Interviewer: Recording again. So first of all, I just wanted to kind of cast our minds back since we're at the end of June now. Back in March, when you took the workshop, and you were in the Python one? Is that right?

Researcher: Yeah.

Interviewer: And were you able to attend both days?

Researcher: Yeah.

Interviewer: Okay. And what was your kind of biggest takeaway from the workshop?

Researcher: Generally, I think it was how the language works. Because I'm really, I do have a little bit of experience with R before, but I was completely new to Python pretty much. And just the way that the language works, I wouldn't say that I got much that I could directly apply to what I want to do, but definitely the kind of language is a good thing to know, to start off with. It was an intro course so I wasn't exactly expecting to come away and be like, Oh, I can do everything I want to do now. So yeah, it was good. And I didn't really know about the GitHub either. That was good.

Interviewer: Yeah. And what did you think of that course, kind of in general?

Researcher: And yeah, I think it was, good. I'm trying to remember, so the first day was Unix and I think an intro to programming. And then the second day was a bit more of the, kind of, coding, and think it was GitHub at the end. And I do think that was good and they took on board the comments. I really liked the exercises, so when they made you do it yourself. Then the second day, they asked you what you wanted more on the second day of and everybody said, "More exercises," and they took that on board. That was very good, because that definitely kind of ingrains it into your mind, and makes you think about what you're doing instead of just riding along with what they're typing on the board.

Interviewer: Yeah. And was there anything that you think would have made it maybe a better learning experience for you?

Researcher: To be honest, no, I think it was a very good introductory course. You couldn't have asked for much more really.

Interviewer: Great, glad to hear it. So now I just wanted to take a look at the workflow that we drew, that you drew. To remind ourselves, yours is actually quite easy to read some of them are a little intense. But so just to kind of go over I think what I got from this was that you do a lot of research with kind of [substance] addiction. You're looking at mice, so you have kind of mice models where you dissect and extract tissue. Then you do some analysis of the genes. And this one seems like it was mostly done off-site.

Researcher: Yeah!

Interviewer: You get the data in Excel and then you were using R for your statistics, looking at different kinds of the data and we're using Prism for your...

Commented [1]: Codes (387-980)
Takeaway
Git/GitHub
ProgrammingLiteracy
Python

Commented [2]: Codes (1355-1760)
WorkshopStrength

Researcher: That was for the stats.

Interviewer: Stats, okay. And graphing, I think I wrote down that you liked it for pretty figures. And then, mostly writing articles and Word and doing PowerPoint. So when we talked last week, we circled this one in red, doing the analysis in R and manipulating data alternatively. So that was maybe something you were thinking about.

Researcher: And I do think I will change it.

Interviewer: Yeah.

Researcher: Because R has a lot of packages and it kind of confines you into what they want you to do, and I don't do that really advanced statistical analysis on this experiment anyway. So I think writing code in Python is actually going to be much simpler. And I've looked on... It wasn't on GitHub. It was on another kind of code storing website and there's some codes that do what I want to do. So that's quite nice.

Researcher: There's lots of resources. Obviously I have to tweak them a little bit to my data, but no, it seems much simpler to do it in Python. Because it's not really that advanced I just want to compare it for groups. So yeah, I think I might do it in Python.

Interviewer: Great! So we're getting to some things you're looking at doing in the future, or a little further down the road. Is there anything that you've changed kind of at this point? Anything that's different than this?

Researcher: Well, currently at the moment, we're on the mouse training, because it takes a long time to get the mice addicted to [substance]. So we're training the mice.

Researcher: And the way that I'm going to extract the tissue has changed but generally here, as I said, that's off-site. Well, not in our lab. It's at UCSF I think, but, not in our lab, so we don't have much to do with that. But other than this, I would say, I would do the stats with the Python now.

Interviewer: Yeah, so we can go ahead and write that in. So we're doing everything that's kind of like changes you're actually in the process of making in pink, so if you want to write that down.

Researcher: And so I would still use this to make the figures.

Interviewer: Still use Prism to?

Researcher: Actually, I don't know because you can make nice ones in Python. I'll have to see when I get the data, but I'm definitely open to using Python.

Researcher: So what would you call it? Make figures in Python. Because if I'm doing sets in there, I wouldn't really have too much need to in Prism, unless we need to make them all exactly the same for the paper or whatever and then I may have to go back because not all of my figures will be from this experiment.

Commented [3]: Codes (3156-3828)
FutureChange
Python
Statistics
BetterMethods
LanguageTransition

Commented [4]: Codes (4468-4523)
FutureChange
Python
Statistics
LanguageTransition

Commented [5]: Codes (4719-5476)
FutureChange
Python
DataVisualization

Interviewer: So you might have earlier figures that you did in Prism and want to make them look the same.

Researcher: That would be the reason, but I'm quite open to doing it in Python now.

Interviewer: Yeah. Great. And remind me had you used Python before the workshop at all?

Researcher: No, I don't think I did one tiny intro course years ago but other than that no.

Interviewer: Nice. So that was really something you kind of got out of it.

Researcher: Yep.

Interviewer: Very cool. What about any of the other things that we covered in the workshop? So we talked about Git and GitHub? Was that something that you think will-

Researcher: I'll have to, yeah. Once I get started, yeah, I'm going to save my stuff. So I would definitely use GitHub and Unix, I would have to learn also, because I use a Mac, so it's just built into the mac

Interviewer: Nice. You'll have to use Unix to kind of run some of those Python commands or...?

Researcher: I think when you download... I'm not a good computer person. When you download the Jupyter, it runs it on the Unix by itself. So I think I don't actually have to do that by myself, the Jupyter does it for me. I hope so. That's the impression I learnt.

Interviewer: Yeah. So with the GitHub piece, was it using it to kind of save your own code, do you think? Or collaborate? Or where do you see that kind of fitting in?

Researcher: I think they'll be my own code.

Interviewer: Okay.

Researcher: And, perhaps in the future collaborate and use it for when I leave the lab. Because if we do this experiments. I think this is the second time, this experiment is run but it's definitely kind of now a feature and so I think it could be, but we don't really collaborate with another lab on this experiment.

Interviewer: Okay so, mostly for your own stuff.

Researcher: Yeah.

Interviewer: And then we should write that down too. GitHub...

Interviewer: So, are there any other areas where you have made changes or you're planning on making changes?

Commented [6]: Codes (5748-6125)
FutureChange
Git/GitHub
Unix/CommandLine

Commented [7]: Codes (6484-7081)
Git/GitHub
FutureChange
Collaboration
VersionControl

Researcher: No, I definitely still will write my papers in Word and use PowerPoint. And, here... Alright, that doesn't really have anything... That's more science again. So that will be the same and but yeah, as I said, we're only changing the extraction but not really anything to do with the data.

Interviewer: Right. And what kind of impact do you think this will have on your research to do things with Python instead of R and-

Researcher: I think it'll make it a lot simpler, because the guy who did it before in the lab, he said that it was a massive pain to do it before, because we do it ourselves. I think you can get these guys to help you. It's better to know how to use it yourself, then you can pass on the information down the line. It's going to make a lot simpler for us instead of scrolling through lots of Excel, where it looks... You can just pick out what you want to know.

Interviewer: So the kind of scrolling has been the... Through Excel workbooks has been the practice up to now?

Researcher: Yeah because you get like 400,000 data points or something. So, that's a lot of Excel.

Interviewer: Yeah, no kidding. So we can get it in a different format or you'll just be reading it into the Jupyter notebook?

Researcher: We get it in ... What's it called? Some kind of like FASTA format I think, and then you can upload it. And that's pretty simple, I think I did look it up, to just upload it into the Jupyter and you can put it into columns and analyze it that way. So yeah, I think that is okay to do.

Interviewer: Nice. I'm trying to look at other pain point. So yeah, I see you mentioned that one of the reasons you didn't really like R was that you didn't like the user interface. Is it that you like the Jupyter notebook kind of interface better or you just think that Python is a lot better?

Researcher: Yeah, the Jupyter notebook was quite good because you could go back and change everything and it would save what you did, which is definitely good for learning. I've still not put it into practice yet. So we'll have to see how it goes, but it makes sense that it would be a lot simpler to know what you've done wrong, because it saves everything as you do it. And that you can change things if it's wrong, you know what I mean? I didn't really like R for doing that but-

Interviewer: Yeah, so when you do these, your stats and things in Python, do you think you'll be doing it in a Jupyter notebook or just running Python in the command line?

Researcher: Definitely Jupyter note book I think. For me learning, I think it's gonna be a lot simpler. Perhaps when you get, when I know what I'm doing. It will be easier just to do it in the Python.

Interviewer: Yeah, when you get to hacker status.

Researcher: One day.

Commented [8]: Codes (7571-8165)
Impact
BetterMethods
LanguageTransition

Commented [9]: Codes (8797-9948)
JupyterNotebook
FutureChange
Python
BetterMethods
LanguageTransition

Interviewer: So one thing that we're trying to figure out is, we teach these beginning workshops, and then two days we're just cramming knowledge in your head, and then off you go. And we're trying to think of other ways that the library or how we can keep people motivated after they leave. So is there anything that helped you kind of stay motivated or, think about how you said you looked up things later to kind of keep doing that after the fact.

Researcher: I would recommend maybe giving them a sample data set. So people can just practice on whatever they want to do.

Researcher: So I found some old data from my lab that hadn't really been analyzed in this way before, and I just kind of played around with it. I need to go back and do some more of that before I use my own data. That was quite helpful into making sure you're still kind of ticking over what you learned in those two days. Because if you don't practice straight away, it's just going to go straight out your head.

Interviewer: So having maybe a more relevant data set that's kind of just either fake data or just public data or something.

Researcher: And maybe, they did do an online notebook during the workshop. I think one of the instructors was always kind of updating it as the other went along. It might be helpful just to have the main points. Just do a printout. Just so you can refer to it on your own.

Interviewer: Yeah.

Researcher: Because people are all meant to update the notebook thing, but I'm not sure everybody did. It might just be useful to have the main points and then you could scribble on that when a question gets asked. They added some extra information or something. I always find it useful to have a kind of reference.

Interviewer: Yeah. No, that's a good idea, to have the main commands or something on a handout.

Interviewer: Anything else? Things that as you... Because what we often see is this is the kind of stuff that people are like, "Oh, this is so important," and then they go back to the lab and then it's like, "Oh no, there's all this other stuff."

Interviewer: Is there anything that can help this stay relevant?

Researcher: The only other thing I think would be some follow up courses? Because I know you do a RNA-Seq for R, so maybe you could do one for Python. And I don't know what other people do before, you might know from the interviews. But just kind of... you'd want to make sure that you'd get enough kind of interest but it might be useful, because I know lots of people do that RNA-Seq so it might be useful to do it for Python as well. Because it wasn't very lab-oriented, the introduction course, so you can even have more lab, like a lab introduction to Python as well.

Commented [10]: Codes (10470-11133)
WorkshopSuggestion
Enabler

Commented [11]: Codes (11133-11741)
WorkshopSuggestion

Commented [12]: Codes (12149-12722)
WorkshopSuggestion

Interviewer: Yeah, that's cool. I like that idea. And what about kind of the opposite side of that? Was there anything that made it harder for you to kind of follow up with the materials or really kind of put it into practice?

Researcher: Yeah, only the notebook thing that said I like to have it opened up just to reference it because you were typing so much, you didn't really have time to take your own notes and the online notebook was hard to find after you left the class because they took the link away so that would help. That kind of hindered me in remembering everything. I did manage to take a few notes. But yeah, it would just be useful to have that.

Researcher: But no, because I was looking to find a data set, so I just worked away from that.

Interviewer: Yeah. And you have support from kind of your team to spend time on this kind of stuff?

Researcher: Yeah, my supervisor wants me to do it, because it's for the benefit of the lab, but nobody else really does it in the lab. It's just me kind of, but I do know some other labs that are helping me. They helped me find the code. They showed me where to look and stuff, so I found some useful information from them.

Interviewer: Cool. And sorry, the code you're talking about that was something you found on GitHub?

Researcher: Yeah, it wasn't GitHub. It was another one. I can't remember the name of it. But yeah, it was just another code storing website. And yeah, they said this could be useful lots of lab people upload their code on to there. I can find it for you

Researcher: Cool website.

Researcher: They wrote a paper on it and then they stored it on this other thing, so I found the paper and had a look on it.

Interviewer: Was that like... No.

Researcher: You can't remember either.

Interviewer: All I could think of was Zotero and that's not it.

Researcher: No, it wasn't that.

Interviewer: Zenodo?

Researcher: I think it began with an 'r'.

Interviewer: And yeah, that sounds great. Good to hear that there's... I like the idea of a handout. Because, yeah, I definitely know that when you're following along you have no time to take notes.

Researcher: And then you get lost.

Commented [13]: Codes (12949-13386)
WorkshopSuggestion

Commented [14]: Codes (13481-13905)
Enabler
Barrier

Interviewer: The last thing is just coming back to this checklist. I'm just going to summarize and think about what you're up to. So here we're talking about the kind of workflow as it is planned or as you're doing it right now. So in this workflow, do you use any programming languages like R, Python, or the command line?

Researcher: Yes.

Interviewer: Have you transformed any step by step workflows into scripts or functions?

Researcher: Already? Not yet.

Interviewer: Okay. Is it something that you think you will do actually?

Researcher: Maybe.

Interviewer: Where would that come in to play?

Researcher: So transforming...?

Interviewer: Step-by-step workflows into scripts or functions.

Researcher: Maybe not actually.

Interviewer: If you have a data cleaning and then a data analysis script could you put them all together?

Researcher: We do that kind of before... is it...? Where is it?

Researcher: I think these guys do some of that.

Interviewer: Oh yeah, the genomics center does.

Researcher: Yeah, there's specific techniques for the RNA-Seq that people read in the data, but I don't think I would do it actually, sorry.

Interviewer: Do use any version control to manage your code?

Researcher: No, not yet.

Interviewer: Do you use any open source software?

Researcher: I do use now, Github.

Interviewer: Do you share any of your code publicly?

Researcher: No, not yet.

Interviewer: And do share your computational workflow or protocols publicly?

Commented [15]: Codes (15217-16007)
QuestionConfusion

Commented [16]: Codes (16093-16177)
Change
Git/GitHub
QuestionConfusion

Researcher: No, but we will do.

Interviewer: Alright. And then the last thing is just kind of a bigger question, which is that. Here at the library, we teach these beginning programming courses and obviously want you to learn beginning programming, but one of the bigger goals is to kind of help people make their research more computationally reproducible. So the idea that somebody else could have the same data, the same code and regenerate the same results. So we just wanted to get kind of your thoughts, if you thought that attending this course had kind of helped you make your work more reproducible.

Researcher: Yeah, I guess so because I now know a platform to save all of my code so that would be much more reliable than me trying to remember it all because I didn't know about GitHub before, so yeah.

Interviewer: So you think that's kind of one of the main ways that it's become more reproducible, is the code saving?

Researcher: I think so. Yeah. I've not really done it yet, so it's hard to say, but yeah, that would definitely help so far. And I guess when you just pick out what you want to know, using Python rather than doing it in Excel which is very kind of time consuming and there's obviously people error in that as well. So it'll probably make it more reproducible there as well.

Interviewer: Great. Good to hear.

Interviewer: Alright, so that is pretty much all we have. I'm just going to turn this off.

Commented [17]: Codes (16364-17637)
Impact
Reproducibility
VersionControl
Git/GitHub
Python