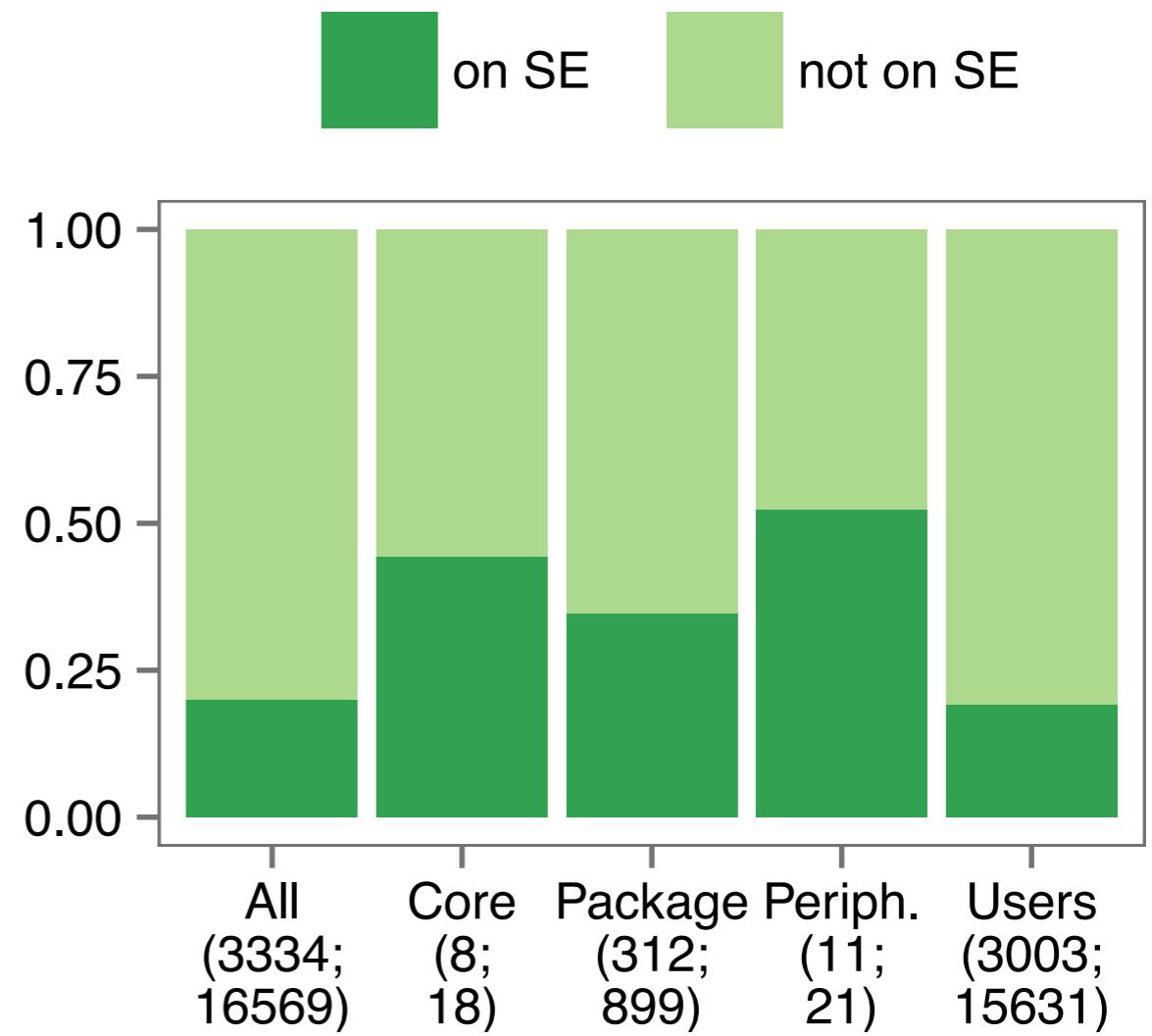
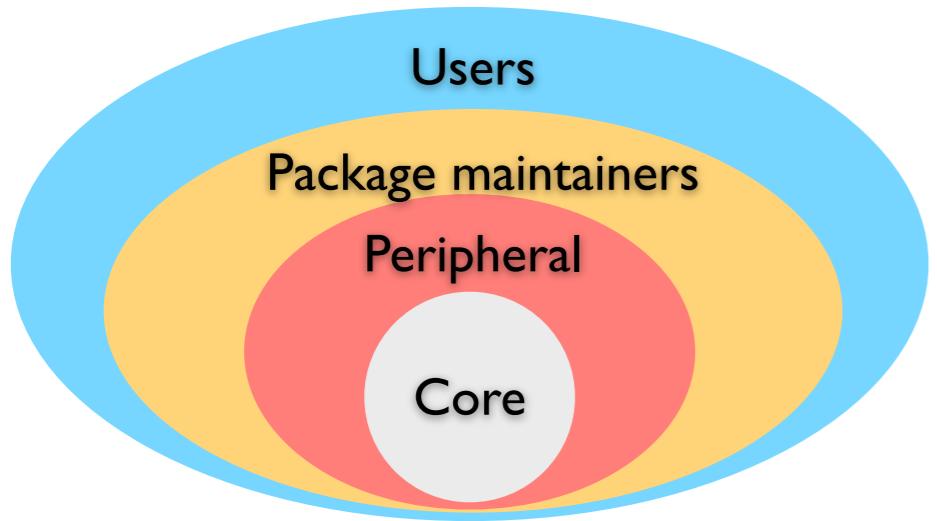
# The R community



Who are they?



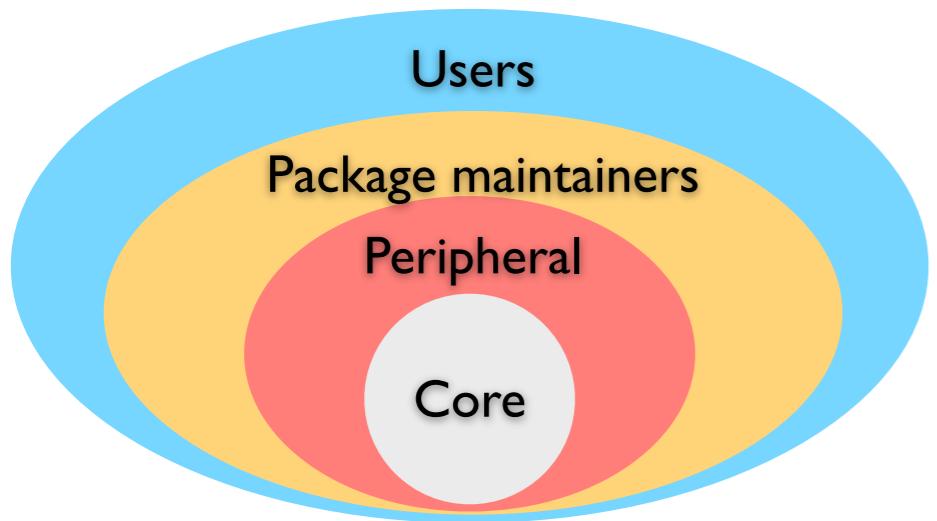
# users on StackExchange



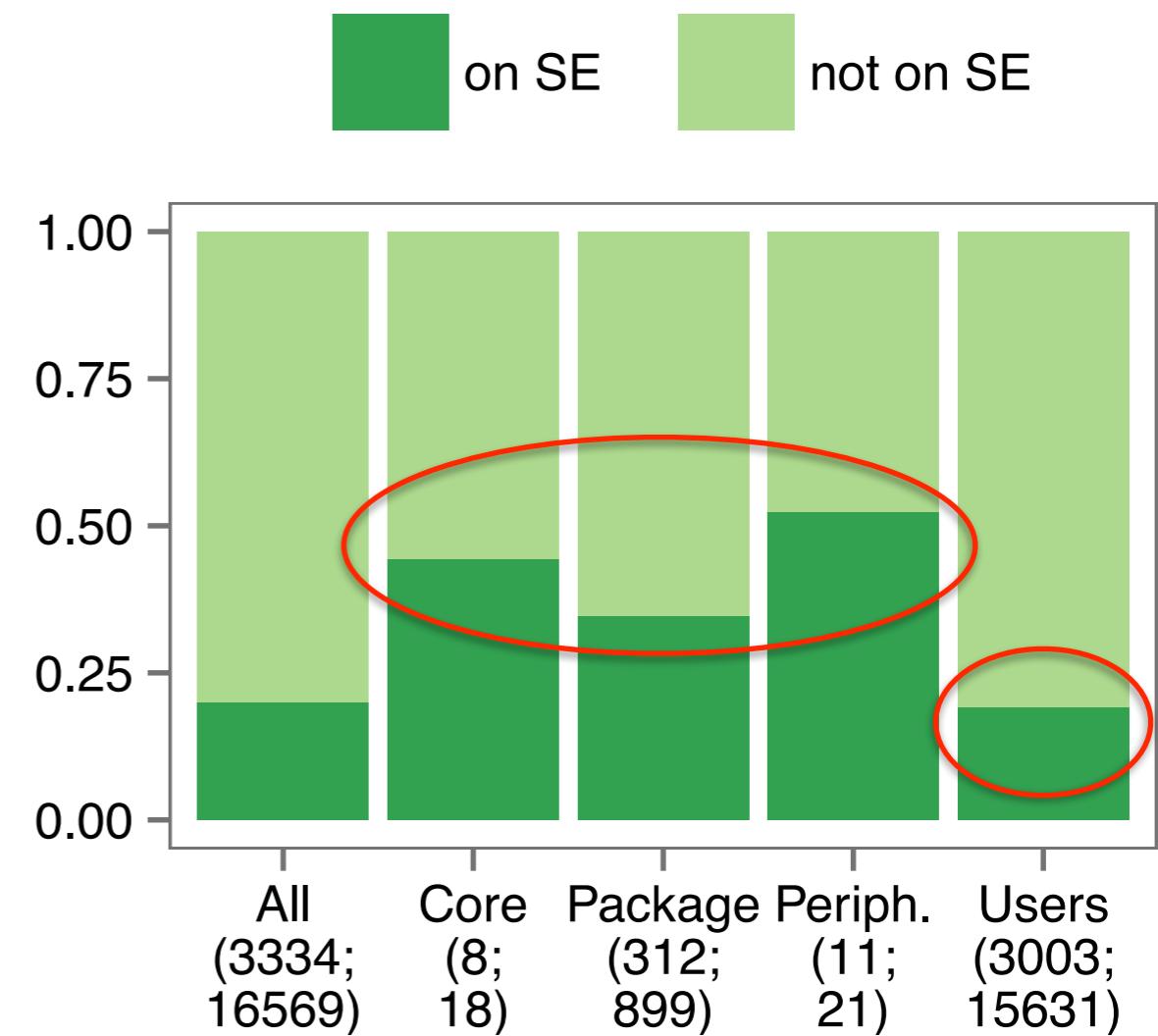
Who are they?



# users on StackExchange



Devs vs users: 1.6x-2x  
higher chance of being on  
**StackExchange**



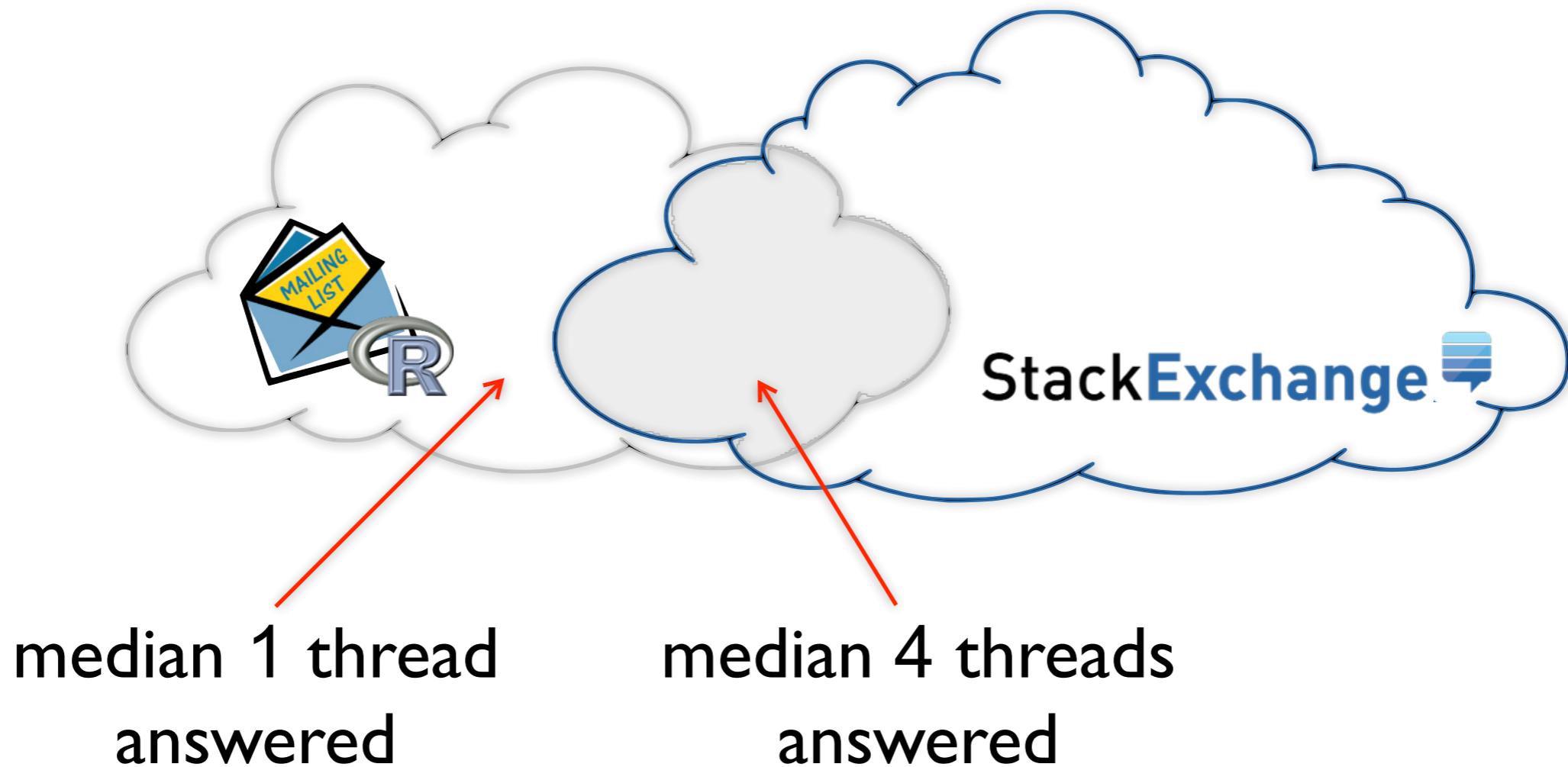
Who are they?



# users on StackExchange

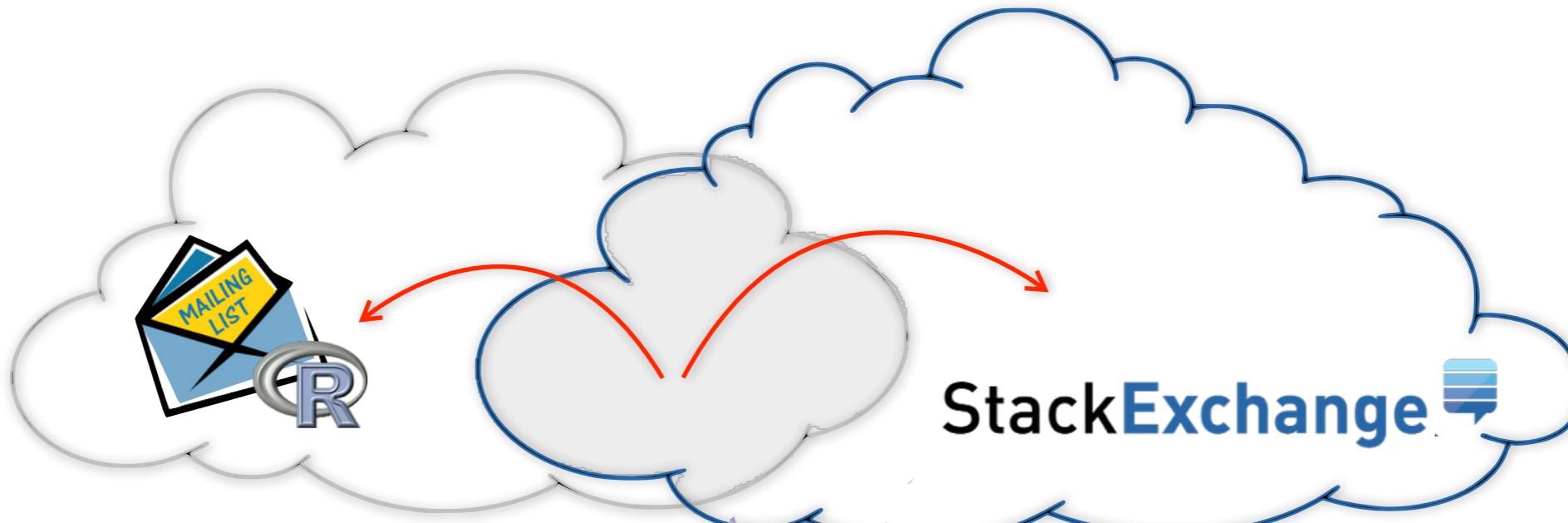


Consistently more active



Who are they?

# Today



Who are they?

Mostly devs,  
consistently  
more active



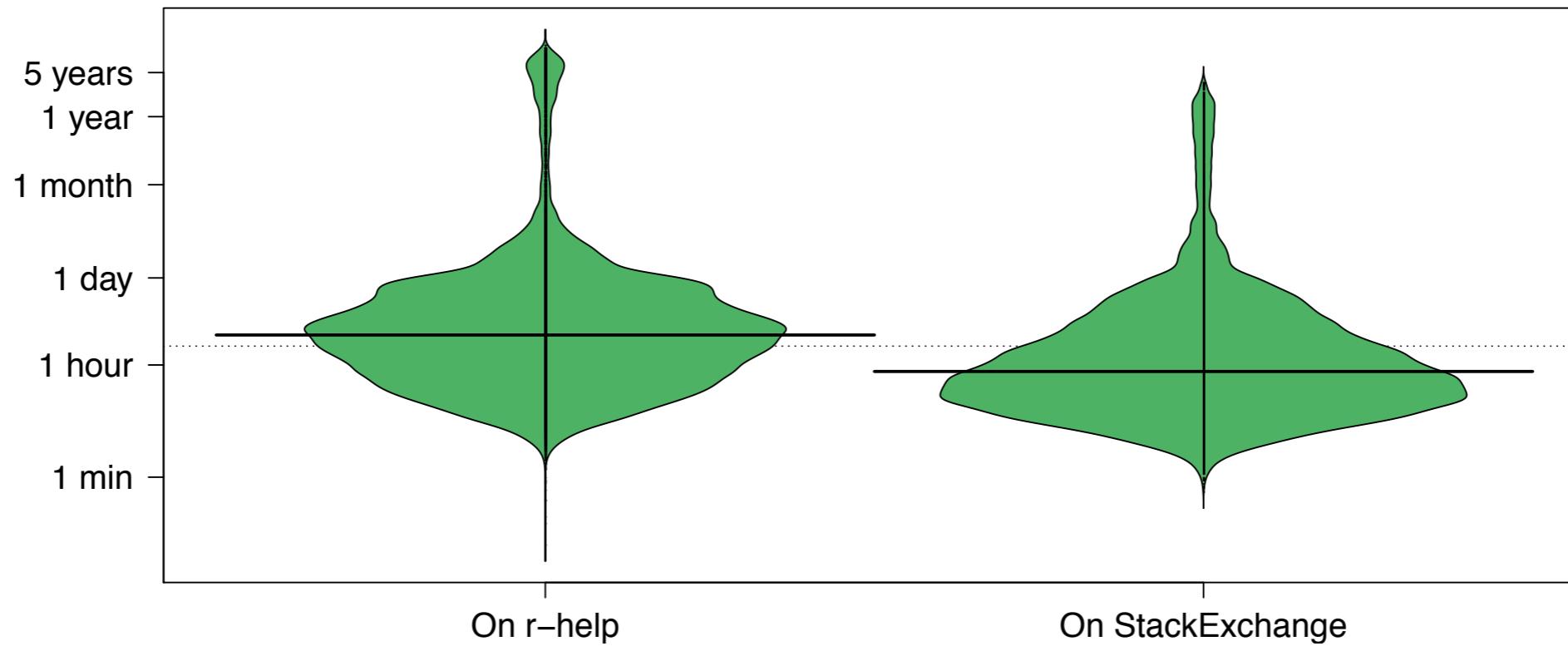
Do they behave  
any differently?



@b\_vasilescu

# The same users answer faster on StackExchange

Speed of answers for r-help participants active on StackExchange



Do they behave any differently?

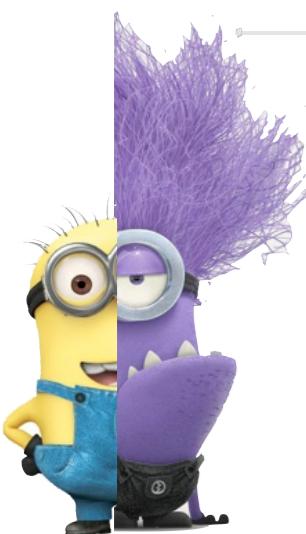


@b\_vasilescu

# The same users answer faster on StackExchange

## Answer Badges

 Enlightened	First to answer and accepted with score of 10 or more	117k awarded
 Generalist	Provided non-wiki answers of 15 total score in 20 of top 40 tags	514 awarded
 Guru	Accepted answer and score of 40 or more	26.4k awarded
 Nice Answer	Answer score of 10 or more	374.8k awarded
 Good Answer	Answer score of 25 or more	92.3k awarded
 Great Answer	Answer score of 100 or more	11.5k awarded



Do they behave any differently?

# Survey

## Motivation to contribute

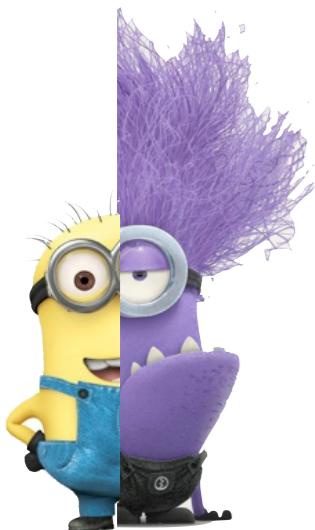


In case of Stack Exchange,  
the **reputation ratings** are  
a nice little incentive.

It's a **game**, which also  
serves a **good purpose**.



**Peer recognition/**  
**gamification within**  
**Stack overflow**



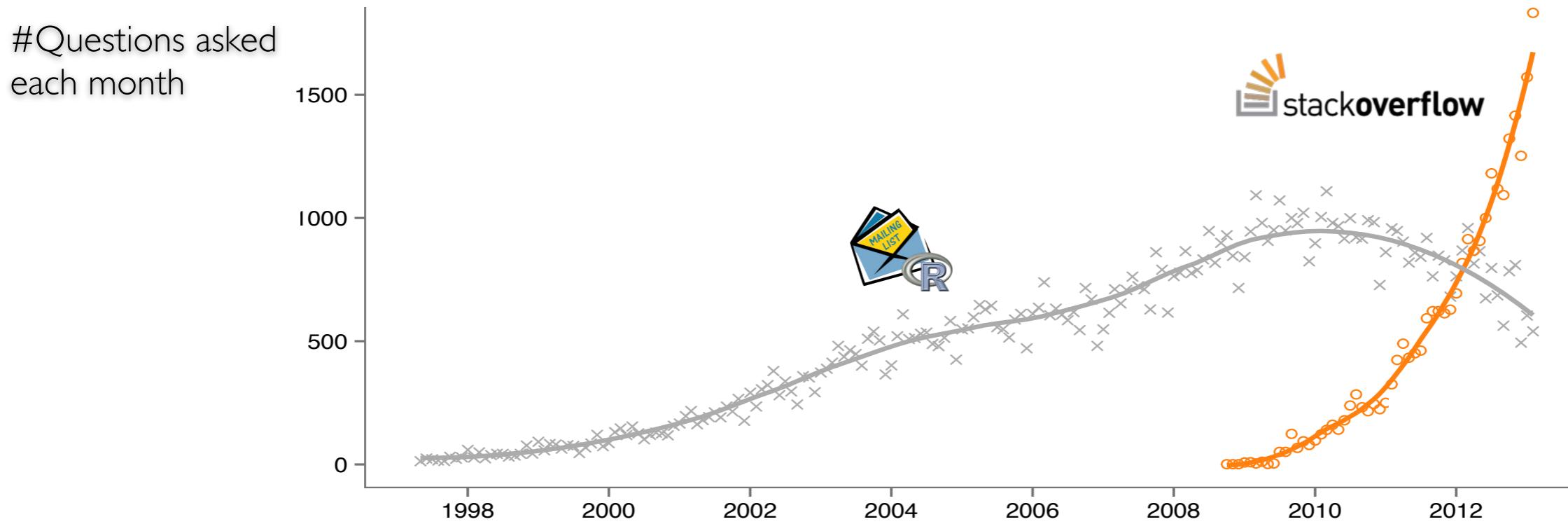
Do they behave any differently?



@b\_vasilescu

# Research goal

How did knowledge sharing in the  community change with the emergence of StackExchange  Q&A sites?

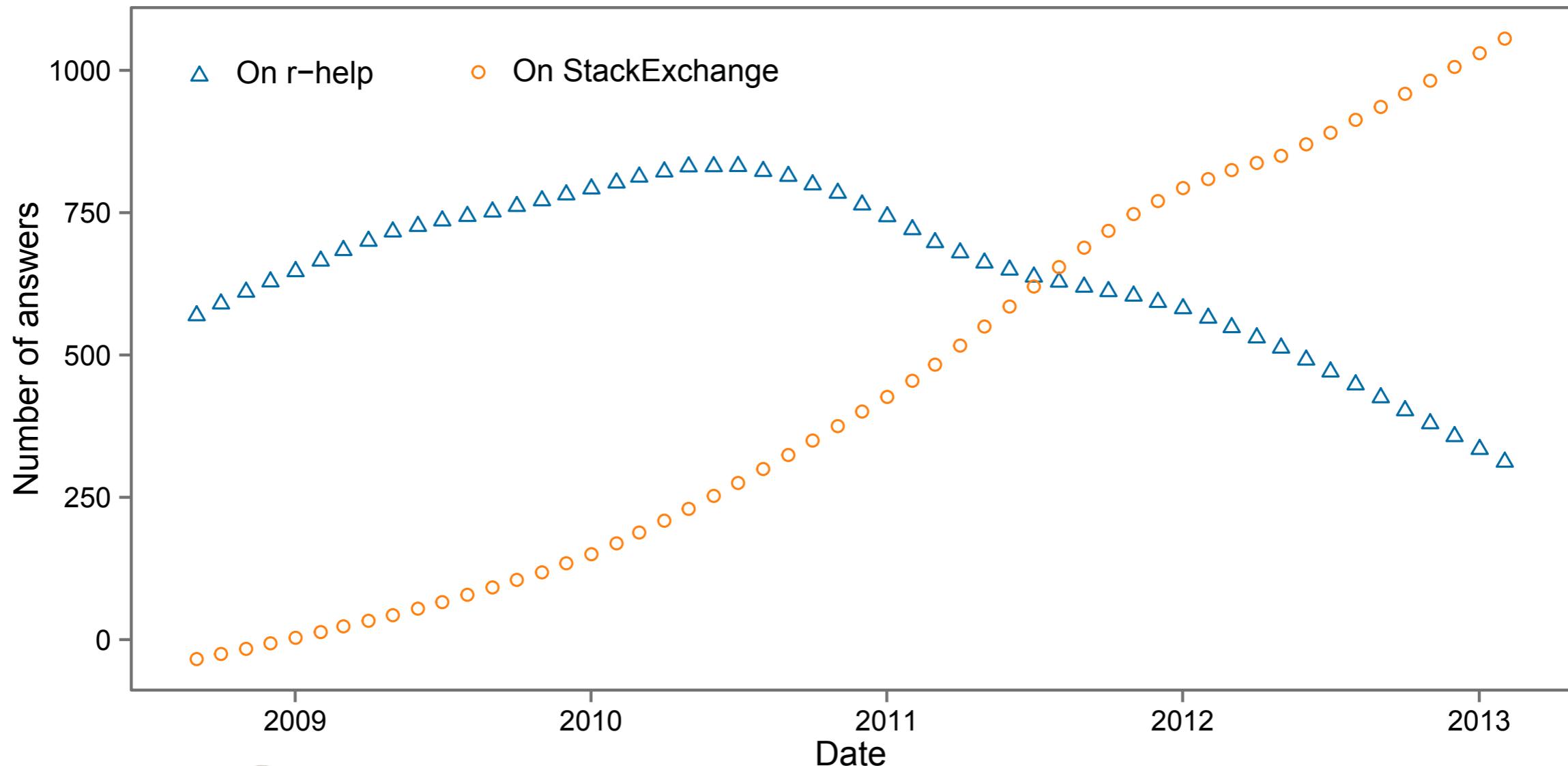


@b\_vasilescu

# Transition to StackExchange



Activity of r-help answerers also active on Stack Exchange



# Survey

## Disengagement from the mailing list



Google is getting better at finding answers related to R so I use it more. I rely less on going directly to mailing lists now.

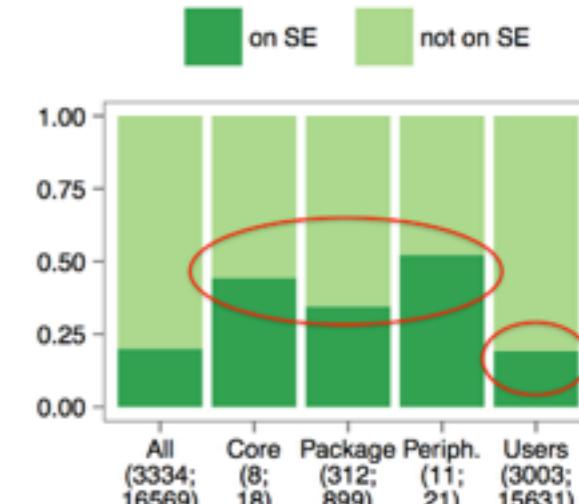
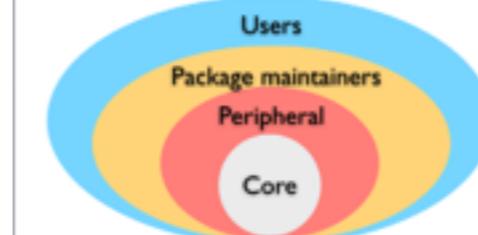
r-help used to be very helpful. But as the number of posts has gone up, I find that reading it is not as useful as it had been.



# Research goal



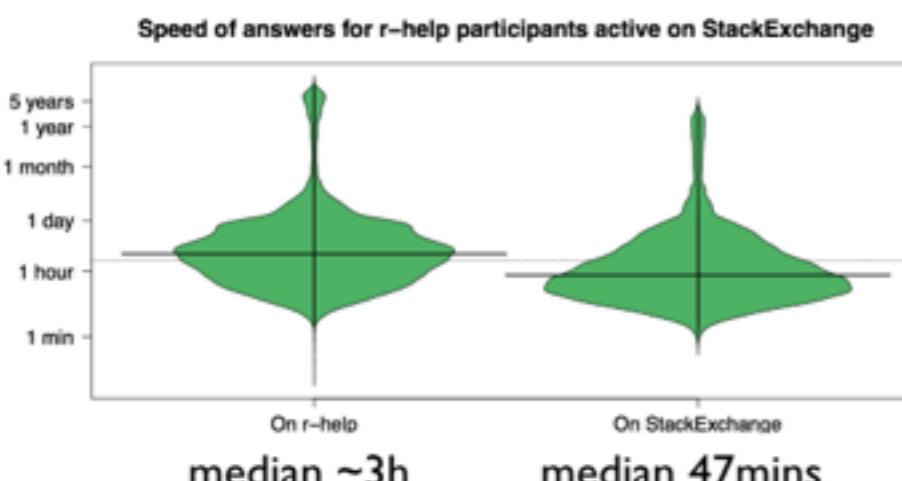
## users on StackExchange



Devs vs users: 1.6x-2x  
higher chance of being on  
**StackExchange** 



## The same users answer faster on StackExchange



Do they behave any differently?

## Survey Motivation to contribute



In case of Stack Exchange,  
the **reputation ratings** are  
a nice little incentive.

It's a **game**, which also  
serves a **good purpose**.



Do they behave any differently?