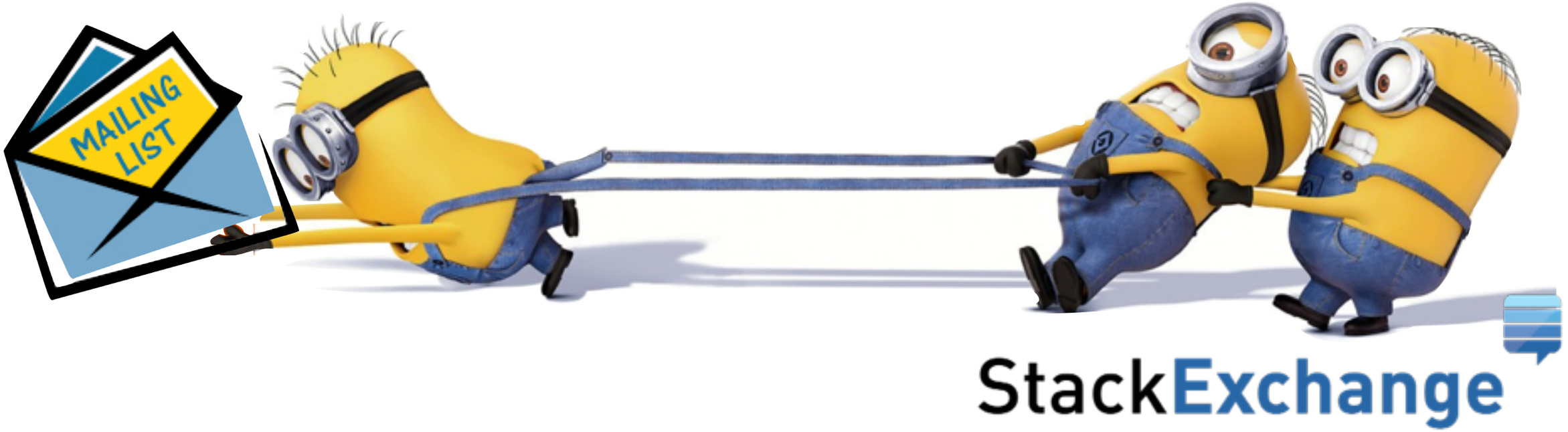


Bogdan Vasilescu, Alexander Serebrenik
Eindhoven University of Technology
@b_vasilescu, @aserebrenik

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

CSCW 2014, Baltimore, Maryland, USA

User support over the years

Dave

(Data analyst)

Dave's problem

How can I adjust p-values for multiple comparisons in R?

User support over the years

Ask a colleague

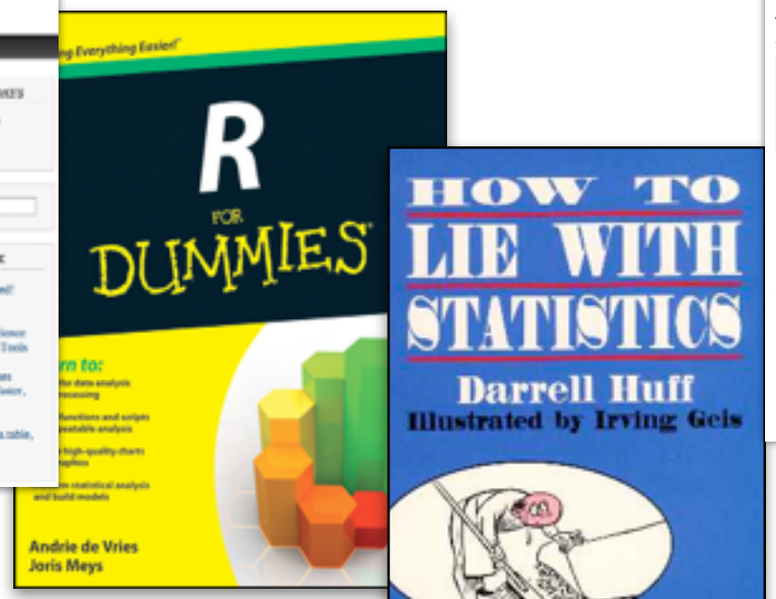
User support over the years

Ask a colleague

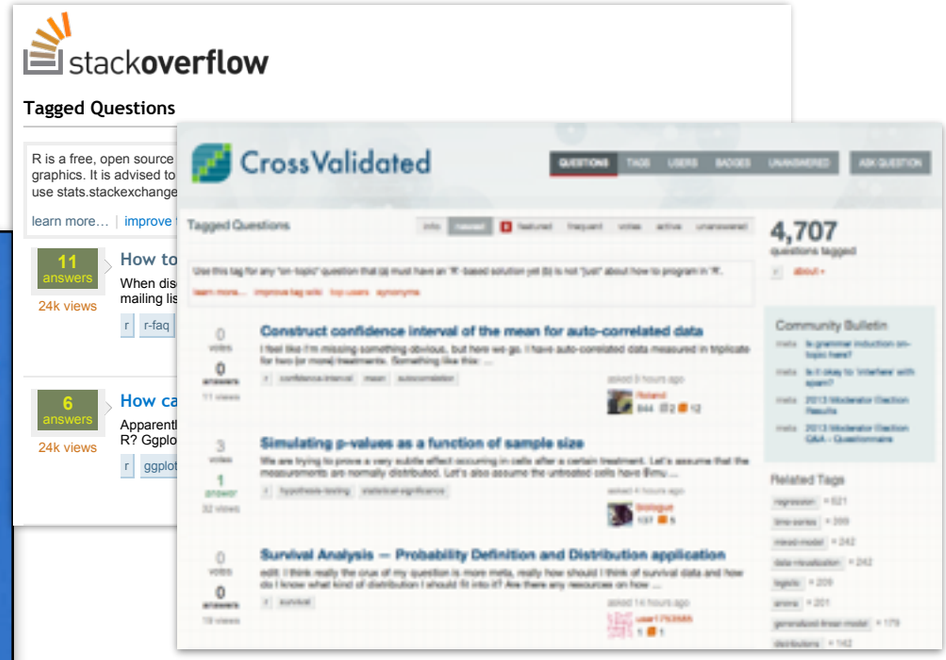
@b_vasilescu

User support over the years

Blogs

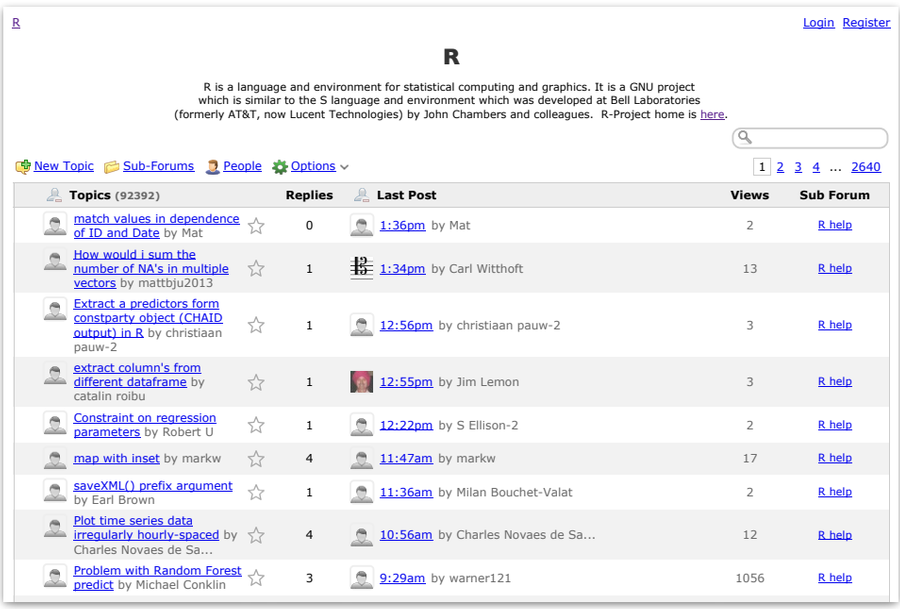


Books



Stack Exchange Q&A sites

Mailing lists



Forums

Mailing lists vs. Stack Exchange

December 2003 Archives by thread

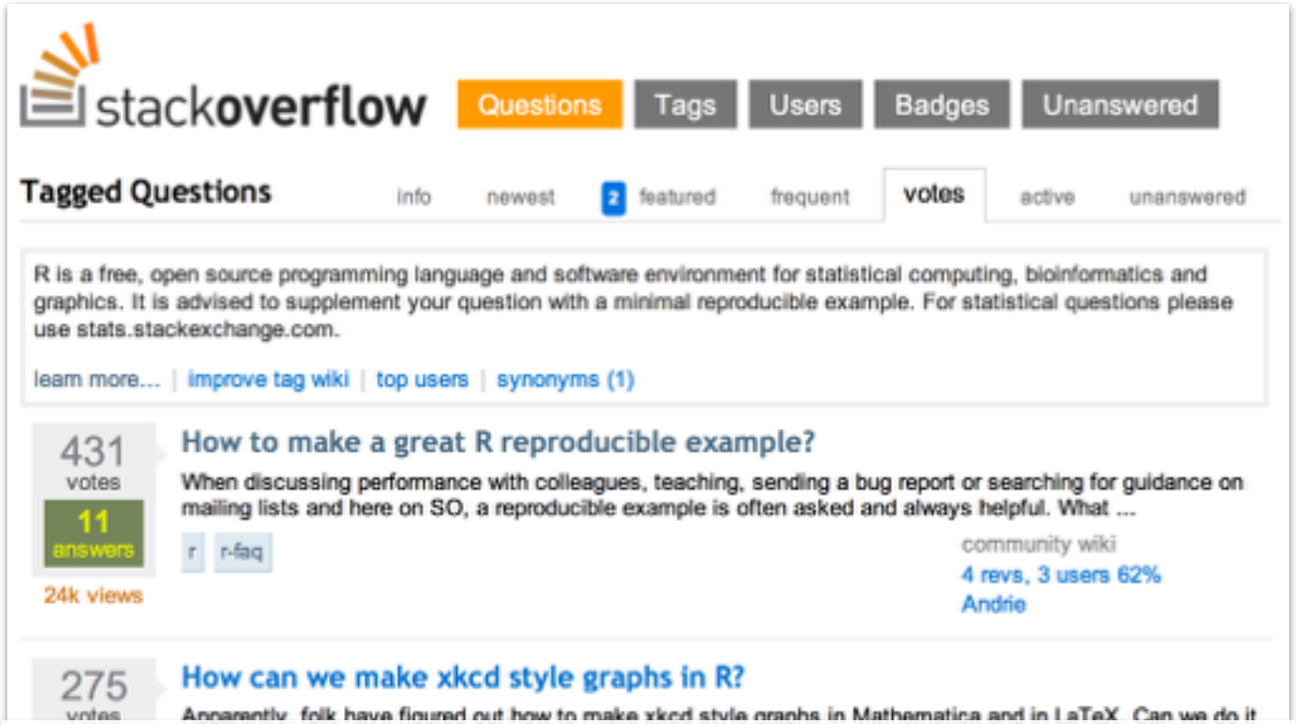
- **Messages sorted by:** [subject] [author] [date]
- **More info on this list...**

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 problema *Torsten Hothorn*
- [R] Indexing ANOVA table *Pascal A. Niklaus*
- [R] wilcoxon-pratt signed rank test in R - drug-effiacy *Niels Steen Krogh*
  - [R] wilcoxon-pratt signed rank test in R - drug-effiacy *Peter Dalgaard*
    - [R] wilcoxon-pratt signed rank test in R - drug-effiacy *Knut M. Wittkowski*
- [R] Changing distance scale in plclust() *Mike White*
- [R] Rd Files? *Wolski*
  - *...il at entelnet.bo*
    - [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



Changing
collaborate, learn and communicate among themselves and with their users

[Begel

Fastest Q&A site

Good technical solutions

Research goal

How did knowledge sharing in the R 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-effiacy *Niels Steen Krogh*
  - [R] wilcoxon-pratt signed rank test in R - drug-effiacy *Peter Dalgaard*
    - [R] wilcoxon-pratt signed rank test in R - drug-effiacy *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*

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)

431 votes, 11 answers, 24k views
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 — community wiki, 4 revs, 3 users 62% Andrie

275 votes, 6 answers, 25k 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)? ...
r ggplot2 — asked Oct 1 '12 at 14:22, jebyrnes 1,615 •3 •7 •12

250 votes, 17 answers, 15k 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 ...
performance r statistics nlp — community wiki, 7 revs, Brian Campbell

236 votes, 11 answers, 156k views
How to sort a dataframe 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 — asked Aug 18 '09 at 21:33, Christopher DuBois 5,856 •7 •28 •70

@b_vasilescu

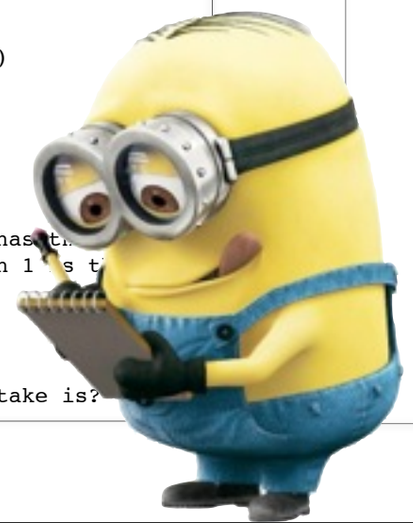
Mailing lists vs. Stack Exchange

[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]

```
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$color by row.names, and force it to has
values that data1. Next it generates a new dataframe where column 1
row.name, and then sort it by the row.name from data1.
data.frame(merge(data1,data2$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?
```



stackoverflow

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

11

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

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]

```
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

stackoverflow

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

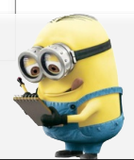
answered May 6 '12 at 1:28
DWin
82.9k ●3 ●39 ●102

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

Modern (fancy) user interface



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]

```
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

stackoverflow

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

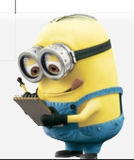
answered May 6 '12 at 1:28
DWin
82.9k ●3 ●39 ●102

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

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]

jdanielnd jdanielnd at gmail.com
Fri Oct 7 15:34:33 CEST 2011

Author

He

I sure if I got it

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?

stackoverflow

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

11

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]
}
```

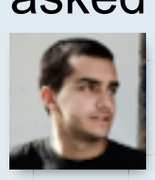
So, as you can see, I have to use both the `x[i]` element, but also the `x[i-1]` element. By using the `*apply` functions, I just see how to use the `x[i]`. Is there anyway I can avoid the `for` loops?

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

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 2:21

Recognition

Research goal

How did knowledge sharing in the R 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-effiacy *Niels Steen Krogh*
  - [R] wilcoxon-pratt signed rank test in R - drug-effiacy *Peter Dalgaard*
    - [R] wilcoxon-pratt signed rank test in R - drug-effiacy *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*

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)

431 votes, 11 answers, 24k views
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
community wiki
4 revs, 3 users 62%
Andrie

275 votes, 6 answers, 25k 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)? ...
r ggplot2
asked Oct 1 '12 at 14:22
jebyrnes
1,615 •3 •7 •12

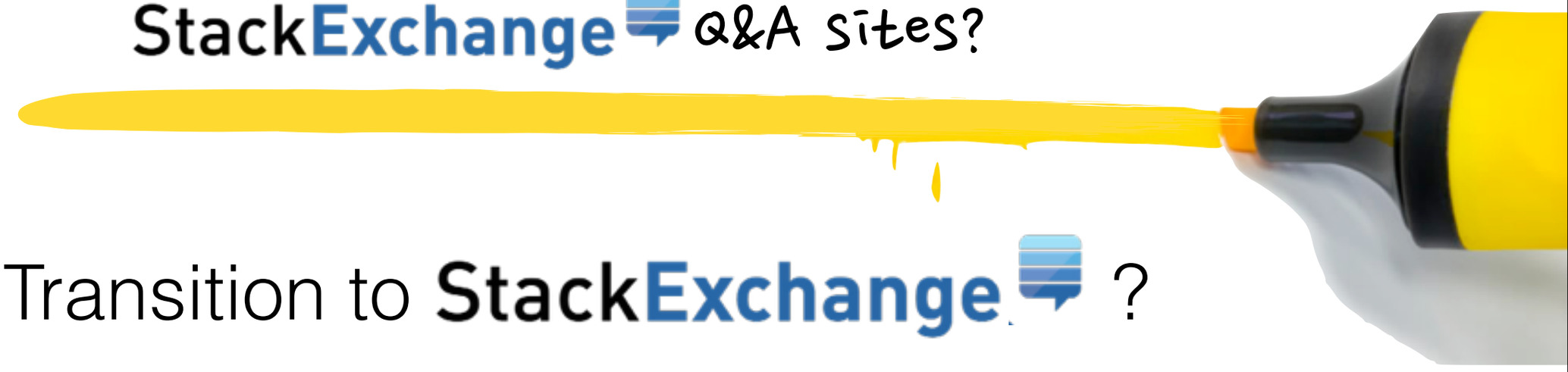
250 votes, 17 answers, 15k views
What statistics should a programmer (or computer scientist) know? [closed]
I'm a programmer with a decent background in math and computer science. I've studied computability, graph theory, linear algebra, abstract algebra, algorithms, and a little probability and statistics ...
performance r statistics nlp
community wiki
7 revs
Brian Campbell

236 votes, 11 answers, 156k views
How to sort a dataframe 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
asked Aug 18 '09 at 21:33
Christopher DuBois
5,856 •7 •28 •70

@b_vasilescu

Research goal

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



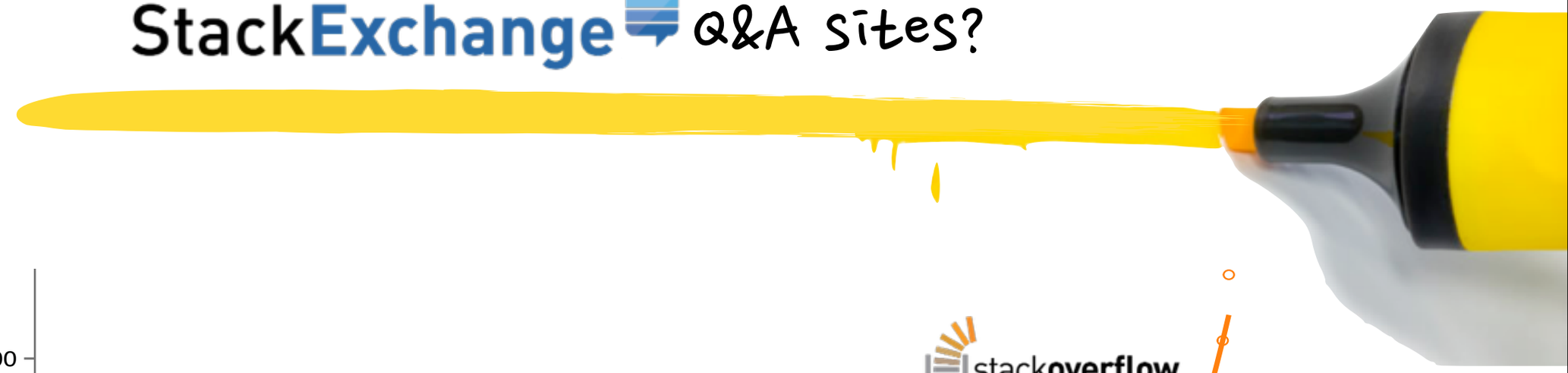
Transition to StackExchange ?

"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."

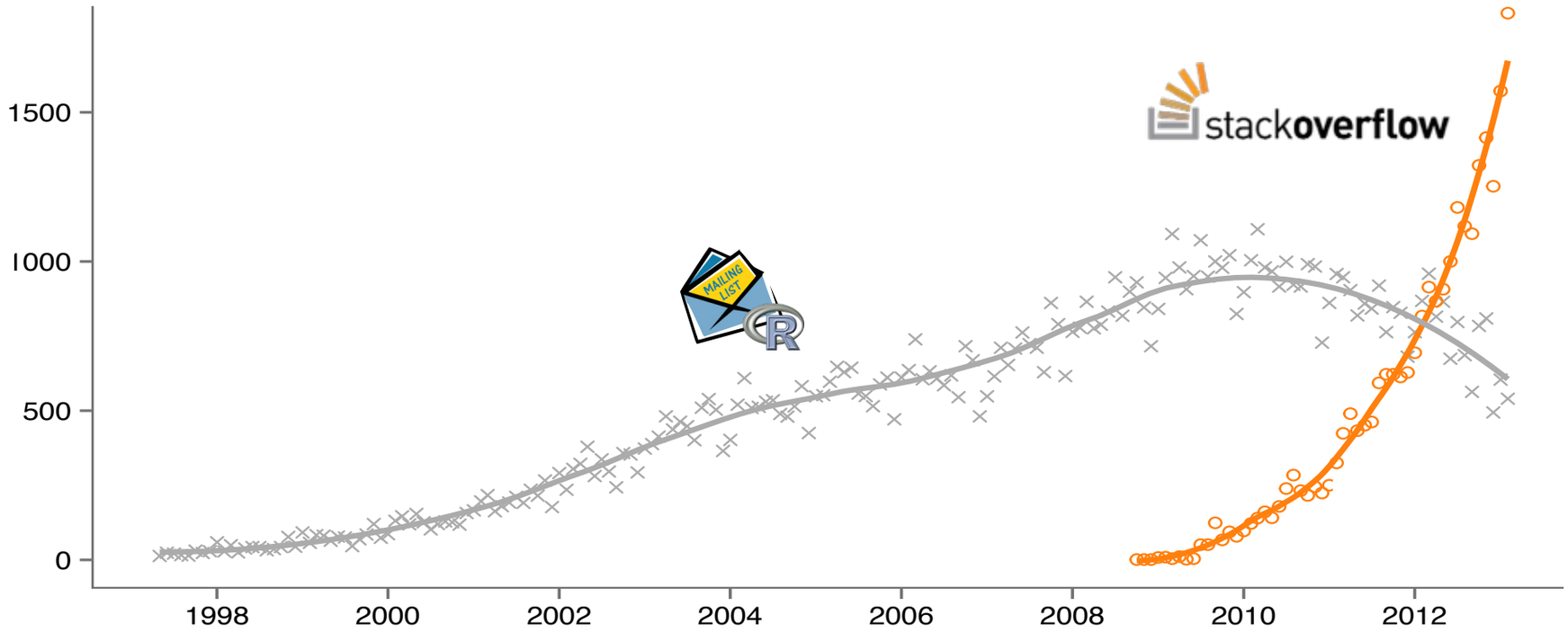
[2010 blog entry]

Research goal

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

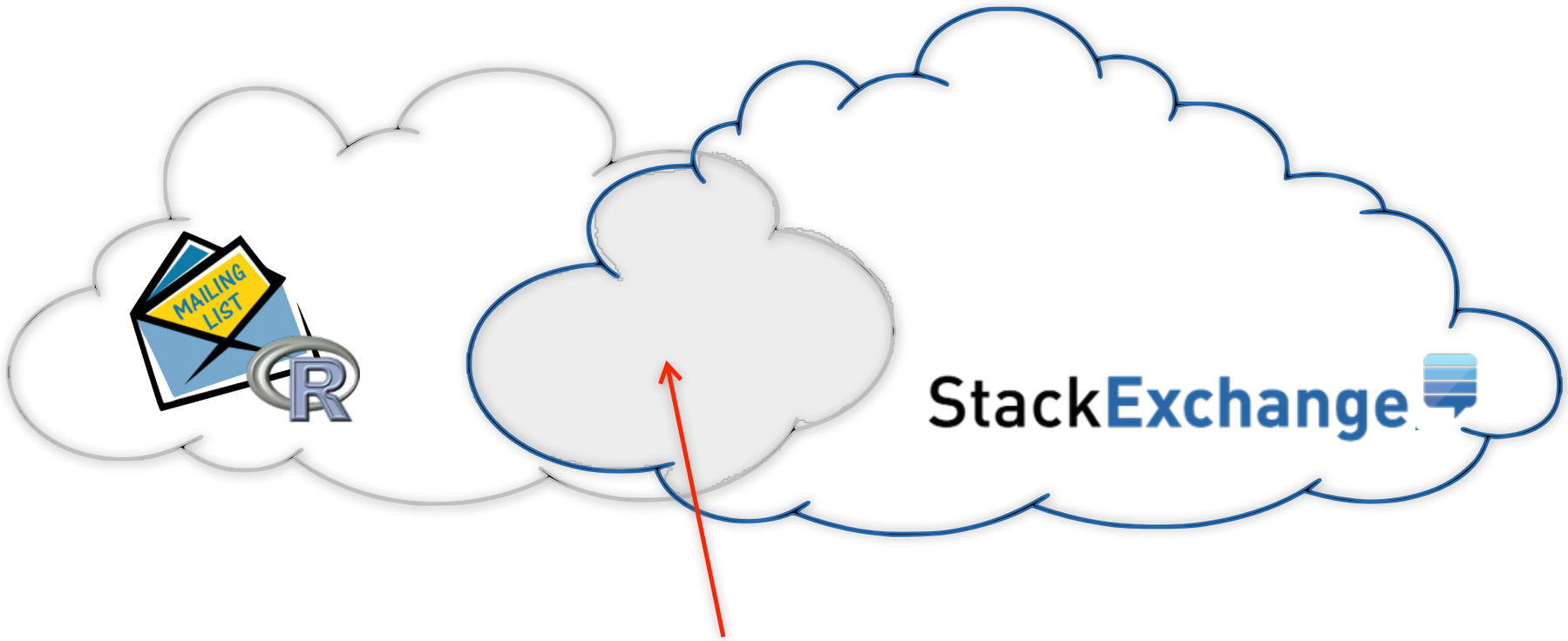


#Questions asked each month



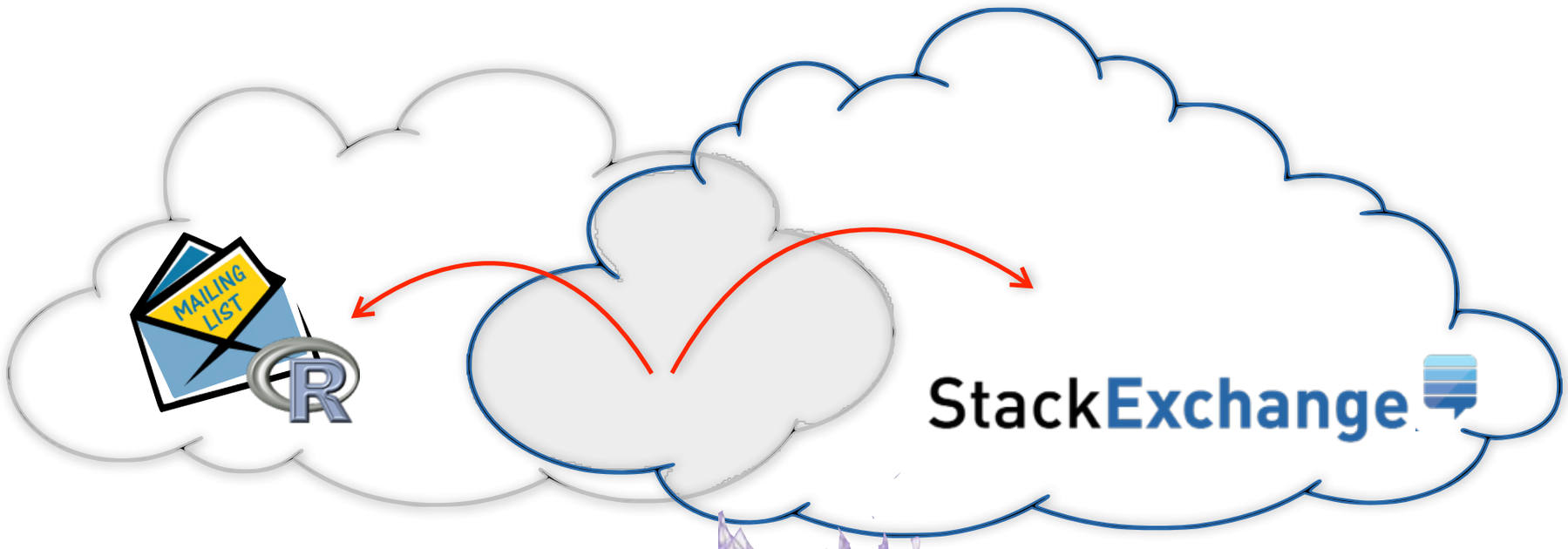
@b_vasilescu

Today



1 Who are they?

Today



1 Who are they?

2 Do they behave any differently?

Today

1 Who are they?

2 Do they behave any differently?

Mining repositories

+

User survey (112 resp.)

Mining challenges

Laurent Gautier

Laurent Gautier - laurent@cbs.dtu.dk

@b_vasilescu

Mining challenges

Laurent Gautier

Laurent Gautier - laurent@cbs.dtu.dk

@b_vasilescu

Mining challenges

Laurent Gautier

Laurent Gautier - laurent@cbs.dtu.dk

Laurent Gautier - s010592@student.dtu.dk

@b_vasilescu

Mining challenges

Laurent Gautier

Laurent Gautier - laurent@cbs.dtu.dk

Laurent Gautier - s010592@student.dtu.dk

@b_vasilescu

Mining challenges

Laurent Gautier

Laurent Gautier - laurent@cbs.dtu.dk

Laurent Gautier - s010592@student.dtu.dk

Laurent - lgautier@gmail.com

@b_vasilescu

Mining challenges

Laurent Gautier

Laurent Gautier - laurent@cbs.dtu.dk

Laurent Gautier - s010592@student.dtu.dk

Laurent - lgautier@gmail.com

@b_vasilescu

Mining challenges

stackoverflow

Cross Validated

Laurent Gautier

Laurent Gautier - laurent@cbs.dtu.dk

Laurent Gautier - s010592@student.dtu.dk

Laurent - lgautier@gmail.com

- lgautier@altern.org

…

9 email addresses, 3 names

Identity merging is far from trivial!

@b_vasilescu

Mining challenges

Laurent Gautier - laurent@cbs.dtu.dk

Laurent Gautier - s010592@student.dtu.dk

Laurent - lgautier@gmail.com

- lgautier@altern.org

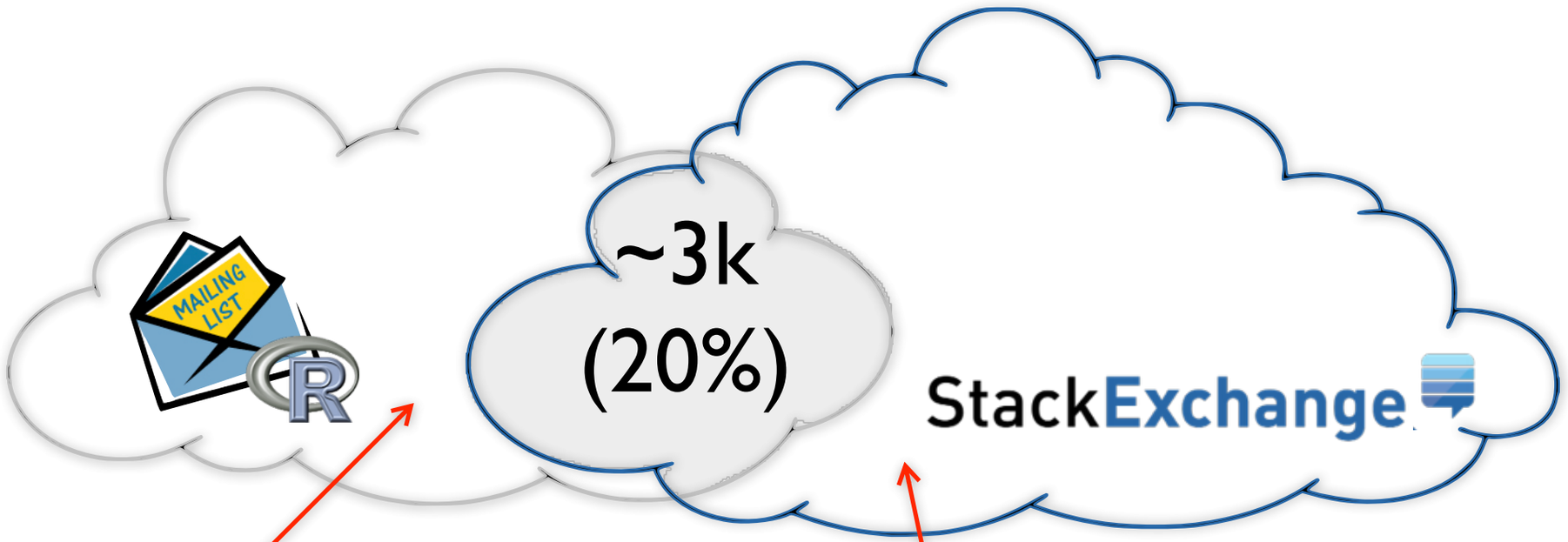
…

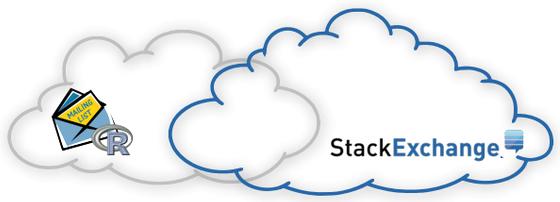
MD5(lgautier@gmail.com)
= 8e5422a173711c086b685140dbc498fe

@b_vasilescu

The R community

> September 2008

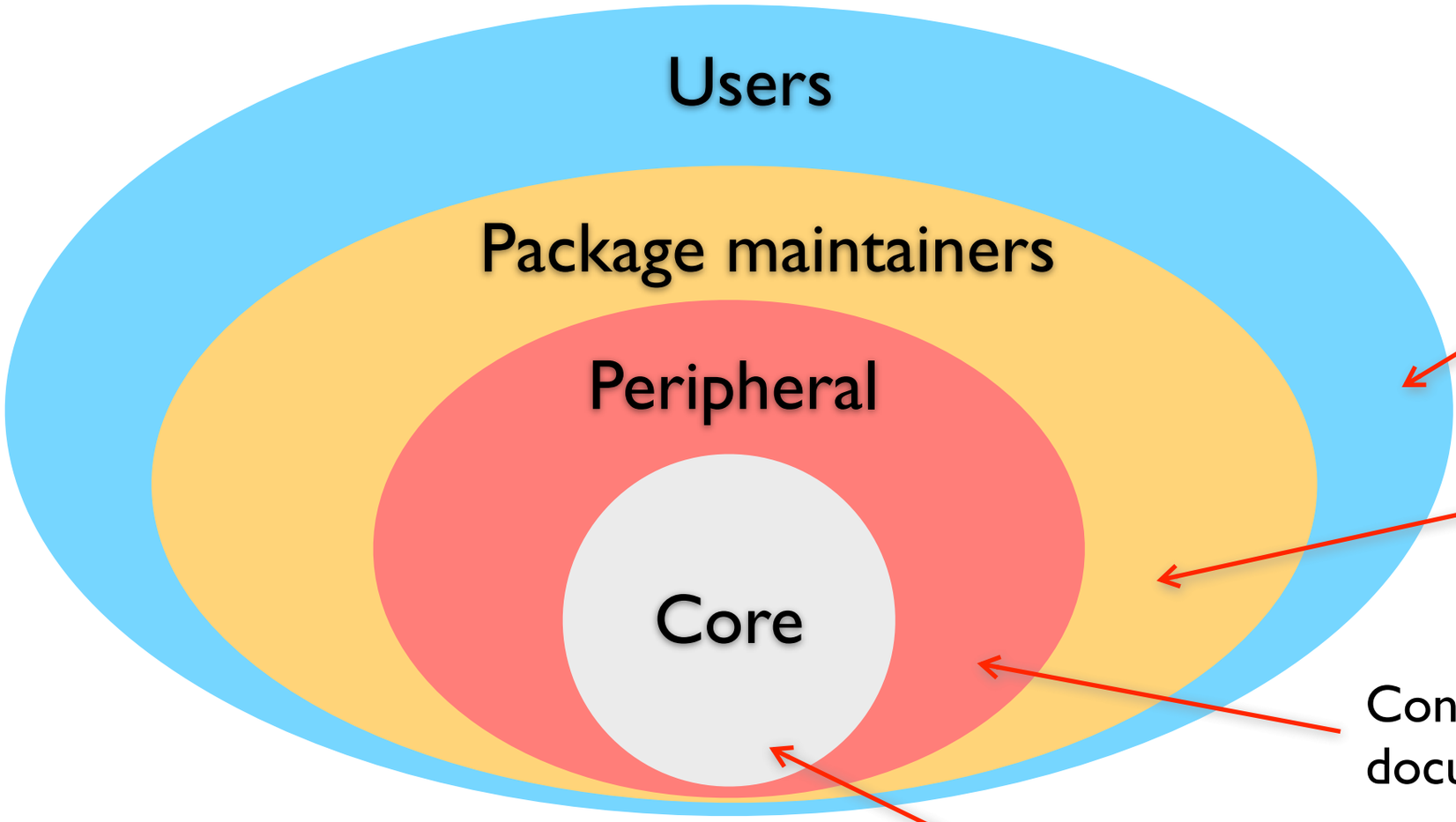




Who are they?

@b_vasilescu

The R community

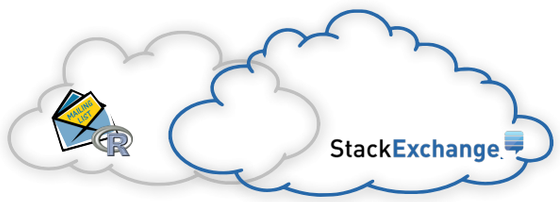


Others

Author or maintain R packages (~2600)

Contribute code, bug fixes and documentation (~40)

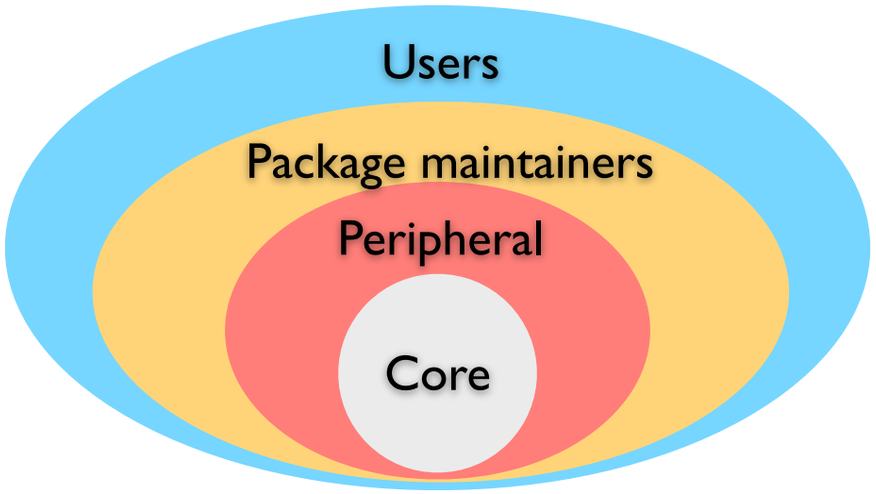
Have write access to the R source (~20)

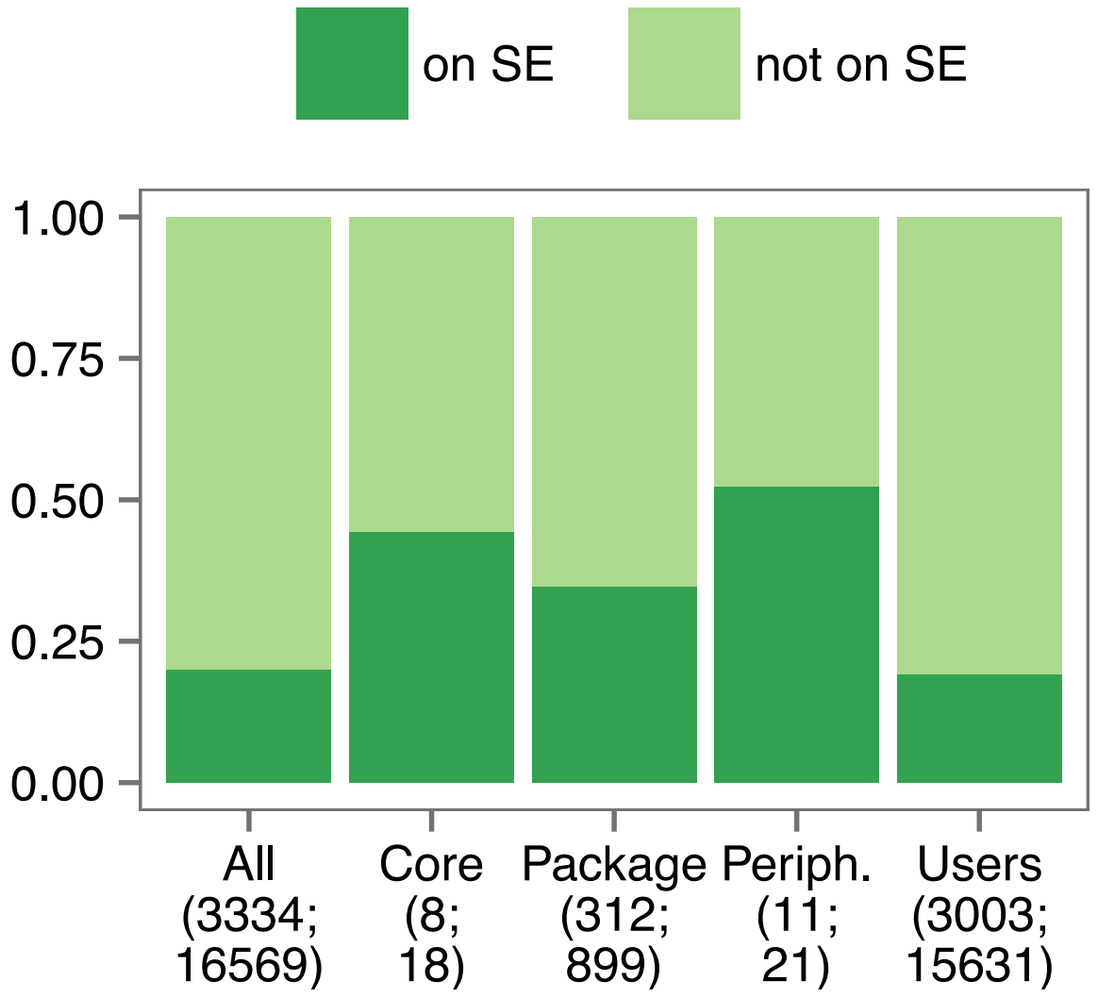


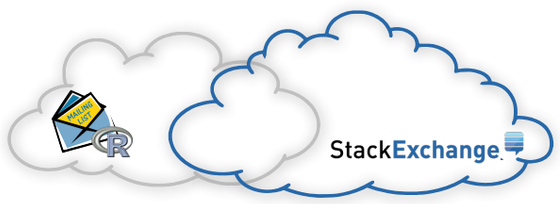
Who are they?

@b_vasilescu

users on StackExchange



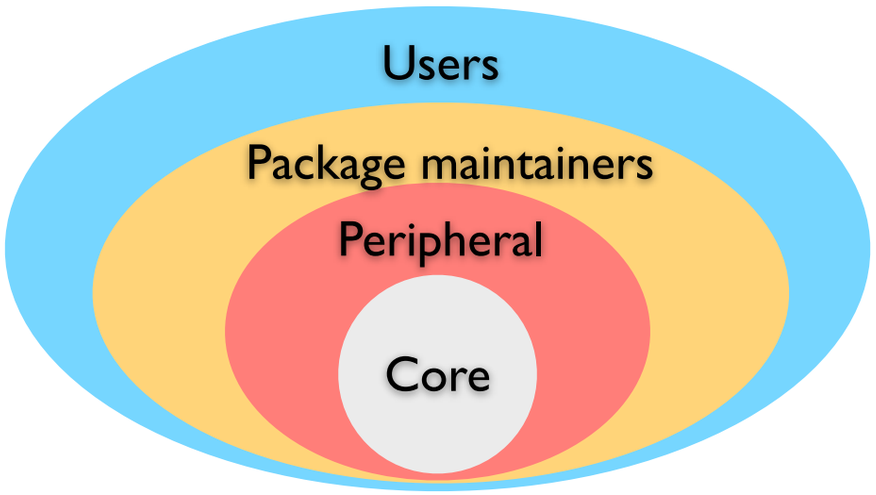




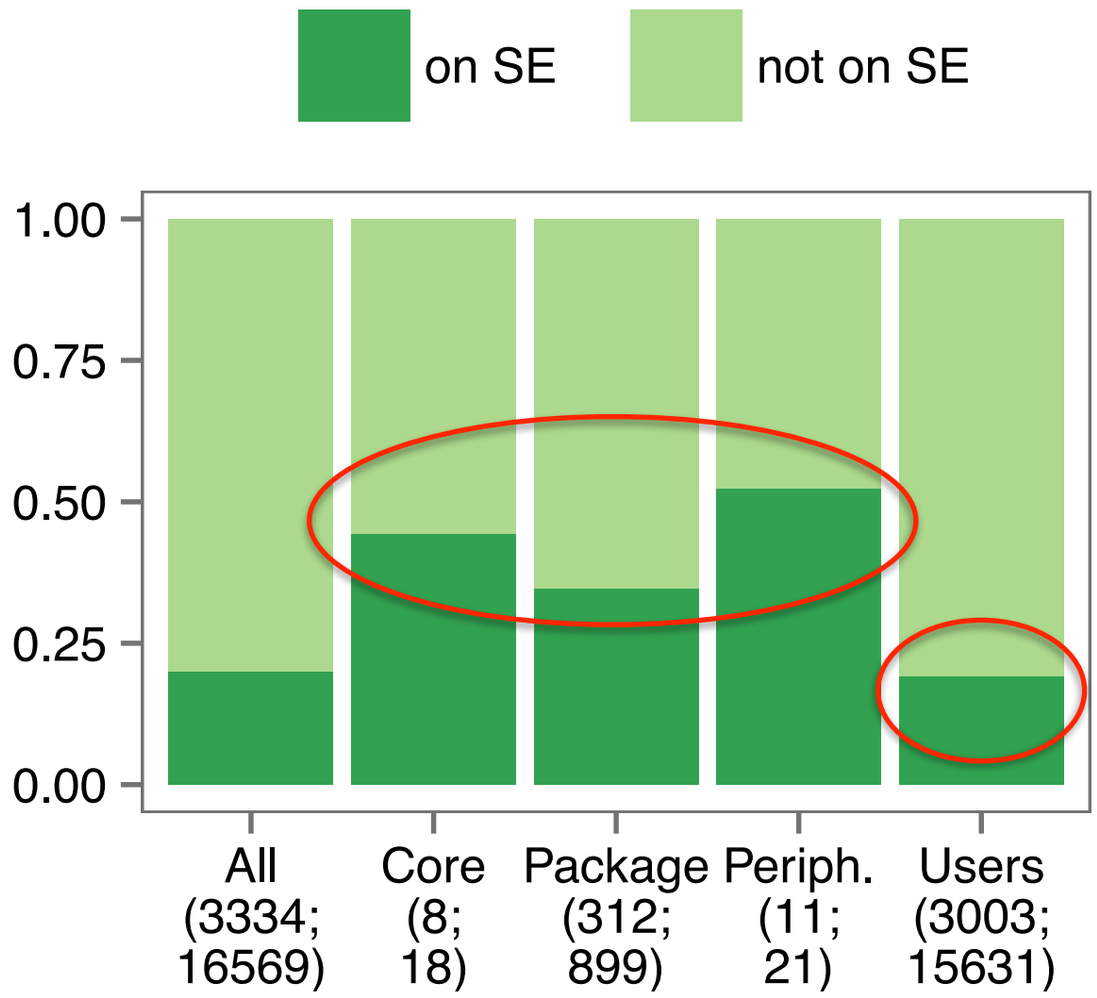
Who are they?

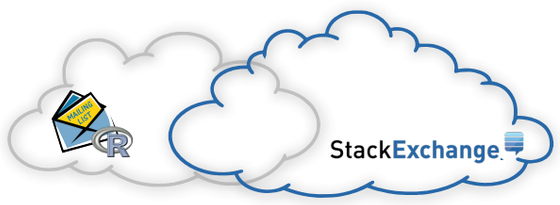
@b_vasilescu

users on StackExchange



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



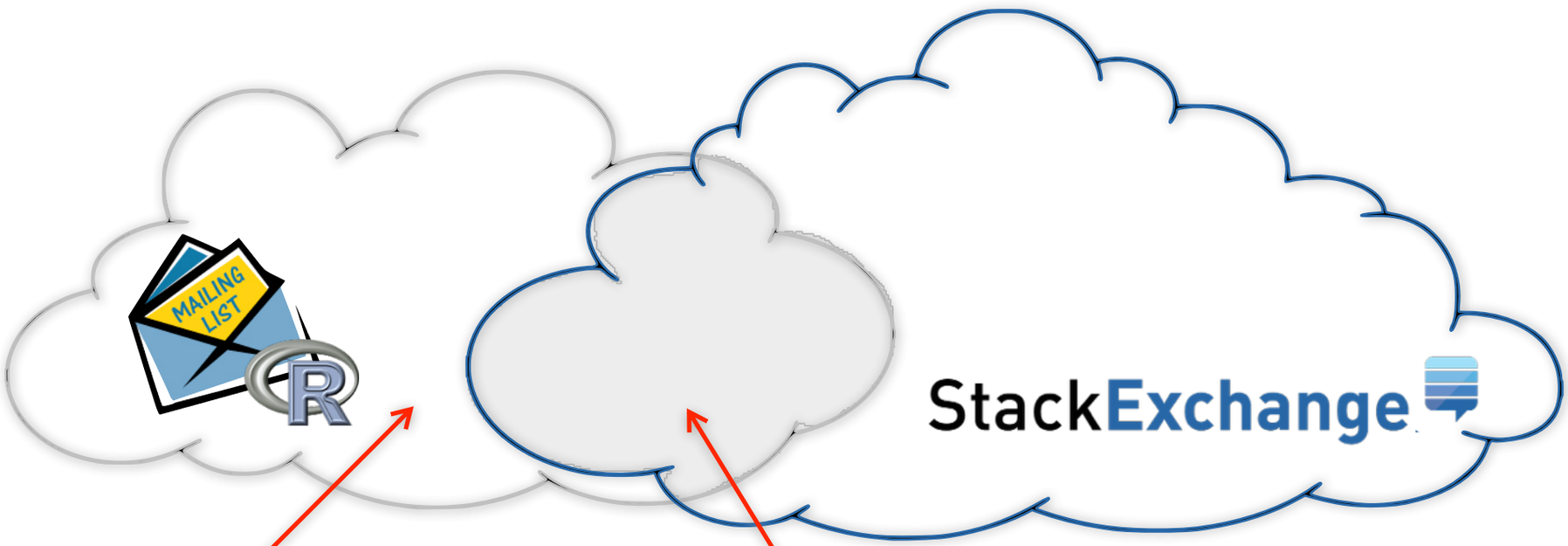


Who are they?

@b_vasilescu

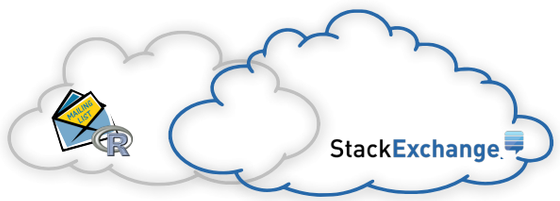
users on StackExchange

Consistently more active



median 1 thread answered

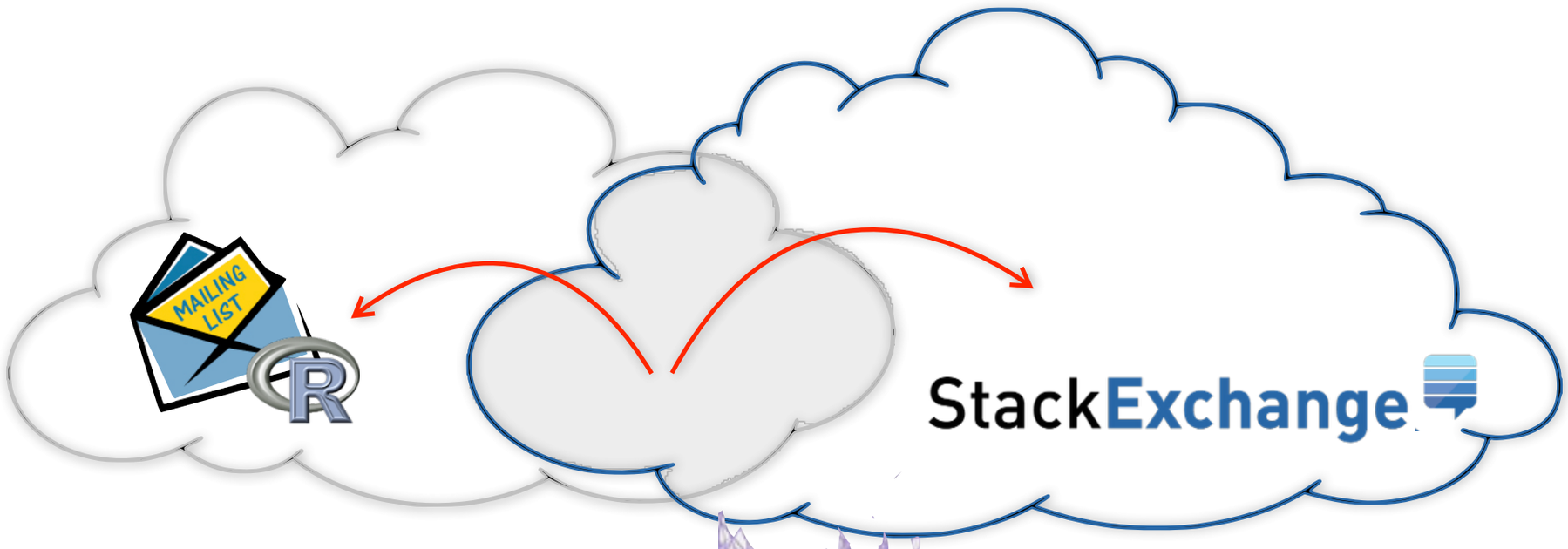
median 4 threads answered



Who are they?

@b_vasilescu

Today

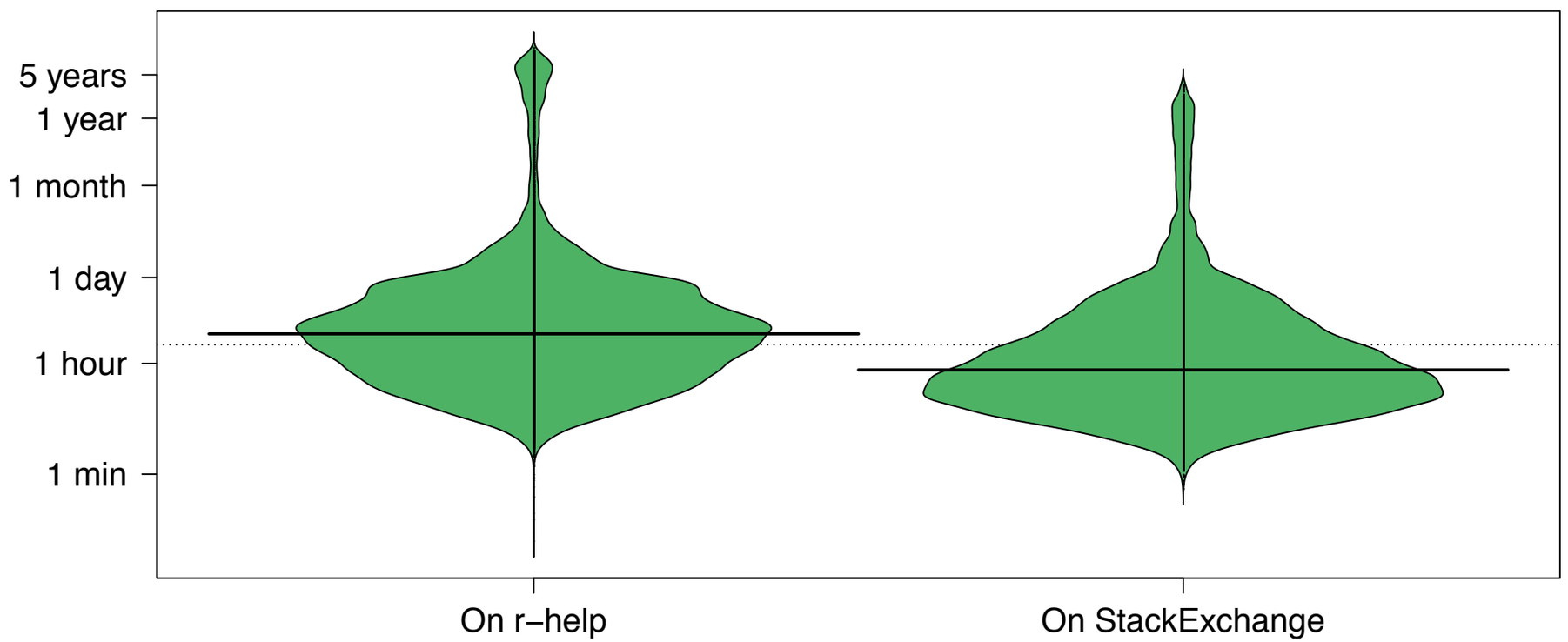


1 Who are they?

Mostly devs, consistently more active

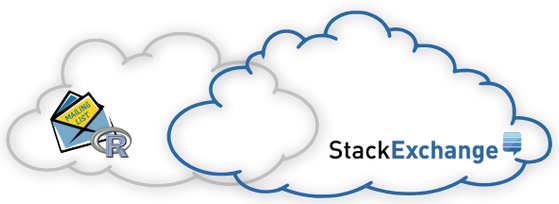
2 Do they behave any differently?

The same MAILING LIST R users answer faster on StackExchange



median ~3h

median 47mins



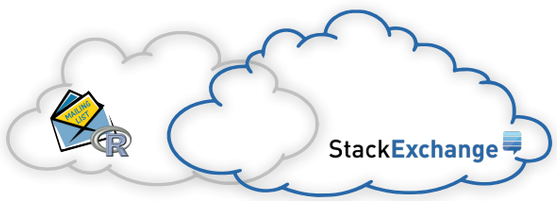
Do they behave any differently?

@b_vasilescu

The same MAILING LIST R 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?

@b_vasilescu

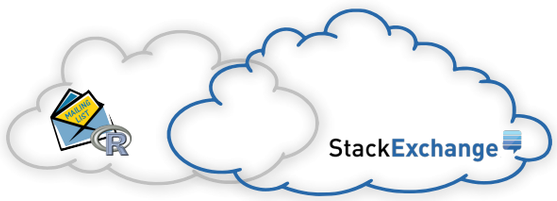
Survey

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

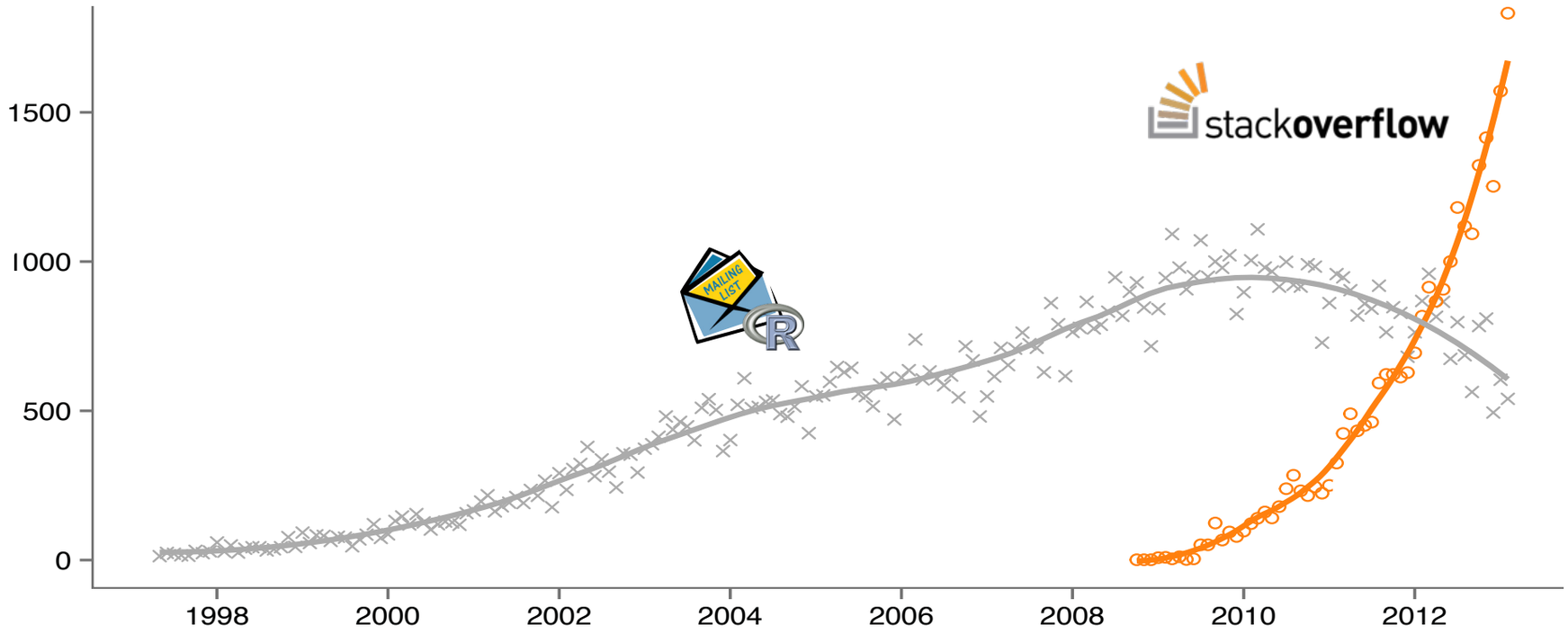


Do they behave any differently?

Research goal

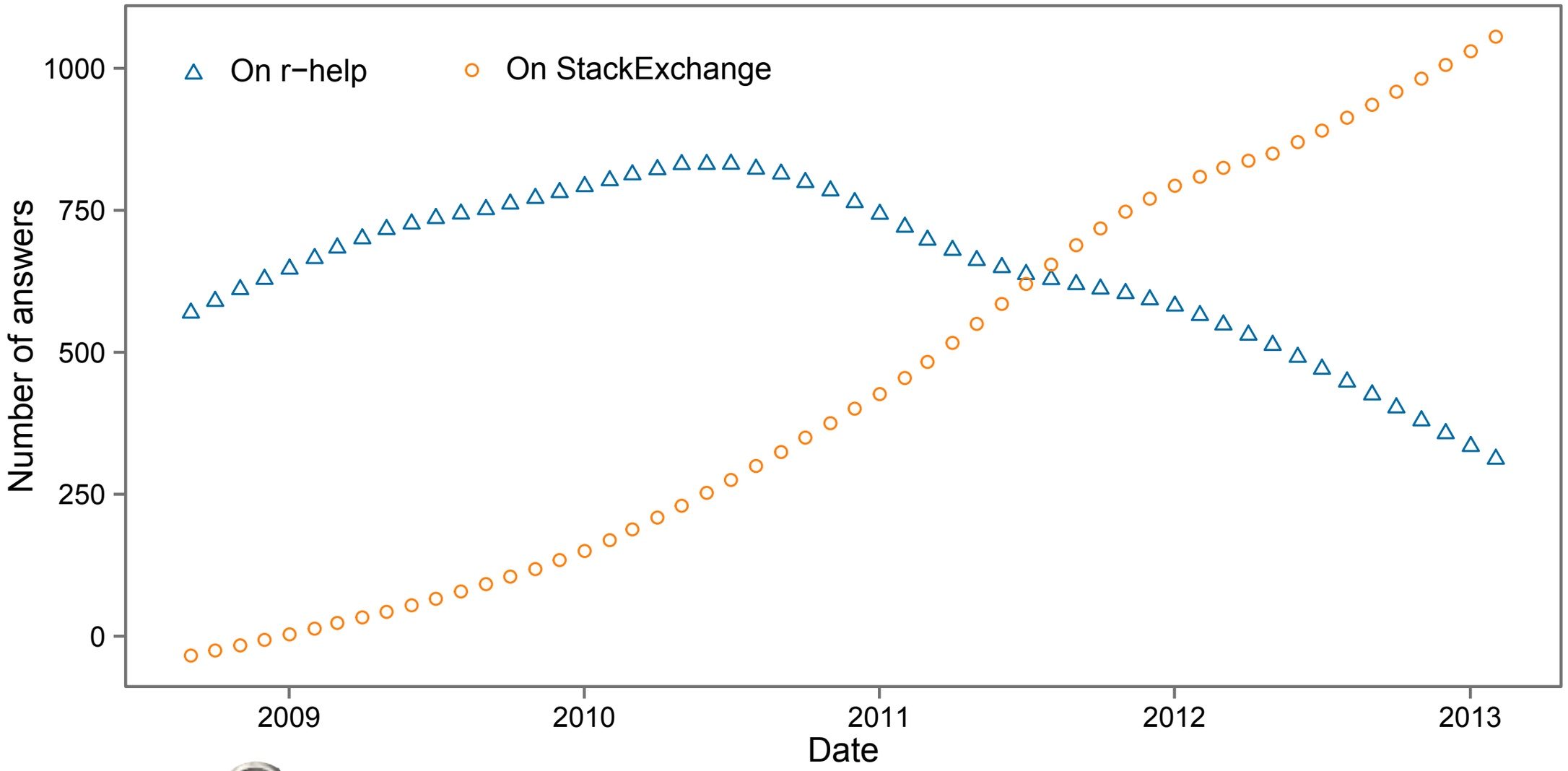
How did knowledge sharing in the R community change with the emergence of StackExchange Q&A sites?

#Questions asked each month



@b_vasilescu

Transition to StackExchange



@b_vasilescu

Survey

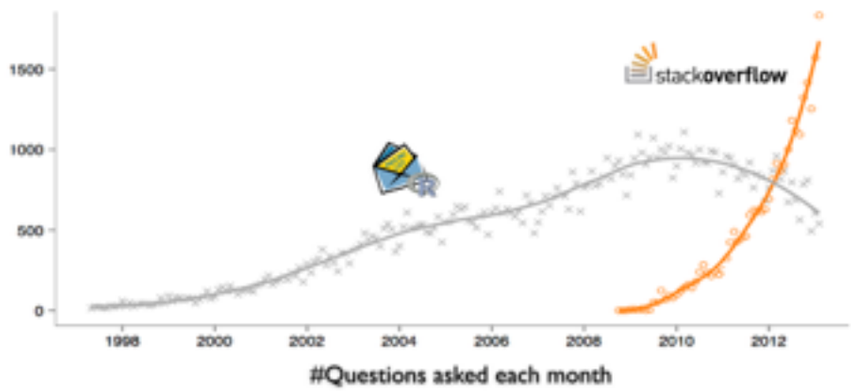
Disengagement from the mailing list

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.

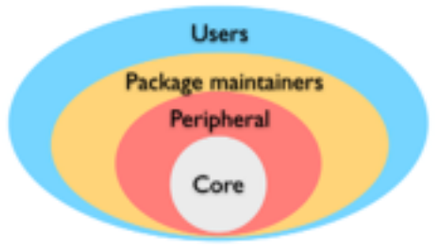
r-help used to be very helpful. But as the number of posts has gone up, I find that reading it is not as useful as it had been.

Research goal

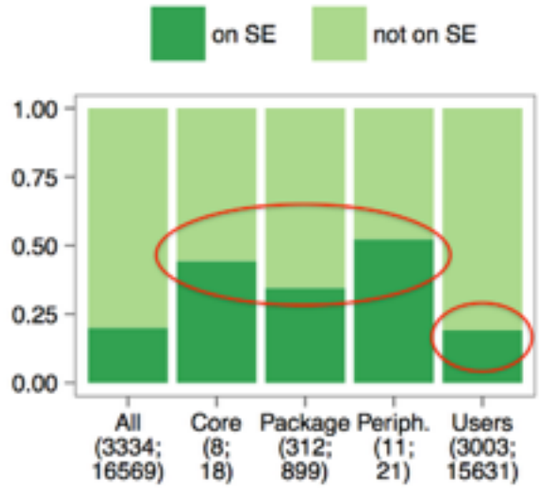
How did knowledge sharing in the community change with the emergence of StackExchange Q&A sites?



users on StackExchange



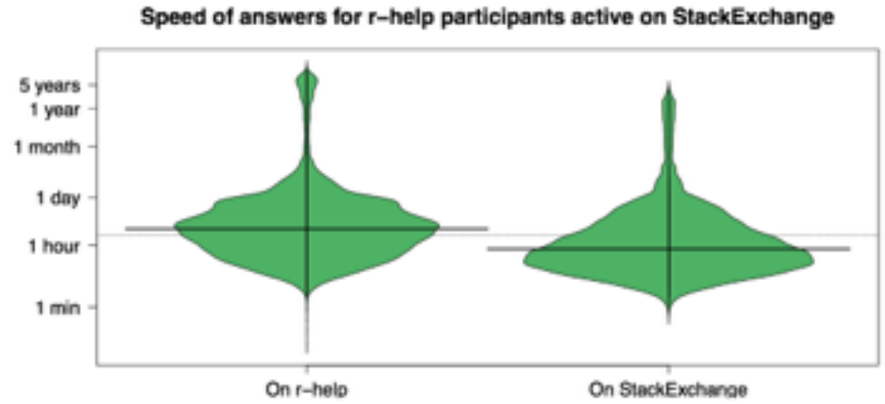
Devs vs users: 1.6x-2x higher chance of being on StackExchange



1

Who are they?

The same users answer faster on StackExchange



median ~3h

median 47mins

2

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

2

Do they behave any differently?