Vince Transcript

Vince [00:00:02] I'm trying to make sure not to not to forget to do that.

Katie [00:00:10] So I, so I need to try and figure out how I'm going to like introduce it and everything as well. Just like I'm going to screenshare a couple of codes with you.

Vince [00:00:23] might be like just saying, because a lot of people, a lot of people, the people involved, have a loose idea. So I might be worth saying like, hi, as you know, my name's Katie. I'm doing my final year you know, my MMATH project on visualisation. Thank you so much for your time. I don't think this should take more than, I said 15 minutes in email. I almost think we should reverse engineer it so it doesn't take more than 15 minutes. But if it does, it can. And then you can say I'm I haven't heard the this is recording thing. Oh no there it is OK, I'm going to record this. But that's just so that I have records that I can transcribe into my project. If there's anything you'd like me to redact for my project, let me know. And I'm just interested in your thoughts. I'm not interested in anything else. Just your thoughts. You know, just to clarify that academics can feel because one of the reasons we became one of the one of the explanations for which we became academics is because we're very good at tests. So any conversation with academic, they can sometimes feel like they're tested. So it always trying like say like. Yeah, this is subjective. I'm just interested in your thoughts.

Katie [00:01:43] So like in the actual survey, I repeated the phrase, this is purely subjective a lot.

Vince [00:01:49] Exactly, exactly. So just that people don't feel that they're being tested or whatever.

Katie [00:01:54] I realised after I said I'm going to screenshare, I'm actually thinking I might. Make a couple of different versions of the code, but with the data. Like reading in and stuff, and I'll just explain what each day.

Vince [00:02:16] I, I am also not sure if screensharing is the best way to do it. I think when you arrange the time and when you like, send the schedule and Zoom link and stuff, I think you should send I would send them as images so that they're not opening up. Don't send the code files because then they'll open up the code files and their editors might do something to the code files, write them in a different. So I would send specific I send pictures as high res as you can get them. So screenshots right. Of of the code.

Katie [00:02:43] I think I also realised that I need to run prettier and stuff on them. I haven't done that.

Vince [00:02:49] Cool. Yes, make sure you do that as well.

Katie [00:02:52] Would you be able to go through that in this meeting? Sure, sure, yeah. After I've had emails back from Andreas, Nikoletta and Henry, I'm meeting with the Nikoletta and Henry potentially tomorrow then Andreas on Thursday.

Vince [00:03:11] But, yeah, the thing is, as soon as you got that scheduled, I would send them beforehand, send them pictures only because like if Zoom falls over if the resolution is not good, etc., etc., and it just gives them time to take a look at it. They might not they might not have the pictures there. And they're ready in case they say, oh, I haven't looked

at it. I can't find your email anymore. Right. But just have. But do send it so they have the opportunity to look at it beforehand. Right.

Katie [00:03:34] That's just. Just need to to figure out how to send the pictured because the codes are quite long.

Vince [00:03:42] Right. Right. OK, that's a good point. That's a good point. Yeah. So you can do don't make it too much, whatever it is, and see one picture file. So maybe it's only a portion of it or something like that. OK. But in a way like don't don't worry about too much, because any bias is like, for example, the fact that they haven't looked at them before or whatever. Is is exactly that right? But but I would send them this picture files, because if you open them in an editor, depending on what the editor is, they might have all sorts of stuff set up that immediately changes what it looks like so or even how it's formatted. OK, so just keep that in mind. Let's pretend we're doing this for real.

Katie [00:04:29] OK, so. Yeah, I'm Katie, this is for my project, everything.

Vince [00:04:36] Yep, great, great. When I think it might be that you transcribe this interview. Sorry for interrupting the process as well, but yeah. In that it's a dummy one, but it might be worth transcribing possibly if there's anything of value in it. But you'll just be very clear. This was with my project supervisor, right. do you know what I mean? Like who is.

Katie [00:04:56] Yeah.

Vince [00:04:57] Like who is familiar with the project. Right. So, so.

Katie [00:04:58] Yeah, and who has been involved in the whole process.

Vince [00:05:01] Exactly, exactly, exactly, exactly.

Katie [00:05:04] OK, so I will share my screen with you now.

Vince [00:05:08] Great.

Katie [00:05:11] So these are the two codes that I used for the visualisation. You can see on the left is the Python code. On the right is the R code.

Vince [00:05:20] Great. So I'm a lot more familiar with Python than I am with R or a lot more fluent in Python than I am with R so that's already one bit of bias in whatever I might suggest.

Katie [00:05:31] And it's just, obviously just looking for your thoughts, it's not it's not a test.

Vince [00:05:37] Yeah, I know.

Katie [00:05:38] Of how well you know each language.

Vince [00:05:39] Yeah, of course Thank you, thank you.

Katie [00:05:42] Um, the first-

Vince [00:05:42] So would you like my. Oh, Go ahead. Go ahead.

Katie [00:05:44] I was going to say first do you have any initial comments on the readability of the code, which obviously I'm going to run prettier and stuff on it, so.

Vince [00:05:51] Yes, yeah yeah, so it might change a little bit. So my my initial. Thing, looking at the Python one is I like that you've put your plotting code within - um in a functional way so within functions. That's good. I think maybe a docstring is my initial thing. Those are missing docstrings, you know, to kind of describe a little bit more what they're what they're doing. But other than that, I guess. But yeah, apart from that, that that all looks like standard matplotlib to me. So. So, so, yeah. Of course now taking a look at the R. And I suppose in a way, I have the the same comment that I like, that it's it's functional and maybe, they don't call the strings in R, but a leading overall bit of documentation at the start would be nice. But I understand as well that that could take up quite a fair bit of the screen, you know. I do love the assignment operator in R ,not that that's relevant to this discussion. The assignment operator, so, so much better than anything else. Apart from that. No, not no immediate comment. Okay I see on line 157 of the R there you've got, or 158 of the R there, you've got that loop going right. ~~You might, you can maybe hear Caitlin crying in the background. I don't know how how how. Well, my microphone is isolating the noise because.~~

Katie [00:07:31] ~~Ah no, I can't hear it.~~

Vince [00:07:33] ~~Cool. Hiya J! Hiya puppy! Oh do you want to come up? Sorry, this is interrupting our chat.~~

Katie [00:07:41] ~~That's okay.~~

Vince [00:07:41] ~~Yeah, that's Katie. Do you wanna say hi? Good job. Do you think she wants to see Caitlin? Yeah. Come here little girl, come here little girl. There you go. She's not very happy right now.~~

Katie [00:07:52] ~~She's so tiny.~~

Vince [00:08:04] ~~What are you doing Julian? Oh you don't want to change Caitlin's nappy? A lot of pens. Do you want a pen? Yeah, you want to take that pen to your Mom? OK, thank you Sweetheart. Yeah you can open it with your mom. OK, good boy, Riggs. Sorry, puppy. sorry, good boy. Sorry about that as well.~~

Katie [00:08:29] ~~It's a welcome- welcome interruption.~~

Vince [00:08:31] ~~The usual. The usual. Yes. My dog has slashed his paw so we've got a big cone as well so just then I had to like, help him get out. So I-.~~ Yeah, like my immediate thing like then like before when I'm like comparing them. Forgive me if I'm going kind of away from your question is I see a for loop on that, in the R code, right on line 158, but I don't see that in the Python code. Is that just because it's not there yet or ah, cool.

Katie [00:09:06] So I'm going to spend today documenting it all up. But this is just because um, so basically in the R code, you if you try and I think truncate it. The bars just don't show up because you have to have zero axis for bar plots in R.

Vince [00:09:33] So where is the equivalent of that in Python? Okay so it's the truncation there. OK, interesting.

Katie [00:09:35] Yeah, so for truncated plots in R [meant Python] you don't have to start at zero you can start anywhere, but in R-

Vince [00:09:41] Ahh yes, you you actually have to truncate the data itself.

Katie [00:09:45] Yeah.

Vince [00:09:47] Yeah, so I guess from a readability point of view, I think apart from that one line that I've looked at, I would suggest those are the only two things, whereas the thing on the the the R code, I think that truncability, truncation? That truncation is perhaps not as readable. I certainly would not have gotten that that was truncation just from a quick glance of the code. That would be my main, main comment.

Katie [00:10:12] ~~Again, I'm going to spend today fully, sort of like, documenting both of these.~~

Vince [00:10:17] ~~Sure, I think the thing it's kind of like stepping out of the interview for just half a second. Because there I was just giving you my thoughts like. It is also not a test of you, right? You know what I mean? Like like, you know, and and. For all the for all the interviewees know, this is not the code you used, this is just some code that you want to their thoughts on, does that make sense, right? So so you're saying you're saying that you're going to document it, you're going to change it. By all means, do that because you probably should anyway. From the point of view of the interview here, it's really about getting a thoughts on these two bits of code, so it doesn't matter really if thoughts are quite negative. OK?~~

Katie [00:11:05] ~~OK.~~

Vince [00:11:07] ~~Right. Because also they might not be right. Yeah. Does that make sense so don't- in the same way that a test of them, it's also not a test of you. Right. So don't don't feel the need to be, I'm saying defensive, not that you were defensive in a bad way, right? But you're not defending yourself. You're just asking for their thoughts.~~

Katie [00:11:25] ~~OK. Afterwards, I'll sort of discuss which codes to take pictures of things, um.~~

Vince [00:11:33] ~~Yeah.~~

Katie [00:11:35] Again you've kind of been through this, um, but how would you feel each code could be changed in your opinion, if there's anything you feel like you would change?

Vince [00:11:45] I think for me, the main one would just be for the Python at least, and for the R, you know, adding some leading documentation in the functionality. But my GGG, my GGG, my ggplot is certainly rusty, so I can't suggest much. But apart from that, the matplotlib looks like essentially the way to do it. I don't normally use those colour codes myself, whatever they are called, the hexes, are they hexes, are they hex code? I never use those I normally use like the RGB thing. So like. In a very subjective way, that kind of stands out to me a bit. But I am beginning to wonder if actually it's nicer, more precise to

use those standard codes as opposed to like the RGB, you know, vectors. But yeah, no, I think I think that's that's all that's those are my main thoughts on the readability.

Katie [00:12:44] All right. And so sort of based on your own knowledge and sort of after looking at these codes, how would you feel each one is kind of suited to visualisation?

Vince [00:13:00] Yeah, that's an interesting question. I mean. So. Just from the fact essentially the reputation that, ggplot, gg- why can I not count to two? ggplot holds, that's the immediate thing that stands out to me. I'm like-

Katie [00:13:18] OK.

Vince [00:13:18] -you do visualisation with ggplot, you must really, you know, you must really know what you're doing. Because ggplot is kind of like the gold standard of of of visualisation. So it certainly seems appropriate, you know, that that that weird issue with the truncation. I would or I would wonder if that does mean maybe there's something else.

Katie [00:13:40] Yeah, that's interesting.

Vince [00:13:41] Because truncating some data is not, you know, that stands out to me. I'm like, oh, I don't know if there's a better way of doing it, but that would be a thing that I'd be like. I wonder if there's a better way of doing that. And then kind of in a boring way, in a boring way on the Python side, yeah, that looks perfectly appropriate. That looks like good old boring matplotlib, you know, but boring in a good way, you know what I mean? So I don't really have much to say about that at all.

Katie [00:14:13] OK.

Vince [00:14:15] One thing, you know, and this might almost be kind of like a slight bit of imposter syndrome because my matplotlib is so much better than my my ggplot plot, you know, matplotlib's the first time I go, I pick up to draw anything anything. Like there's a bit of me that's like, because it's ggplot it's probably better, you know, but but I don't know. I don't know. But that's all that's really just the reputation of ggplot. So you know, in Python. A lot of library, a lot of other visualisations, libraries come up and they kind of promised to be the ggplot of Python. Whereas in reality, matplotlib is perfectly fine. So those are my thoughts between the two of them. Sorry if that was a bit vague again.

Katie [00:15:10] No that's good. And then again, comparing the two visualisation libraries, which do you feel will be easier for a beginner to pick up if they have an equal amount of R and Python experience and do you have any reasons for that?

Vince [00:15:26] Yeah, yeah. So it's hard for me to answer that question from the code you're showing me. OK, well, let me try to answer that question from the code you showed me. But I think what I'm trying to say is I'm not sure I'm answering the code from the code, from the question and the code you're showing me. I think I'm answering it from my own knowledge. OK, so looking at the code you're showing me, you're seeing, we're seeing these additions. We're seeing these these specific grammatical calls, the ggplot library of scale, like continuous scale, y discrete, scale_x_discrete, et cetera, et cetera. Whereas the code on the left, the Python code, seems to be: create a bar plot, create an x label, create a y label and possibly modify it, modify the text. So I think I think the matplotlib plot code you're using is probably the more straightforward one for a beginner to pick up if they just need to draw something like that.

Katie [00:16:24] I can definitely see that, yeah.

Vince [00:16:25] But you are using there, you are using there the pyplot interface to matplotlib, which is kind of-there's actually, there's two ways of plotting in matplotlib, using matplotlib directly or using pyplot and pyplot is actually just meant to be matplotlib's interface to make it like Matlab's plotter. So it is meant to be relatively straightforward. So I think in a way whereas to plot with ggplot you essentially have to learn the grammar of graphics, right? That's kind of what ggplot is all based on, is the grammar of graphics. So I think my answer is twofold. If you just need to get a bar plot or a histogram or something like that, I think that matplotlib's pyplot is probably what I'd recommend to someone with equal knowledge of both. If you want to become very good at visualisation, then I think my advice is to learn, but not because it's easiest, but because by learning it, you'll not only learn syntax for visualisation, but you'll learn the graphics of visualisation as well. So I think, I think that's my answer.

[00:17:43] So kind of Python be more user friendly, but ggplot is kind of a good one to learn.

Vince [00:17:48] Python. Yeah, yeah. Python being more- the problem with matplotlib, right, is because it's got those two interfaces to matplotlib, which by the way, Python prides itself on not having. Python prides itself as being a language where there's only one way to do things. And it's unapologetically like, no, no, no, there's no confusion. There's one way to do something, OK? And that is indeed wonderful about the language. But matplotlib is a wonderful example where that doesn't apply because it's this pyplot and there's matplotlib. And and in a way, having those two options is both nice because you've got that like gateway drug, which is pipeline at the same time as someone who really, really only ever uses pyplot. The few times now where I have to use matplotlib, it's almost harder. Whereas if you just kind of say learn ggplot. You know, then you- Whilst it's more user friendly, I'm not sure it's more user helpful, if that makes sense, I think in the grand scheme of things. Yes, learning equips you equips you better than anything else.

Katie [00:18:48] Yes, that makes a lot of sense. I don't know for this one I'd probably have to show pictures of the plots, which language you feel provides a more publication ready output.

Vince [00:19:04] Yeah. Yes, I would do that. I would just I wouldn't send these pictures beforehand, though. I would just put them up on the screen. Here are some examples of visualisations created. Which one do you feel is more publication ready? Do you have any there?

Katie [00:19:23] Yeah, I'll just do the basic that control scaling for each one.

Vince [00:19:28] Yeah

Katie [00:19:42] So these plots were made kind of as close to the-

Vince [00:19:46] You're only sharing- hold on you're own sharing VScode with me, so I can't I can't see your other windows, I think.

Katie [00:19:55] Should stop sharing and then re-share?

Vince [00:19:57] If you click. Yes, stop share. Or I think if you just click the share button again, it brings you up the option.

Vince [00:20:07] Right.

Katie [00:20:09] So these were both made, sort of as close to- well as the default settings as possible with each of the languages. With minimal sort of editing.

Vince [00:20:24] Yeah, yeah, I, I, I mean, I recognise which ones which, but I can't say. I suppose the one on the right is slightly nicer that it doesn't have the box around it. OK, so out of the box ggplot is kind of slightly more publication ready, but I don't feel there's a big difference between the two there.

Katie [00:20:53] OK.

[00:20:57] I wonder if it might be nice, Katie, if you have a plot that's the same, but that is actually markedly different or maybe even a plot where you're getting different results from the data, whether where there was a difference between the Python and R questionnaire data. You could show one of the biggest difference there is because it might not be much statistically. I want to show you this one as well as this one.

Katie [00:21:23] OK, I was gonna say I have found that the biggest difference so far I found was in the log scales, because in Python, when you use- convert the y axis to log, that uses sort of 10 to the notation which I've left in because that's the default, what Python defaults to and a lot of the response and very off or people, a couple of people even just wrote, I don't know.

Vince [00:21:47] Perfect. That's really interesting. That's fantastic. Great. Show those, show those. Because even if you get both ones, the same response, which I expect is I don't really think there's much difference there. Right. I think that'll be an interesting thing. Go look. However, the data says this, whereas my discussion with what we're going to call them subject matter experts, I don't know. No one no one noticed such a difference. Right.

Katie [00:22:13] So you can see in the python one. This is what it defaults to when you change this to a y sc- to a log scale. Which confused a lot of people.

Vince [00:22:24] That's interesting, that's interesting that the one on the left confused a lot of people and not one on the right because I find one on the right more confusing there.

Katie [00:22:31] I mean that one did also sort of um-

Vince [00:22:32] But it would be interesting then to cross reference that.

Katie [00:22:35] Pardon?

Vince [00:22:39] If you see the cross-reference that with background, right?

Katie [00:22:42] Yeah, definitely.

Vince [00:22:43] Because some people literally ten to a number could mean nothing like they might be like 'what does that mean again?' You know what I mean. Which obviously sounds weird to us.

Katie [00:22:50] But because, obviously because, in the dissertation survey exchange thing that I went into there are a lot of sort of like sociology, psychology students there doing sort of tests who would potentially have different- reactions to that then sort of mathematicians and engineers that I sent the survey to.

Vince [00:23:09] Yeah absolutely.

Katie [00:23:14] Yeah, so in general, the two of them are kind of similarly publication ready?

Vince [00:23:20] Yeah, I certainly wouldn't- but I wonder if I'm asking that if I'm answering that question because I know how easy it is to make any potential minor modifications with each language that any journal, cause journals might be like, oh, we want a double boundary against every other graph, every graph or something like that. Right. So but I think we're looking at those they're essentially the same to me.

Katie [00:23:45] Yeah. I was also the thing is, I know we'd originally taken this question out, but kind of on that note, kind of a follow on could be: sort of how easily do you feel like you can modify various features of the plot? Or like, I don't know.

Vince [00:24:05] I think yeah, I'm not sure. Yeah, I feel that I could modify both of them relatively straightforwardly because I'm familiar with both, so. Yeah. And I mean, in both senses, either using the ggplot grammar where you just don't include something or change the option of something include or or, that's quite similar to the R one so I don't think I see much difference there.

Katie [00:24:29] Okay, um, do you have any other comments that haven't come up.

Vince [00:24:39] Nothing else, nothing else.

Katie [00:24:40] OK.

Vince [00:24:42] Nope, nope, no other comments from me. Yep.