Interviewer: ... recorder. Awesome. Basically, what we're going to be doing most of the time is you're going to be actually drawing out your research process and highlighting tools that you use along the way. We'll just be talking about it. Before we start that, can you just give me kind of a high level, what is kind of the goal of your research, how long have you been at UCSF, that kind of stuff.

Researcher: I've been here for about [time]. We started up the lab, so we were starting to establish new techniques in the lab. One of the ones that came up that we didn't know going in but we realized was important for us to learn in the very beginning was single-cell RNA sequencing analysis. With that, no one in the lab had experience with R or Python, and we realized that we needed to get some sort of programming experience in order to establish that technique and use that to move our projects forward.

Researcher: We actually had a group, we knew people at the [institute], so we're collaborating with them. Because we didn't have any expertise at the time, we kind of used their expertise to help us get started.

Researcher: Then, I would say the other stuff, it's mostly from 10x Genomics because we use their machine. I read all the stuff from their website to better understand how it works. Then, where the programming part comes in is really R for Seurat, so we use a Seurat package. In that, at first, it was just using the pipelines that are already published on the 10x website and by Seurat itself.

Researcher: We went mostly off of that, but as we realize that we want to tweak certain things, we needed to learn how to change the code to get what type of graphs we want, so that took a lot of ... I mean at first, it was really just Googling to figure out what people have said because we didn't have experience. Then, slowly we talked more and more [institute], people who are more in bioinformatics, and we started to collect more commands or collect pieces of codes, code that were really useful for us to get the graphs that we wanted and the type of analysis we wanted. Then, over time, we created a notebook that then more people, when they join the lab, would just use that notebook.

Interviewer: Cool. I'm actually just taking a even bigger step back. What is kinda the main goal of your research?

Researcher: Oh, that's right, I went straight into ... The overarching goal is to better understand the heterogeneity of the cells in a specific organ that we work on. We're taking an unbiased approach by using single-cell RNA sequencing to see how the cells are transcriptomically different, seeing what type of hypotheses we can make using that technique and then test them physically in the lab with wet lab skills.

Interviewer: Cool, sounds really cool. Awesome. Let's get to drawing. Basically, what we have here is kind of a template, but it may or may not kind of match your flow, but we kind of start with where does the data come from, who's gathering it, what

kinds of cleaning or processing does it go through, if you analyze it what packages are you using for that. Then, what comes out of it at the end, paper, one generally assumes, but is there a code that's shared, is their data that's shared, any of those pieces?

Interviewer: Just to some examples, this is an example of more kind of like a data science pipeline, but here you see like two different datasets that are merged together and then the data is cleaned. They then get all of these other things that they add to it and it's kind of like an iterative process, but they use Stata. Then, up here they write their paper using LaTex. Most people write their paper in Word.

Researcher: Not sure what that is...

Interviewer: It's like a programming language for paper writing basically. Then, they go ahead and submit the paper. Just to give you a kind of a sense of what this can look like, yours will probably look very different. That's kind of the things we're thinking about is what kinds of data is it, what tools we're using to analyze it? Mostly that. Feel free to start drawing anywhere you want, and we'll just talk about it as you go.

Researcher: I don't know where exactly all these things fall on here. I can just start with the process?

Interviewer: It's a pretty loosey-goosey, so wherever makes sense for you.

Researcher: Just any example experiment?

Interviewer: Yeah. What is kind of like the main ... Is there like a main kind of pipeline?

Researcher: Yeah, there's a main workflow just to get data, but then there's several projects going on.

Interviewer: Is there a kind of like a main project that you're helping out with or what is your role in all of that?

Researcher: Yeah, I have several projects. I'll just pick one of them.

Interviewer: Is there a project that you're more hoping to use your new, let me remind myself-

Researcher: Python.

Interviewer: ... Python skills on?

Researcher: Yeah. For one of them, it just recently got published, the one that we're collaborating with the [institute]. They're now shifting towards Python because it's much faster at doing the same things. I haven't learned it yet. It seems like

since they are really pushing for that, we're going to do the next project we do with them in Python instead of in R.

Interviewer: Interesting.

Researcher: That's why I'm thinking of attending this workshop.

Interviewer: If we're going to be comparing three months afterwards, would it be good to start with that pipeline to see pieces of it that you do and R and [institute] versus-

Researcher: Mm-hmm (affirmative). It depends though because I don't know if I will necessarily be working on it in the lab or not.

Interviewer: That's a good question. I think 'cause you're a lab manager is kind of your role?

Researcher: Yeah, but I'm also a research associate.

Interviewer: You're the first non, just postdoc. It's definitely interesting to think about the different role, but we have a lot of people in your role in the class, so I do wanna kind of see what you take out of it.

Researcher: My project, which is just my project, so this one is a collaborative one and then I have my own one that I've been doing for the last year and a half; that one involves the same process.

Interviewer: Let's talk about that one, then.

Researcher: I don't know if I would use Python necessarily for that one, but the lab is shifting towards that.

Interviewer: We could do-

Researcher: It may enter into mine.

Interviewer: ... both of them, also. Why don't we start with your personal project?

Researcher: All right. Where we get our data from, it's really cells from mice. First, we have to harvest mice. Should I draw pictures of it or just write?

Interviewer: You can just write it or if you wanna draw pictures you can draw pictures, as well. I've gotten all different kinds of things. Just write cells for mice if that's easier.

Researcher: Trying to think how to draw a mouse.

Interviewer: Are you doing that harvesting?

Researcher: Mm-hmm (affirmative). Literally, we just cut open the mouse, take out the [body part] and then specific [body part] go into specific types of ... Oh, no, sorry. This is the second part of it.

Researcher: We do two things in this. We do single-cell RNA sequencing to make these, generate these hypotheses, but then ultimately wanna test it. Then, when we harvest [body part], that's mostly for like embedding it into paraffin or into cryo to get tissues that we can section and then stain. If we see a trend that ... mRNA is what we're really getting transcriptomically through single cell. If we see a trend in mRNA, then we can check if that's transcribed at the protein level, and those changes are actually happening when we stain in for it. This is actually the validation part. This is the second part, but this is what I'm currently doing, so that came straight into my-

Interviewer: Maybe we could start it ... Where, what's the very first part.

Researcher: I guess the very first part is more a perfusion. Wait, but there's more pages, right?

Interviewer: You can start, yeah. You can use any of those if you want.

Researcher: Might just start. In the very beginning, so I'm going to just go back one and a half years. In a perfusion, we connect up the mouse to more tubing and create a flow to get rid of the blood. The organ we work with is the [body part]. We flush all the blood out and then we digest the [body part] cells so that they'll more easily be released because they're all in this very tight form in the [body part]. Let's see. How do I write this. Perfusion and digestion. All right, I guess I'll make that into a box.

Researcher: Then, from there, we make single-cell suspensions of these cells. There's so many different types of cells in the [body part] that you can do specific protocols to isolate specific cells. That itself is a long process. Sometimes that may involve sorting on flow cytometer, as well. This arrow is really long, but you finally get to your single-cell suspension that you can submit to 10x Genomics, which is what we use.

Interviewer: Sorry, that's a company or that's a tool?

Researcher: Both. That company made this little machine that is I think part of the [center] or it's part of another core here, so we literally just give them our sample, and they run everything for us. Our work stops once we get the single-cell suspension, we give it to them, then they do all actual steps from there until we get our data back.

Interviewer: Then, what does it look like when you get it, like a spreadsheet or-

Researcher: When we get it back, it looks like three types of files: a matrix file, a bar code file, and a gene file. Then, we input that into Seurat directly. All the sequencing alignment, all that, everything is done beforehand by them, so it's like a black box, we don't see it. I guess would that be in this section?

Interviewer: Yeah or the analysis kind of. Some people don't have that much in that section.

Researcher: 'Cause we don't do that part of it, so it's just like a step we skip.

Interviewer: Could go right to analysis, then.

Researcher: Then, when we get all these, we get the raw data back basically. Then, we have to clean it up. Then, you clean up the raw data. This is mostly just figuring out good cutoffs for things because there's nothing really established. We need to find ways to figure out which cells are dead, which cells are not dead, which cells are doublets maybe, or are there things that we know we should be seeing We aren't seeing? How well did this process actually go? Was it read depth okay? All sorts of questions like that.

Researcher: I guess the quality control part of it that it's not really streamlined in the process. It's just we take a look at the data. Does it make sense? We go back, change some cutoffs and then keep doing that until we're satisfied. There's no, there's nothing in the field that's established right now as far as to just use as a cutoff.

Interviewer: Got it. What do you use to do that?

Researcher: For that, in Seurat, there's specific things you can look at, like number of genes express or the percentage of mitochondria. It's all within Seurat.

Interviewer: Sorry, I think you said this already, but Seurat is an R package?

Researcher: Mm-hmm (affirmative). It's an R package specifically for single cell, analyzing single-cell data.

Interviewer: Cool. Why don't we write that.

Researcher: Should I write that out: in Seurat? Then, sometimes we use Scatter, as well, though, it's kind of new in my lab.

Interviewer: What's that?

Researcher: It's a package, similar to Seurat, but it has, it allows you to do certain things that help you clean up the data and it has its automated system of determining what's the outlier, what's not. It kind of is like a second glance at it to check if you've identified some dead cells, does it also identify those outliers or not.

Interviewer: Got it. That's also an R package?

Researcher: Yeah. I don't have too much experience with that, but some people in our lab are bringing that into it. I mostly been using Seurat. Do you want me to go into detail exactly about what we do?

Interviewer: I think that's good if we know that most of the kind of cleaning and data is done in that package.

Researcher: All right. Then, from here, once we have identified our actual cells, then sometimes because we do single cell on a lot of cell groups at once, we may be wanting to focus on one of those cell types. Then, it goes to annotation to figure out what types we have and then subset on the ones that we want. I guess from here it goes to cell annotation. For annotation, we mostly use markers that we know that are established in their field, so just genes that we know.

Researcher: If that doesn't work, we use a package called SingleR, which is another way of identifying the cell types because they look at reference gene expression databases from bulk RNA-seq data from pure cells. That database will help you identify a new cell and put it into a category, like, oh, this is a macrophage or maybe this is a dendritic cell. If you really have no clue, that is something that gives you a hint as to, oh, this may be the cell type. That's what we use SingleR for. Otherwise, it's just markers that we know.

Researcher: Then, after annotation, it really depends where the project goes from there. Usually, we subset on the cell type we want and that's another function in Seurat. Then, see, I'm trying to figure out which project to focus on because up till now, all of this is the same.

Interviewer: Do you want to focus on your project?

Researcher: Yeah. All right. I'm trying to figure out which avenue to talk about because I've taken it so many ways. Well, let's see. I'll focus on the [cell type], for now. That's one of the cell types. I've analyzed all the cell types.

Interviewer: Got it.

Researcher: All right, if I'm going to focus on that one. Then, what happened is I found out a new population of cells that we, I'm looking in a specific disease and ... We compare: we did this whole process for disease and then we did this whole process for non-disease. We found one population in this section that was not there before, so it showed up in disease. Then, we subset on that population of interest and then we tried to stain for it. Then, it goes back to the second section over there. I'm gonna start it here. I might run out of space.

Interviewer: You got another page, paper underneath there. It'll be a two paper thing.

Researcher: I guess this is still in processing, I would say. I don't know. This doesn't really fit very well.

Interviewer: That's totally fine. You can just ignore it.

Researcher: All right. Well, this is basically taking out the [body part] from the mouse and then we process it so that we end up embedding. We embed it in two different ways. It's just like two different ways of fixing the tissue and embedding it in different materials. There's two ways to do it later on. There's like different techniques that are better for one type versus the other.

Researcher: We usually end up doing both. Then, from here we ... Actually, this is combined because other things you can do in Seurat are if you find a new population, you can find new markers for that population. Should I pretend this subset-

Interviewer: The only person this diagram needs to make sense to is you 'cause you're gonna look at it again in June.

Researcher: I was like how am I gonna put this so that someone else looking at this would understand.

Interviewer: You just need to remember what it is. We're just using it to talk about your process.

Researcher: All right. That's helpful because then I can draw arrows that ... I guess I'll go from here. Find. Then, this goes back, so the markers you find here, you stain for those here. Stain with ... These are connected, so I guess I can draw an arrow like that. Then, there's two ways of doing this. Since all of this is transcriptomics, it's on the mRNA level, so we can do in situs. It's a type of staining that will stain the mRNA. Then, there's also antibody staining, which will stain the protein, which is the step that comes after mRNA. Sometimes there's a gap or there's a change in what is that one step versus the next step.

Interviewer: Got It.

Researcher: Usually, it's easier and cheaper, and most people just stain for the antibody, but then if there's issues, we go back to in situ. This has two. This is either in situ or antibody. I've been doing both of these. I've used these, then to go back to the populations I'm finding here, compare them, see do the clusters of cells that Seurat is giving me actually map onto a [body part] or not, and what is changing in wild type versus my disease. I'm at that stage right now where I'm comparing them.

Interviewer: All right. When you have that, what kind of, how are you looking at that? How are you comparing it? Is that still in Seurat or is that-

Researcher: Well, there's just a lot of different questions that we can ask between the populations that Seurat is clustering. I guess both.

Interviewer: when you're eventually done, I assume there's gonna to be a paper?

Researcher: Yeah.

Interviewer: What goes in that paper, are there figures?

Researcher: Yeah.

Interviewer: Is there-

Researcher: There's gonna be lots of figures, there's going to be a lot from Seurat and there's definitely going to be a lot more from image analysis, As well.

Interviewer: Maybe we can write some of that over here and do like a dashed line or something as like this is the stuff that is coming.

Researcher: The thing also is for this project, we have one main question, but we don't know what will lead to answering that, yet, so I don't have a straight line here.

Interviewer: Got It.

Researcher: It's gonna keep changing as time goes by and if I find one question that's, if I find some finding that's more interesting, I might pursue that, drop the main question and change the question.

Interviewer: Got it.

Researcher: It's not very straightforward.

Interviewer: Maybe we can do what you think you're going to do, right now. We can do like a dashed line or just a-

Researcher: This will eventually lead to some imaging. Imaging analysis. Then, somehow that will be quantified. That will probably turn more into bar graphs or more sort of-

Interviewer: Do you know yet what tool you'll use for that or is that still something...

Researcher: Yeah, we're mostly going to use ImageJ, but we may change that because I haven't gotten into it yet. Let me write that. Then, let's see. I feel like this is already the outputs that we would, it's already more towards this section because whatever graphs we produce from here are already ready to be in the paper.

Interviewer: Nice.

Researcher: We just don't know which ones to use yet because we don't know what the main question is that we're answering.

Interviewer: Got it.

Researcher: We have a lot of data and a lot of analysis available. Just, I have to pick and choose what's the most important and what to continue focusing on.

Interviewer: When you actually write your paper, are you gonna write it in Word or use another tool?

Researcher: I think so 'cause I don't know that there's more that exists. The other paper that ... Oh, here?

Interviewer: Yeah.

Researcher: Paper?

Interviewer: Write paper.

Researcher: I guess this will be a dotted line, too, like figures for paper. This'll be interesting 'cause in a few months I will probably be at this stage. All this that I was doing was a pilot study and now we're doing the real version of it right now in a month. I'll probably have the data back, be doing the analysis and reaching the stage-

Interviewer: Cool. That's awesome.

Researcher: ... the next time we talk.

Interviewer: Do we have most of the main stages and tools, especially? Are there any other-

Researcher: Tools?

Interviewer: ... tools you use along the way? It sounds like you're doing most of her analysis in kind of ImageJ and then the Seurat?

Researcher: Yeah. It's mostly been Seurat, for now. Also, in R because sometimes we just use ggplot to make graphs that we want. It's not necessarily Seurat.

Interviewer: That's great.

Researcher: Does that count as-

Interviewer: Yeah, totally. Let's see. Can we write that somewhere? Sometimes.

Researcher: Or just code. It's not really even a ... We recently had someone that knows R really all joined the lab, and she has been able to just help us write code whenever we want to create some sort of graph.

Interviewer: Very cool.

Researcher: That's what I mean.

Interviewer: I was gonna ask about that. Besides just these packages, is there any place where you're kind of writing code from scratch?

Researcher: Yeah, a lot of the times if there's something that didn't exist as something I could call, then I would try to create it. Now that she's joined, it's a lot easier.

Interviewer: You have done some of your own coding and all that?

Researcher: Yeah, I've tried. Oftentimes, even just in Seurat, they have something, but then I have to change specific things to make it work the way I want it to. It's not just changing one word, but it's actually changing, like I have to change the cells that are going inputting into it, then I had to create some new vector of those cells. There's a lot that go in.

Interviewer: Do you do any, do you share that code with anybody else within the lab or outside the lab?

Researcher: Yeah, whatever I end up figuring out, it'll stay with the lab. It'll go to everyone because-

Interviewer: Be passed down.

Researcher: ... all the people that have, yeah, because most of our lab doesn't have experience with programming at all, so whatever we establish as a pipeline is used for everyone's analysis.

Interviewer: Got It. Do you know if you do any kind of version control on that code, in terms of using Git or GitHub?

Researcher: No, we haven't done that yet, but it's mostly been just dated, so then we know how it's changing over time. It's very new in our lab, so we haven't reached that stage where we put it into Git.

Interviewer: Do we feel like we've gotten all the main points down here, in terms of a workflow?

Researcher: I think maybe over here I would add sorting 'cause we do this sometimes, and I may do this. I didn't do it in my pilot study, but I may do it in the real one.

Interviewer: Very cool.

Interviewer: Now, I just had a couple of questions about your thoughts on this workflow. Is there anything here that feels like it's not quite working or is like a pain point in the way that it's kind of set up?

Researcher: Pain point? I feel like the hardest, or the thing that takes a long time is when we find markers for the cell types, for our clusters in Seurat. I wish there's a more easy way to identify which of those markers would uniquely mark that cluster and not other clusters. Also, at the same time, have a good antibody that works with a known concentration that works because it just takes a long time to identify a good marker and then have it actually work. So, this arrow.

Interviewer: Got it. Do you wanna circle it in red?

Researcher: Yeah. This arrow takes a long time.

Interviewer: How much time was a long time?

Researcher: I'm in that process right now, so I don't really know, but it takes longer than the rest of this.

Interviewer: Is it like days, weeks?

Researcher: Yeah, maybe months-

Interviewer: Months.

Researcher: ... because it depends how many populations I'm trying to find.

Interviewer: That does sound frustrating. Anything else?

Researcher: I feel like all of this took a long time to learn, but I'm really, really comfortable with it, now.

Interviewer: The data acquisition part?

Researcher: And this, yeah. All of this, 'cause I was completely new to this before I joined this lab. It took a while, but then I set up protocols for that. Now, this whole thing is established.

Interviewer: Nice.

Researcher: Now, we're trying to establish this section.

Interviewer: The analysis piece is more-

Researcher: Yeah.

Interviewer: What are you hoping to learn in software carpentry, in the workshop?

Researcher: Well, I really liked ... For R, I didn't have as much experience in it before, and I tried to self-learn a lot along the way. Then, I went to the workshop after having tried to self-learn for so long. I felt like a lot of it was really basic for me, a lot of the intro stuff, but I wish someone had taught me that when I was starting.

Researcher: That's why I feel like for Python I'm at that place where I'm completely new, I'm really starting, haven't tried to even self-learn it. If someone will show me that the way they did for the R course, then it would be more proper timing. I feel like the R course kind of happened after I needed it.

Interviewer: What are you hoping to do with Python as part of this?

Researcher: Like I said, every time we find a function or something that's useful to find out about our dataset, we have other people in the lab who are doing other projects that also have single-cell data. I know that since our collaborative, like someone we're collaborating with is really pushing for that, I feel like that's something that's going to end up coming into the lab. I may not be the main person for it or it may not be my project that gets affected by it, but I feel like that will bring in a whole new set of functions and things that I will eventually end up working with. It might be more useful for me to learn that. Since we appear to be shifting that way, I feel like I don't exactly know how it'll be useful, but I know it's coming up.

Interviewer: Sorry, which part of the process do you think that'll come up in?

Researcher: All of this.

Interviewer: Then more data analysis?

Researcher: Yeah.

Interviewer: You said maybe instead of using these R packages, you'll use something in Python?

Researcher: Mm-hmm (affirmative).

Interviewer: Cool.

Researcher: Or maybe we'll use both, or it's not really clear, right now.

Interviewer: It is interesting that kind of Python/R back and forth. Great, well, that's really helpful.

Interviewer: Just to wrap up, I have a little checklist here. It's basically going over some-

Researcher: Shoot, I realize this part is missing here. That's the name of the technique that we're using. Didn't mention that anywhere.

Interviewer: That's great. Just a checklist of some behaviors that we kind of hope to teach in these workshops, just to get a sense of kind of what you're doing now. We've already talked about a lot of these, but as part of this workflow, do you use any programming languages like R, Python or the command line?

Researcher: Mm-hmm (affirmative).

Interviewer: Do you use or have you-

Researcher: Oh, the command line. One of the things for that from the last course, the R course, they had a section on that. That actually really helped me because I hadn't seen that before, but then, for not my project before another project, when we had to download a big set of data, I had to use command line for that. I was a little bit more familiar as to what I was doing and what everything looked like because I had attended the course, so that helped me.

Interviewer: Oh, that's excellent. Being able to grab that data, that's great. You don't use that for any of these pieces?

Researcher: I haven't used it for my project, yeah.

Interviewer: Do you have any step by step workflows that you've changed into a full script or function?

Researcher: In this section?

Interviewer: In any piece of this?

Researcher: Wait, can you say that again?

Interviewer: Yeah. Are there any kind of workflows that require, like do this, then do this, then do this, that you've changed until like one long script, or a function that brings it all together?

Researcher: I feel like we've standardized our process more into protocol, not really-

Interviewer: You can't just say like run?

Researcher: No, no, no, these are all like hands-on things.

Interviewer: I think in this case-

Researcher: This is more in this section, then. Run. I feel like we, there's a lot of sections that we run, but we haven't sped it up by combining things, no.

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

Researcher: No.

Interviewer: Not through Git, but you are using some kind of-

Researcher: Well-

Interviewer: ... date system?

Researcher: Yeah, yeah. I guess that counts, but not through Git. It's not very formal, but we know when things happened if we need to go back.

Interviewer: Do you use any open-source software? I think, well, R is open source.

Researcher: Oh, yeah.

Interviewer: Is ImageJ? I don't know about that one.

Interviewer: You are using R, so we'll say yes.

Interviewer: Do you share any of the code that you create publicly outside your lab?

Researcher: Oh, yeah. For the other project that got published with the [institute], they publish our code or every group's code, which is actually standardized.

Interviewer: Oh, cool.

Researcher: We had different organs, but all the organs had mostly the same code and that is published online.

Interviewer: Oh, nice.

Researcher: It's available for everyone.

Interviewer: For this one, is that something you were planning on doing?

Researcher: Maybe. I guess it depends where the project goes.

Interviewer: How to answer that one 'cause we're just trying to focus just on this one right now? Are you planning on doing right-

Researcher: I feel like I probably would. I feel like it would help others to see exactly and replicate it, yeah.

Interviewer: All right. Do you plan on sharing your kind of computational workflow, or your protocols publicly outside the lap?

Researcher: I feel like that's up to my PI.

Interviewer: Not quite?

Researcher: I don't know.

Interviewer: Do you think are they pro that or anti that?

Researcher: I really don't know.

Interviewer: We'll just do that question like not sure.

Researcher: I feel like it doesn't hurt, but yeah.

Interviewer: It's not always up to you, though. Awesome. All right, that is basically it. Hopefully, that was relatively painless. We'll come back in June and we'll talk more about this to see kind of where you are, if you feel like you've learned Python in a way that can replace this, or if it's still exactly the same, and you're like, "Python isn't for me," we can talk about that, too. That's all we're up to, today. This is for you.

Researcher: Thank you.

Interviewer: Thanks for coming out.

Researcher: Should I put my name or something?

Interviewer: Nope.

Researcher: No.

Interviewer: We've got an ID number for you, so we'll put that on there.

Researcher: All right, thank you.

Interviewer: Thanks for coming. I can lead you out of here, so you can find your way back. This one's for you, too, if you wanted to keep that. Have fun at the workshop.