





How many Stan users are there?





Dec '19

6

6

Dec '19

6 V O

6

Dec '19

It seems we could look at a number of pre-existing data sources (eg discourse views and contributors. papers, StanCon attendance etc) to inform an inference of how many people use Stan (and/or use things that use Stan). We could also generate new data (eg via surveys etc). Do we know the answer and/or how best to work it out?



lauren

s.maskell

lauren

I've thought a tonne about surveying the Stan community or potentially capture-recapture techniques, but generating new data in this way would be a lot of work and my current funding wouldn't cover it (that'll change next year though!). I'd be interested in collaborating if anyone is interested. :)

As is hopefully obvious, i am keen to help. I can probably deploy people on this who can help make this happen if we know what it is that we needed them to do.

avehtari Stan Developer Dec '19

How about asking who has registered to discourse? From that we could extrapolate the population size for those who we could reach by survey?

s.maskell Dec '19

@avehtari: who would we survey and how would we get the survey to them? I feel I am being dim.

From memory the hope was to utilized a snowball survey (once you finish you forward onto other people in the population), which is generally what you use when you have a hard to define population. One of the stats grads here, Jonathan Auerbach and I talked about starting multiple snowballs (Andrew's blog, Discourse, Twitter, StanCon mailing lists, etc.) and then tracking what snowball people were recruited by (and if multiples) as a way of measuring coverage.

1 Reply V 200

Dec '19 andrewgelman

This is a fascinating statistics question. Maybe I could post a blog on this and see if there are any thoughts.



Dec '19 That's a fantastic name!!!

@andrewgelman you can, it'd be interesting to see what folks think

Bob_Carpenter €

Dec 2019 1 / 23 Dec 2019

Jan 2020



What's your definition of a Stan user? Is it someone who once downloaded Stan and ran a model? Or is it someone who uses Stan regularly, and if so, how regularly? What about people who use packages like brms or rstanarm or prophet that are built on top of Stan?

There are over 3K users registered on Discourse, but that doesn't mean they're regulars. Many of them only showed up once.

1 🗘 🕜



andrewgelman

Blog post on the topic scheduled for Monday.

Dec '19





s.maskell

Dec '19

For context, UK academic departments (eg the one I sit in) are assessed (as part of the "Research Excellence Framework", REF) on the basis of some criteria: the next assessment is imminent, but the one after that will be in 2026. When aggregated over each entire University, the outcome of that assessment process modulates the amount of (Quality-related Research, QR) income that the University receives from the UK government. So, it's important to inform the assessment process with pertinent information.

One of the three assessment criteria is "impact", which relates to the uptake of academic research outside of the academic discipline it came from and is metricated in terms of "reach and significance" during the census period. Unfortunately, "reach and significance" is not defined quantitatively. However, the notion of the census period is defined as people using the research during a specified period of time (eg 2021-2026).

The specific motivation for my question is that I'd like to understand how we (locally to my department) could quantify the "reach and significance" of any enhancements to Stan that (we hope!) might come out of our work between now and 2026. It seems natural to start by finding out what we (as the Stan community) know about how big we are now.

So, in answer to your question, I think I'd ideally like to know how many people and/or organisations are making use of specific subsets of Stan's code (including in any packages that use Stan) during a specified period. I'd also like to know where they are based geographically, whether people work in academic, industry or government, the demographic of applications they are working on, etc etc.

That's clearly hinting at scope creep, but hopefully helps explain the specific reason I asked the question, which I see as an important step towards quantifying the impact of our work to help enhance Stan.





I think this would be relevant outside of the REF as well - a useful addition to many grants/impact sections.



s.maskell



So, in answer to your question, I think I'd ideally like to know how many people and/or organisations are making use of specific subsets of Stan's code (including in any packages that use Stan) during a specified period. I'd also like to know where they are based geographically, whether people work in academic, industry or government, the demographic of applications they are working on, etc etc.

We were at the point where we were focussing on questions around this, plus an emphasis on barriers to entry. I can dig them up if there's renewed interest. I believe the conclusion last time was that the time-cost of doing it wasn't worth the expected benefit, but it could have changed! :)







andrewgelman

Dec '19

I posted my question here: https://statmodeling.stat.columbia.edu/2019/12/09/how-many-stan-usersare-there/ 48







avehtari Stan Developer

Dec '19

This paper might have useful ideas Using an Online Sample to Estimate the Size of an Offline Population 5 how to estimate the number of those users who would be unlikely to be reached by survey.





mitzimorris ®

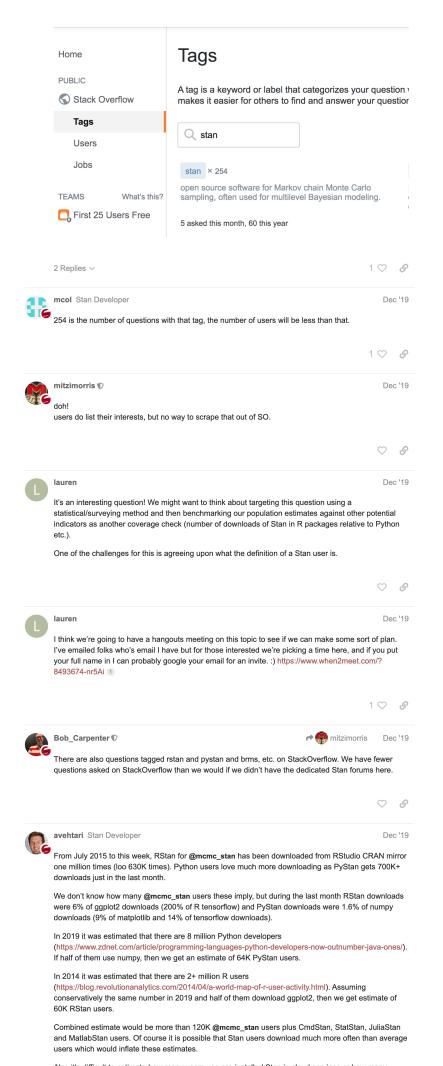
Dec '19

Stack Overflow has tag "Stan" and 254 users (if I'm interpreting their graphics correctly) have listed Stan as things they're interested in. about once a week someone asks a Stan question

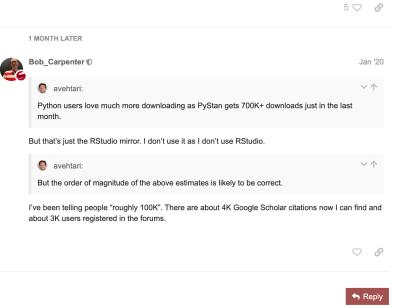


Products





Also it's difficult to estimate how many users use pre-installed Stan in cloud services or how many "users" are students who are forced to use Stan in a course (e.g. https://github.com/avehtari/BDA_course_Aalto/) and then never use it after that. But the order of



Suggested Topics

Topic	Replies	Views	Activity
Hello,I am new user and I need help.Inside E[i, m] what will it add each time?! The first time and with (j in 1:1) it will add E [1, m] (I'm not sure). The next ones?! ■ General	4	76	Nov '20
Creation of a bunch of small paid jobs to help Stan operate ■ General	5	523	Dec '19
Issues with building a package with Rstan ■ General ■ rstan	1	172	May '20
Stan for multi-omics data integration ■ General ■ rstan	8	213	Oct '20
Constraints on parameters/data General	3	74	Dec '20

Want to read more? Browse other topics in ■ General or view latest topics.

WinBUGS and Friends Downloads

These count downloads from the R packages that either are interfacees to Stan, winBUGS, JAGS, Nimble or OpenBUGS or are directly dependent on these packages.

```
library(tidyverse)
library(knitr)
library(cranlogs)
```

Specify packages and get download statistics.

```
packages <- c("rstan","rstanarm","MetaStan","StanHeaders","R2WinBUGS","R2jags","rjags","runjags","rstantools","jag
sUI","nimble","nimbleEcology", "nimbleSCR","nimbleSMC","R2OpenBUGS","BRugs")
dat <- cran_downloads(package = packages, from = "2013-08-01", to = "2020-12-31")</pre>
```

Downloads by package:

kable(dat %>% group_by(package) %>% summarise(downloads=sum(count)))

package	downloads
BRugs	141688
jagsUI	109241
MetaStan	13784
nimble	64031
nimbleEcology	8559
nimbleSCR	1738
nimbleSMC	663
R2jags	326454
R2OpenBUGS	166839
R2WinBUGS	342887
rjags	929072
rstan	2040196
rstanarm	528787
rstantools	695631
runjags	166738
StanHeaders	1766344

Total downloads:

kable(dat %>% summarise(downloads=sum(count)))

downloads