


How likely are you to encounter snark on Stack Overflow? Part 1: Comparison of statistical software

PNIS Editorial Staff¹

1 - PNIS

After `compress` ing my data, I have several string variables with storage type `str4` or `str1` and format `%9s` . I would like to revert them all to the default display format, which `help dformat` reports should be `%#s` for `str#` . Is there a quick way to do this?

asked Oct 8 '14 at 17:09

 F 5,567 ● 1 ● 9 ● 32

"structure of my best guess" -- not what I actually expected to work. I wanted to demonstrate that I had thought through part of the solution (as is commonly expected on SO). Sorry if that was unclear. – F Oct 8 '14 at 17:41

"Stata doesn't truncate display strings just because they don't match the display format" Ah, that's good to know! Thanks – F Oct 8 '14 at 17:42

1 Try this: `di %1s "F asks lots of Stata questions" .` – N Oct 8 '14 at 17:43

Note: This conversation was edited for length. The entire exchange can be found [here](#).

INTRODUCTION

Stick Overflow (SO) is an online forum, launched in 2008, where over [3.6 million registered users](#) pose and answer questions that are, for the most part, related to computer programming. As of this writing, there have been over [8.9 million questions asked](#) and SO users are estimated to have answered 92% of them ([Mamykina et al. 2011](#)). Alexa currently ranks SO as the [58th most popular website in the world](#).

To some users, the conversation pictured above between the questioner “F” and the answerer “N” is a typical example of an SO entry. Typical not in that the question was answered and accepted by the original poster, but that the answerer used a certain level of snark bordering on insulting. Instances of rude or hostile answers, as well as accusations of overt censorship on SO appeared quickly after the site went live (e.g., [Bishop 2012](#)). In fact, it only took 5 months before SO moderators were labeled as Nazi’s

(FSK 2009). The atmosphere became so rancid that the site’s founders Joel Spolsky and Jeff Atwood proclaimed that the summer of 2012 would be the Summer of Love, with the goal of restoring SO as a “welcoming, friendly place” and hunting out the snark (Spolsky 2012).

The success of the Summer of Love could be called tepid at best. Not one month afterwards, bloggers continued to call out the hostility on SO (e.g., Hadaway 2012, Sergworks 2012). And judging by this steady stream of blog posts (Mendoweb 2013, Richter 2013, Schreiber 2013, Sammaye 2014) and Quora questions (one, two, three, four, five, six), the hostility seems to have remained.

So how bad is it really at SO? A mega survey was conducted by Spolsky and Atwood as a prelude to the Summer of Love (the Spring of Inquiry?). Seven thousand randomly selected comments were rated by 20 experienced raters. Overall, 75% of these comments were rated positively (Rogers 2012), with 8% judged as really friendly (that is, rated as “nice” by 75% or more of the raters), and 2.3% judged as really unfriendly (rated as “unfriendly” by 75% or more of the raters) (Spolsky 2012). So, you are 25% likely to get a neutral or negative comment on your question, and you (or your mother) are 2.3% likely to get called a ‘fucking lamer’.

Our goal in this paper was similar to Spolsky and Atwood’s, except that we were interested in determining if snark rates differed based on what type of statistical software (e.g., stata, SAS, R, SPSS) the question was regarding. Do certain statistical programs breed a snarkier user?

METHODS

We focused on 6 of the most popular statistical programs (Muenchen 2012) that employ a command-line interface: MATLAB, R, SAS, SPSS, Stata and Wolfram/Mathematica. We searched for each program tag at SO and randomly selected 100 questions. Then, we judged each question/answer set for two criteria:

1. Was there a snarky/insulting/rude comment or answer to the question?
2. Was the question accepted by the original poster (indicated by a checkmark next to the answer or if the questioner said ‘thanks’ or something similar)?

RESULTS AND DISCUSSION

We made a table of our results (Table 1), as well as a redundant figure (Figure 1), although we feel Table 1 is easier to read.

For starters, the SPSS folks are the nicest you will ever want to be around. They had the lowest snark rate at 3%, and if we had kept a score of “Most Positive Encouragements”, they would have somehow reached 111%. A typical entry was reminiscent of the *Discover It Card commercials*. The next *Chicken Soup for the Soul* should just be a random collection of Stack Overflow SPSS entries. What we’re trying to say is that they are nice people, probably Canadians.

On the Comcast end of the customer service scale, we have Stata, which ended up with a 14% snark rate, the highest of any of the 6 types of statistical software. Although we didn’t collect data on this, most of the snark seemed to originate from the same user, “N”, quoted at the beginning of this article. However, because N answered so many of the Stata ques-

Table 1. Incidences of snark and number of accepted answers for the six statistical programs.

Program	Snark?	Answer Accepted?		Totals
		No	Yes	
MATLAB	No	28	62	90
	Yes	6	4	10
	Totals	34	66	100
R	No	19	69	88
	Yes	2	10	12
	Totals	21	79	100
SAS	No	32	64	96
	Yes	3	1	4
	Totals	35	65	100
SPSS	No	41	56	97
	Yes	0	3	3
	Totals	41	59	100
Stata	No	22	64	86
	Yes	7	7	14
	Totals	29	71	100
Wolfram/ Mathematica	No	22	74	96
	Yes	2	2	4
	Totals	24	76	100

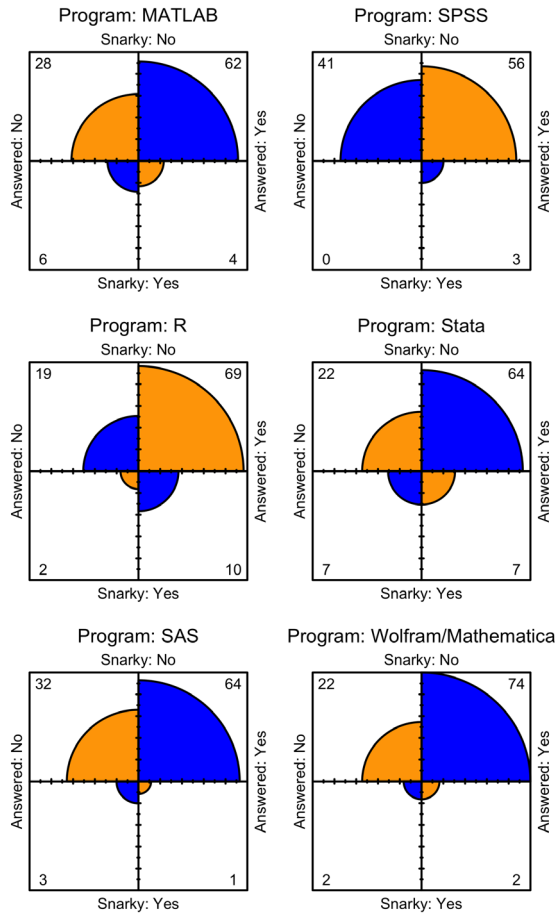


Figure 1. Four-fold plot of results. Radius of each quarter circle is proportional to the frequencies in each specific category.

tions, it was unadvisable to remove N’s answers from our analysis. We should also point out that N knows a lot about Stata and that many times N’s answers were accepted by the original poster. We just wish N didn’t have to be so sarcastic.

There didn’t seem to be much of a relationship between rate of snark and rate of accepted answers. Although R users answered the greatest percentage of questions (79%), 10 of those came with snark. Thus, the winner in this analysis should be Wolfram/Mathematica, whose users answered 76/100 questions successfully, while resorting to snark in just 2 of those answers.

As stated above, Spoolsky (2012) found a 2.3% “really unfriendly” rate in their massive data analysis, while our snark rates averaged about 7.8%. This difference could have arisen from a number of differences between the two studies (such as: sample size, using all possible comments vs. using only comments pertaining to statistical software, ratings done by 20 different people vs. ratings done by us, definition of “unfriendly” vs. definition of “snark”, etc.). But the larger point is that there is still snark on SO, and you are about twice as likely to run into snark on SO than you are to be a [victim of a crime in the United States](#).

It may be that there will always be snark and hostility on SO, no matter how many calendar seasons are dedicated to its eradication. But SO will also likely remain an amazing place to get your questions successfully answered for *free* by professionals, sometimes in the matter of minutes. That might be worth the potential exposure to a little snark.