

Note: This interview was done via video conferencing rather than in-person

Interviewer: Awesome. Well, thanks again for calling in. It's really good to touch base. So basically, today I just wanted to cast our minds back and think of the programming workshop, talk about that a little bit, and then talk about the research workflow that we drew out back in, probably, February at this point. So it's been a while. Just to see what you're up to. And that's pretty much it. I don't think it will really take any longer than maybe 15, 20 minutes, so-

Researcher: Okay.

Interviewer: ... but we have 30 just in case. Yeah. Just to start out, just wanted to think a little bit about the workshop itself. Were you able to attend both days at the workshop?

Researcher: I was, yes.

Interviewer: Okay. What did you think of the workshop?

Researcher: That was a little while back. The instructors were really great. Friendly, approachable. It was a lot more of the Unix and Git than I thought it would be coming into it, since I was there mostly for R. The Unix was actually the hard part for me, I feel, or the Git part in getting it on to my computer. But for the portions of R that we did cover... I think his name was [name] that covered it or something, but he did a really good job walking through the intro.

Commented [1]: Codes (840-1304)
WorkshopSuggestion

Commented [2]: Codes (1304-1621)
WorkshopStrength

Researcher: I felt it was very introductory level, not too many intense functions or intense statistics going on, I guess. Even though it was still challenging for me as a beginner in ways, with the challenge questions. But yeah, for \$25 and two days of class, getting everything that we did, I loved it so much, so-

Interviewer: Good. Glad to hear that. What do you think your biggest takeaway was?

Commented [3]: Codes (1704-2024)
Takeaway
FurtherTraining
LearningCurve
ProgrammingLiteracy

Researcher: My biggest takeaway would be that it takes a lot of time to really get full understanding of the programming language and practice, as well. I can't read cookbooks and understand the theory behind it. I'd have to really go deeper, I think, to use programming to leverage the research I want to do, I think.

Researcher: Talking to the teachers, it took them several years to get to where they were, which was unbelievable since it just seems like, "Oh, they probably were always gifted." But they took a lot of time and started on intro level, too, so that was encouraging.

Commented [4]: Codes (2024-2289)
Takeaway
FurtherTraining
LearningCurve
ProgrammingLiteracy
WorkshopStrength

Researcher: So takeaway would definitely be needing to practice and learn the theory, which is why I ended up trying to reach out to take the biostatistics course through the Data Science Initiative, too. Just working on ways to better my overall picture.

Commented [5]: Codes (2289-2545)
Takeaway
FurtherTraining
LearningCurve
ProgrammingLiteracy

Interviewer: Yeah. So you reached out and registered for another class another the library class?

Researcher: Mm-hmm (affirmative). Yeah. I attempted to. I think it might have been either Intro to Biostatistics with R, or intermediate level or something, with what's her name? [name]...

Interviewer: [name] maybe?

Researcher: I don't know. Oh, [name] It might be [name]. I think I'm just saying names.
[name] Yeah. I tried several times, but I was always waitlisted for that one, or it wouldn't work with my schedule. emailing [name] about it, and she was so kind to reference me some resources. And I still need to...

Researcher: I'm just..

Researcher: Hi.

Interviewer: All right. I think I got you back. So you're following-

Interviewer: Hi. So you're following back up, looking at some resources. Is that what I heard?

Researcher: Yeah. While I was trying to learn coding, I was taking a statistics course with my program, and I really wanted to make sense of the two, so I'm trying to find courses that help me understand the stats and the R together. Through my [program] that had the biostats class, we learned it through Stata, and totally different from R. And when I compared using the Stata language to R, I realized R was a lot more complex. Stata's just like click and go. It's so different. And then when I tried to do some online stuff with R in statistics, it's like, you need to really create long functions and formulas it seems, from my initial knowledge. So, very different.

Interviewer: Yeah. And thinking back about the workshop a little bit more, was there anything we could have done to make it a better learning experience for you?

Researcher: I don't know if Unix and R could be separated more, but that was kind of beyond me, I think. And we spent time on GitHub, which is really cool to know, because my actual engineer friends use it all the time, and so, I can converse with them on that element. But for me, I'm very beginner and my classmate who I was there with, we were both like, "Oh. It would have been nice to have more R exposure, I think, since that's what we came for, but I know that class we had was trialing the Unix thing is what we hear. So, yeah. That was for us.

Interviewer: Yeah. Yeah. Good to hear. Okay, so now, I just wanted to take another look at the workflow that you drew. So actually, I think I might turn off my video and share my screen. Just, I don't want to overwhelm the line here. So let's see, share screen. There we go. Okay, so, whoa. So can you see that screen?

Researcher: Yes, I can. It looks great.

Interviewer: So here's your beautiful drawing. Just to kind of recap what I took away with this is that you were working on an [drug class] research project where you were pulling data from [database], so de-identified data about basically patients who were prescribed [drug class]? I can't quite remember the... Is that kind of your patient population?

Researcher: Yes. Mm-hmm (affirmative). Yeah. Sorry if I'm responding late. It's kind of cutting out here and there, so.

Commented [6]: Codes (3272-4039)
FurtherTraining
RProgramming
Statistics

Commented [7]: Codes (4201-4754)
WorkshopSuggestion

Interviewer: Oh, yeah. Sorry. And then you were doing your data analysis. You weren't quite sure yet. You were kind of leaning either towards Stata or towards R and then you were planning on writing a paper in word, and maybe doing some data visualizations in Tableau as well.

Interviewer: So at this point, I'm kind of interested to know, so it's been maybe three or four months. Have you made any changes to this workflow? Is there anything that you're doing differently now?

Researcher: So I didn't hear your overall summary to edit, but from an overview, it's not changed too much. We're still using the same data resource. When we're still using it with Excel. I don't know if I put Stata before. We're definitely just going to do R now.

Interviewer: Okay.

Researcher: My preceptor might use, or my PI might use Stata to double check the algorithm that we come up with. But for me it's...

Researcher: Did anything else change? I don't think so. All I know that's new to me is we have a biostatistician and he said we might use the MyResearch Portal. And because there might be a lot of data, we need to look into storage and not sure what's going to happen on that end. But that's about it, I believe.

Interviewer: Yeah. So sorry. I think you cut out there a little bit. So you, before, were not sure about Stata or R, and now you're pretty sure you're going to use R, is that right?

Researcher: Yes. That's correct.

Interviewer: Okay. And was that basically because of the R that you learned in the workshop, or did something else kind of make you lean more towards R?

Researcher: I definitely think having the workshop available helped my PI feel comfortable in moving forward with using R, just because I can build ... in Stata, much that I would be learning a different language from her just so that we could cross-reference, I guess.

Interviewer: Mm-hmm (affirmative).

Researcher: Yeah.

Interviewer: And maybe you mentioned this, but what about the Tableau piece of that? Is that something you're still planning on using?

Researcher: For Tableau, I guess it depends on time, but I do still like Tableau and I will definitely try to incorporate it if I can somehow in diagrams and figures. But it's mostly dependent on time, since I'm not very familiar with Tableau yet. And so if it's hard to use, then I may not get to use it for those great graphics. But I can-

Interviewer: Yeah. And so thinking those are changes you've kind of already made, are there any changes going forward that you're thinking about doing differently? Any future plans?

Commented [8]: Codes (6065-6481)
Change
RProgramming
Statistics

Commented [9]: Codes (6794-7432)
Change
RProgramming
Statistics

Researcher: For this research topic?

Interviewer: Yeah.

Researcher: I think, so, we kind of already have the data, and they're handling all the management. I guess I don't have a say in that, but from a retrospective point, I would have. So if it's just future-wise, I don't think too much will change. But your question was what I would work on? Or, sorry.

Interviewer: Yeah. Was there anything that you would do differently going forward, or that you will do differently in future workflows?

Researcher: Oh, okay. For future workflow, I guess, since it's kind of a group effort it's hard to say. Because, for me, I would have loved to have been more involved in the beginning parts with data management and ... But I guess I wouldn't change it too much. I would just want to be more involved with it. Changing the workflow. Trying to see if I could just say yes or no, because I feel like this is the standard workflow where you collect the data, you process it, and et cetera. So I don't think the workflow would change too much. Yeah.

Interviewer: Yeah. No. That's totally fine. So it sounds like maybe you took a little bit of R from the workshop, in that you're able to implement. What about, we talked a little bit about Unix and Git. But is there any place where you could see using Unix for this workflow? Or Git?

Researcher: For this workflow, I feel maybe not, just because I'm not the most familiar and my PI has no idea ... those at all. But in the future, if I were running my own projects, and I were more familiar, I would definitely try to use GitHub and all that because from what I saw, it's a good way to connect with other people on projects and I know people who use it as well. So just for project and practice, it would have been good for a personal growth thing. But for research with my PI, yeah. She's not familiar, so I don't think we would use it.

Interviewer: But maybe something to think about in the future, in terms of publicizing your work? Is that the angle you were thinking of?

Researcher: Yeah. I think so. Yeah.

Interviewer: Yeah. So thinking about, just your research practice in general, is there anything else you feel like sharing about the ways that you're working now that might be different, or outcomes from the workshop?

Researcher: I definitely feel like the workshop gave me a good first layer understanding of R and programming in general, with all the different... I'm trying to reflect on when I had the class a few months ago. But different things, the issues you come across, and how to troubleshoot with character types and things like that. And I noticed, from the workshop, it made learning another language like Stata a lot more manageable and simpler to understand, so it's definitely a good intro. Yeah, and the class just validated my feelings of, "Oh, this is useful to have and is going to be really cool for research." And it made me really curious to go to other events. I think I saw you at Berkeley-

Interviewer: Yeah.

Commented [10]: Codes (9467-10021)

FutureChange
Git/GitHub
Collaboration
Barrier

Commented [11]: Codes (10413-11112)

Takeaway
ProgrammingLiteracy
FurtherTraining

Researcher: ... for the UC Berkeley [event]. And that was really interesting, so I hope my direction goes sort of there, too. So, yeah.

Interviewer: Yeah.

Interviewer: No that's great. So maybe it even opened some new potential areas of research for you, is that what you're saying?

Researcher: Oh, no. Yeah, maybe. I can't tell if you were giving me thinking time or ... connected.

Interviewer: No. I think that's fine. So the last couple questions are, so you take these workshops and then we set you free. So one of the questions we have is, is there anything that helped you or motivated you after the workshop to keep going with programming, or anything was supportive in helping you learn?

Researcher: Oh, I see. It was definitely having my biostats course where I was learning a bit of Stata. When I was learning intermediate biostatistics, it definitely was like, "Oh, learning how to make algorithms and modeling." I couldn't wait to start trying to do that with coding a merging the two. That and seeing events out like the [event]. It came in my mail, from this description, I think.

Researcher: Little things like that that would trickle into my life. Knowing I had this project to do with R... It was definitely hard trying to find time to continue learning R and practicing. So that's something I'm still working on, and yeah. But I've been kind of slow on learning the R portion, since I'm trying to master the statistics right now...behind it.

Researcher: But, yeah. There are a lot of things going, like school-wise, UCSF event-wise, that reminds me, "Oh. I've got to keep up with this." Even here at the [institute], just today in orientation ...getting a database and scope, but I was still able to talk a little bit about it with them, and just little experiences. It's kind of everywhere. It's ubiquitous wherever I'm going, I feel, to keep learning R.

Interviewer: Yeah. That's great. So you mentioned that one of the barriers was maybe just time? Was there any other barriers you faced in implementing and using R for your research?

Researcher: Yeah. Maybe time. And just having the resources to... Maybe it's just not familiar to me, but I tried some online videos too, since they were free online courses. And they're very different. They were a lot harder to understand and follow than the workshop we had, I feel. And the codes are long and lacked explanation, so that was difficult. But, I always go on a tangent, and I forget with the question is.

Interviewer: Just any barriers to using R.

Researcher: Oh. Oh yeah, R itself is kind of hard for me. I think there may be personal barriers because I just want to know the statistics before I try to code mindlessly. But otherwise, even maybe just knowing where I can get practice data sets with some sort of key, or something.

Commented [12]: Codes (11827-13005)
Enabler

Commented [13]: Codes (13187-13608)
Barrier

Commented [14]: Codes (13651-14122)
Barrier

or I don't know. Or like a guide to walk through things on my downtime. I don't really know where to go for that, but if that were available, I could use it and practice that way, maybe?

Interviewer: Mm-hmm (affirmative).

Researcher: Yeah, I don't know if that exists, but, yeah. If that were just readily available, I'd probably use it more, and that would help.

Interviewer: Yeah. That's a great idea. So just wrapping up, I have that checklist that we filled out originally. So I was just going to run through that with you.

Researcher: Okay.

Interviewer: So this is looking at the workflow as it is now, as you have it planned. So as part of this workflow, do you use any programming languages, like R, Python, or the command line?

Researcher: I would say, just R.

Interviewer: Yep. Okay. Do you have any place where you've transformed a step-by-step workflow into a script?

Researcher: I have not done that yet, but I might have to.

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

Researcher: I do not.

Interviewer: Do you use any open source software?

Researcher: I guess so? Would that... R?

Interviewer: Yep, I think R counts. Do you plan on, or do you share you code publicly?

Researcher: So that, I think I originally said I wanted to, but after chatting with my PI, she's pretty strict about not sharing it. You wouldn't, unfortunately

Interviewer: Yeah, and why was that?

Researcher: I guess maybe it's a personality thing, too. She's just very protective of the data set, and she had us all sign contracts to not share or disclose information about people. I think she's just worried about having patient health information. Even though it was de-identified. I guess she's concerned with the risks. And I feel like she may not have had any discussions about how to proceed with data sharing in a safe way, so I think she has uncertainties. As a student, I hadn't pushed her out of her comfort zone because I am very small on the totem pole. So, yeah. It's pretty unfortunate.

Researcher: So I think if... It's like the older researchers, maybe they'd be more open, too. I don't know.

Commented [15]: Codes (15051-15087)
CodeSharing

Commented [16]: Codes (15087-15297)
CodeSharing
Barrier

Commented [17]: Codes (15297-15941)
Barrier

Interviewer: Yeah, no. I think that's a pretty common thing. Hopefully as you work with other people, you'll have more of a chance to do more of that public sharing.

Interviewer: And the last question was about another sharing question. Share your computational workflow or protocols publicly.

Researcher: I haven't had that discussion with her... to it mention of data. And we do have to describe our data collection process and everything, so it seems like that might be more on the yes side, but that's just from my limited knowledge.

Interviewer: Right. So do you think, if you had to go one way or another, yes, or no? What do you think?

Researcher: Do you mean in a code way, or just from ...?

Interviewer: Yeah. Sort of like, what was your whole pipeline of gathering the data, and cleaning the data, and analyzing it?

Researcher: Oh. Me, I would love to. But my PI, I'm afraid she would not, because I should have changed my questions from the beginning. But yeah. It's just... She wouldn't even let me be involved with the data processing process, so I wonder if she would have it written out for the public, or for publications. I think it would be limited to that part in the publication where you write your methods and how you did you inclusion, exclusion. That kind of thing.

Interviewer: Yeah. That sounds in line with the other things you were saying. So then, one last question, which is kind of just a broader question. We teach these workshops to teach you basic programming, but we're hoping that it helps you make your work more reproducible.

Researcher: Sorry you are cutting out a little bit. Could you-

Interviewer: Oh, sorry. The question was, basically, do you think that participating in the workshop has helped you make your work more reproducible?

Researcher: I think so. I guess, having someone check over the code, because we have someone else who is also going to be doing R on the team, and we're kind of cross-checking with each other in addition to Stata. So hopefully our results are reproducible. And that's the goal, so, I think if I know how to do R correctly and right, we should be reproducible. Yeah.

Interviewer: Cool. Glad to hear it. All right. That is basically all I had. Thanks for chatting with me. And just the kind of output-

Interviewer: I'm glad we made it through the conversation. I got everything I needed. So thank you for coming on and chatting, and... I think [inaudible 00:26:44].

Interviewer: Let me know when you get back to UCSF so I give you your gift card.

Interviewer: Well best of luck with your summer. Have a good one.

Researcher: Thanks for all of your workshops. They're really amazing. I love-

Commented [18]: Codes (16216-17340)
Barrier
CodeSharing

Commented [19]: Codes (17677-18193)
Reproducibility
CodeSharing

Interviewer: Oh good. Glad to hear it.

