

Nikoleta Transcript

Katie [00:00:46] Thank you for your time as well, um-.

Nikoleta [00:00:47] No worries, no worries.

Katie [00:00:49] I'm going to start by asking if you have any particular bias towards either language, R or Python.

Nikoleta [00:01:00] Yes, I am familiar with both languages, but I am very biased towards Python because it's the language I use mostly and I'm way more comfortable in Python than in R

Katie [00:01:13] OK, yeah. Um, and with regards to the code, um, I can- I can share them as well, if you want? Um, can share my screen, but did you have any initial sort of comments on the code? Sort of like any similarities or differences.

Nikoleta [00:01:31] Err, yeah similarities a thing, also is the way you the way that you wrote the code, right, the way that it's structured? Um, pretty much if you are more fluent in one language, you can always translate it very easily to the other one. Um, so there were commands I didn't necessarily know what they meant in R, but then I could look to the right to the Python code. It was- so, and I guess that means they're quite similar, they use similar syntax for several things. So, for example, some of the pre-recorded functions such as, you know, changing the scale and things like that. This is not exactly the same, but it's very similar. So, yes. And what was the question? I can find similarities in both, in those two languages.

Katie [00:02:25] Okay, so um. And do you feel any code could be changed in any way, in your opinion?

Nikoleta [00:02:38] Yes, for sure. Again, I'm not too familiar with, with R. So I don't know if I have any effici-. I guess if you argue to me that this is the most efficient way to do this, maybe I would have- I would agree. But if you go to bar plot one, for example, you are overwriting the plot each time, so that's an error, in my opinion. And at some point in both codes, you're referring to these things called names and ntimes, which are not defined anywhere past, even in the functions.

Katie [00:03:25] I think um, so, the point of them was I was kind of- so I defined them higher up, and I'd done sort of a load of data manipulation. But, erm, I spoke to Vince and he said just kind of just leave them out, just get the opinions on the-.

Nikoleta [00:03:38] Okay, fantastic. So my opinion is that the both codes right now, there are things that wouldn't work necessarily, right? Like if I run this it wouldn't work and there is some it seems to be some bugs as in the plots are writing themselves. So you're not returning each plot, so that seems to be- so from functionality, I think the functionality is a bit broken. Now, depending on how things are written, I think both languages, it's very good as in it's very obvious what is the colour is very obvious that you're changing the y labels and the x labels and things like that. So, yes.

Katie [00:04:25] OK.

Nikoleta [00:04:29] But the way your code is broken, the language themselves make them a bit more readable, I guess, and, yeah, yeah, I think that's, my humble opinion.

Katie [00:04:44] And sort of based on these, as well as about your own programming knowledge, how well suited do you think each language is to visualisation. Or sort of in particular matplotlib and ggplot.

Nikoleta [00:05:07] Um, so the problem with R is that the type of plotting that I've personally used in R was more pre-recorded things, pre-recorded things, sorry. So, for example, I used the stat package and it would put the arrows and things like that just with a function. But that made me feel back then that in order to manipulate the plot further, it was a hassle for me because it was a recorded thing and then I had to go on top and force things. Where with matplotlib when I plot, I do everything from scratch. So I have, yes, but that was because of my limited knowledge of R back then. And I do know that ggplot is a very powerful library in R. So how I gather when it comes to plotting both of the languages are quite powerful. And I know, for example, for plotting graphs R is fantastic where I haven't found something equivalent as good in Python. Plotting wise, both of the fantastic people will argue that one is better, but it has to do with bias, in my opinion. I would say matplotlib is fantastic, but I don't know R ggplot, so. Yes.

Katie [00:06:26] Alright, that's great. And this is going to be quite a biased question, but do you feel like ggplot or matplotlib would be easier for someone to learn if they had sort of similar R and python experience?

Nikoleta [00:06:40] Yes, I am biased as in I find- I find Python to be more readable, also, as a person whose English is not the first language, I find Python to be more readable. So for any beginner starting something, I would say that Python, in my opinion, is slightly easier. Now, I'll say yeah plotting with the same.

Katie [00:07:03] All right.

Nikoleta [00:07:08] Yes, because I think with matplotlib it's very clear, right? You have the plt dot or figure axis as each command and you know what you're doing where in R you have these weird plusses and, weird.

Katie [00:07:20] Yeah

Nikoleta [00:07:24] So, yes, I think, like, oh, are we adding things to the things. Yes. Is not as clear to me.

Katie [00:07:34] Yeah I guess cause with ggplot you kind of, again Vince was saying you've got the whole- you have to learn the whole grammar of graphics and stuff, and whatever.

Nikoleta [00:07:42] Yes.

Katie [00:07:43] Yeah, all the pluses weird as well.

Nikoleta [00:07:44] All the pulsses are weird as well, yeah. And I guess you can't necessarily see- so the way that I'm set up, so right now I have bar plot 1 in both languages in front of me. And Fantastic, like, I understand that geom_bar is under ggplot, so it has to be part of ggplot, but apart from the fact that it's underneath it, there's no other

indication that it's part of ggplot plot where with the Python code, I can see that is part of the plt library. You know, no library, but you know what I mean. Everything's after that plotting instance, and I'm like ah, OK, well.

Katie [00:08:23] OK, that's interesting, I hadn't thought of that, because I'm very biased towards R so this is interesting.

Nikoleta [00:08:29] Yes, but yes. So I think that's the thing. And also, you know, yeah that with Python.

Katie [00:08:38] Um, I'm going to show you a couple of the plots now, and it's just getting your opinion on whether you feel like either of them is more sort of publication- produces plots that are more sort of publication ready than the other. Um. So this is just to sort of just the default, well, not default. Sort of like control scale so that I have done nothing to alter scales on this one. Um, and I've tried to sort of other than altering the scales in the plots tried to sort of keep them as close to sort of like, I don't know, default as possible.

Nikoleta [00:09:21] Yes, makes sense. OK, I think a lot of people would argue that the plot on the left looks more beautiful, OK? I personally argue the I would- for my publication I would probably use the plot on the right just because the bar plots are not elevated. They're floating like they're floating on the left. I think if the left plot they were not floating and zero started from zero, I would probably go for the left because I think it looks more appealing.

Katie [00:09:52] Yes, OK.

Nikoleta [00:09:54] But I don't like the floating part.

Katie [00:09:57] Yeah, I see that actually. Um, I will show you these two as well. Oh, I've got the wrong one there, sorry.

Nikoleta [00:10:11] ~~That's okay, no worries. Someone is using Windows, what? No I'm joking.~~

Katie [00:10:20] ~~Huh?~~

Nikoleta [00:10:20] ~~Someone is using windows~~

Katie [00:10:23] ~~Yes, I've never got into the whole sort of Apple thing~~

Nikoleta [00:10:31] ~~Not even the linux thing, Vince didn't?~~

Katie [00:10:34] ~~No, he actually hasn't commented on my use of windows, yet.~~

Nikoleta [00:10:41] ~~Fair, I don't complain, I used to game a lot, I still game I used to game a lot on computer so windows are fantastic for that purpose.~~

Katie [00:10:49] ~~My friends are trying to sort of get me into a gaming at the moment being stuck at home.~~

Nikoleta [00:10:53] ~~You should, it's good, it relaxes you.~~

Katie [00:10:56] And I've been playing Minecraft, which I don't think really counts as gaming.

Nikoleta [00:11:04] It is. No, Minecraft it is counting as gaming. I'll take that, I'll take that.

Katie [00:11:10] Okay, I'm a gamer.

Nikoleta [00:11:14] And said these two are different- they're the same.

Katie [00:11:18] Yeah, so they're both- they both use a logarithmic scale. Um, yeah again, I haven't done anything to. So this is just so I have the control plot to change the scale that a plot uses to logarithmic. And this is the output that it gives without any other sort of.

Nikoleta [00:11:42] OK, then the plot to the left. I think the plot to the right at least needed some sort of explanation, that is the logarithmic of, so yeah. So if these the standard that it returns I think you know for purpose of corrections, I think the left one on these in this case.

Katie [00:12:02] OK.

Nikoleta [00:12:02] But the plots are still floating.

Katie [00:12:02] Yes.

Nikoleta [00:12:02] And now they start from one right before they start from zero, and if you put those two plots next to each other, it would have been so confusing.

Katie [00:12:16] So this one did- this one particularly confused a lot of people in the survey, I've done some initial analysis. And I think some people have just written, 'I don't know', instead of like a number, which is interesting.

Nikoleta [00:12:30] No it would make sense right, if you're not familiar with logarithmic, you're just like what am I looking at?

Katie [00:12:35] Exactly. And just kind of like a final thing. How much freedom do you think each language allows for sort of like customisation of features and sort of.

Nikoleta [00:12:52] A lot, they both do open source languages. Yes, R is open source language, so I think a lot of course, I guess you would need a more- if you want to use the features that are there, fine, you can do it as a user. But if you were meant to implement something, maybe you would need a bit to be more comfortable with coding. So maybe a beginner wouldn't be there. But given that you know, you have some knowledge, I think it's very easy for both languages.

Katie [00:13:27] OK.

Nikoleta [00:13:31] So, I wanna argue, I know a lot of people, a lot of people that I know that are not very good at coding, whatever that means. No, are not very good at software development, use R a lot and they're very happy with plotting. So I guess maybe in R without software development. You have more power over rewriting some of the functionality.

Katie [00:13:55] Mm hmm.

Nikoleta [00:13:56] I don't know why that is so because I again, I don't know that much about R. In Python if I want to change something, I feel like I am comfortable to do it, so I don't think is impossible, which is good, where if we're talking about other languages like matplotlib for like Matlab. That are not open source an issue. So both of them, I feel like I have a complete super power over it. Maybe in my head right now would say maybe R is a bit easier, maybe better given to the people I've spoken to.

Katie [00:14:31] Yeah. And have you got any sort of like final comments that haven't come up?

Nikoleta [00:14:45] Yes, sure, I guess let me just double check something. Yes, I think it's funny for some plots how so, bar plot 2, right, the Python code is way smaller than the R code, but then for bar plot, ay like, three, the opposite happens. I think that's I think that's I think that's funny. Like, some things take longer than one language and then they take less in others so, yes.

Katie [00:15:21] OK.

Nikoleta [00:15:22] I think that's funny little thing as in, alright, so it's do this in R, but then it'll take a few more lines in Python. But then I can argue that the bar plot is two lines.

Katie [00:15:35] OYeah, obviously that could also be down to sort of like my um-

Nikoleta [00:15:37] The way you write your code?

Katie [00:15:40] yeah, the way I write my code, the fact I have bias in R, and I'm not the most efficient at writing code anyway. Okay yeah, that is is interesting.

Nikoleta [00:15:52] But I yeah, I like both languages and I can't comment on the R aspect too much, but I know it's very beautiful, but I don't know, then again then that's another comment. I guess R comes out with out of the bag more beautiful plots. Beautiful, more appealing to the eye like they look better. That does not necessarily mean that is a good thing. When we're talking about plots, sometimes raw brutal visualisation is the best thing, and sometimes you're trying to make something very beautiful, that's when you start losing information.

Katie [00:16:33] I guess aesthetics can make it misleading sometimes.

Nikoleta [00:16:35] Yes. The floating right, like if you again, you put the plot in the one started from zero, the one started from one because they were both elevated. You would have thought that this started from the same point. So, yes, definitely R a bit more pleasant to the eye out of the bag, right, but if you put care in to do something beautiful matplotlib, I'm sure you can. But again, sometimes being appealing is not the best thing when you are talking about data visualisation.

Katie [00:17:03] OK, yeah, that's a really good point. And I'm not going to be able to unsee sort of like floating ggplot bars.

Nikoleta [00:17:13] How did you not? Like the first thing I was like, what? They're elevated.

Katie [00:17:16] I don't know, I mean on my placement- I'd never done any visualisation before and I had to learn ggplot. OK, this is how visualisations, this is what they look like.

Nikoleta [00:17:25] Everywhere.

Katie [00:17:32] I think that is everything.