**L**: Bigger picture, I have this model of microbial metabolism and it basically is like microbes in a box and some sugar and some oxygen and I define the rate at which they respire. And then I define the amount of energy that they need in order to maintain their biomass. So then there’s like a mass balance, so the details of the model are extraneous. Basically, what it gives me is a timeseries data set of a bunch of different variables: biomass over time, dissolved oxygen over time, and I’m looking at this box of microbes, and sugar and oxygen and then trying to characterize how they are changing over time, using some constraint-based modeling approaches. So that’s the bigger picture and then moving forward from here what I would like to be able to do is take a dataset where I actually have timeseries environmental data and then infer some of the parameters in that model using a nonlinear regression approach. So, this is kind of like the first step in trying to do that. [This model] was to characterize the uncertainty in my model parameter estimates and so I used this Monte Carlo approach where I ran the model forward with a known…The parameter I was trying to estimate was the amount of energy that a microbe needs in order to maintain its biomass. So, I ran the model forward with a known value for that and then I have this like known dataset. And then I added a little bit of error to that dataset with a certain standard deviation. I think I used like three different standard deviations to compare the results. Then I ran the nonlinear regression to try to use that synthetically dataset with random error to infer the what I call “maintenance energy demand”. So that’s what I did for three different synthetic error values and I think I did like 100 iterations of adding synthetic error and then estimating the parameter and then I can get this like parameter estimate distribution based on different values of synthetic error standard deviations. Does that make sense?

**A:**  Uh huh.

**L:** Yeah, cool.

**L:** So that was kind of the goal for this analysis.

**A:** So, if we walk through it

**L:** So, I walk through it, let’s see. This is just defining the times, vector of times.

**A:**  These three source files are three different types of functions that you’ve written?

L: Um,

A: You have functions from a homework and clever named functions, but are you subsetting these are functions for this purpose in this file and these are functions for this other purpose in this other file and then you’re importing them?

L: Yeah so, like these last two are kind of like part of a future R packages that my lab has been working on, so they are more generic utility functions for any kind of wrapper that you would want to write or any kind of…um, the model that I am using is called a gangsta, that’s why it’s a gangstawrappa. So there’s like different [my lab is so weird] that why we name these things because they just make sense. It’s a generalize algorithm for nutrient growth stochiometric and thermodynamic analysis, which is gangsta. So, it actually does have a like acronym. So, these are just like utility functions that are more generic that could be used for different kinds of analysis and then these homework 5 functions are for homework 5 that I did for Rob’s class. And those [hw 5 functions] are more specific to this analysis that I am doing here. So, I think this gangsta iterate is a function from there.

A: So you specify your sequence of times and then you construct your known model.

L: Yeah, and I guess what that gangsta iterate function…So my model is like a set of linear equations and then you can solve it for one time-step. But what this does is it takes all of the time-steps and then it takes the results from the end of the time step and inserts them as the initial conditions for the next time-step and then re-runs the model and does that for however many time-steps. So that’s what this does here. I just specify the number of realizations, and then makes them empty things to store and then this is like the Monte Carlo part where I add some synthetic error and then this is the nonlinear regression. I just used nls for this, but we also did some work with optim so that you can write a custom objective function and not use the minimum sum of squares. So you could use log-likelihood or we did like an absolute difference. We tried different types of objective functions. And then I do the nonlinear regression and I just extract the parameter estimates and then extract the predictions and store those. And then extract the predictions with some added errors so I can make some prediction intervals. I can show you some plots of what the output looks like.

A: So, you are extracting all of these and then you are plotting them here.

L: Yeah!

A: I guess I couldn’t figure out specifically from your code whether they were caterpillar plots or if they were just line plots...

L: (showing plot) So this is like what my model output looks like, is just like times and all of the different variables on the y-axis. And then these are the density distributions for the parameter estimates for the three different values of synthetic error. And then those are my confidence intervals and prediction intervals. And then I plotted in transparent each realization.

A: Nifty. Cool! Okay, so, I guess with that, um, like if we could think big picture wise of where you learned some of the pieces that are in here. I guess you are writing your own functions and then you’re…I’m trying to separate the stats from the coding, even though it’s coding for stats. But, so, it is interesting that you are using an NLS, but like that requires some amount of understanding of formulas and arguments that you are putting into there. So where did you see thing like that? Like functions and using NLS the plotting, the extracting, the looping for Monte Carlo, things like that?

L: Well, I was an engineering undergrad, so I had some Matlab and I remember we had to take an Intro to Matlab class freshman year and I remember being so frustrated during that class. It just didn’t make sense, like this coding stuff there is a jargon associated with it and I was just like…and I don’t think that it was taught very well. But still I just remember being really frustrated and towards the end kinda getting it. And then I had to use Matlab not like a ton, but a little bit throughout my undergrad, so I got some practice writing functions and for-loops. But I didn’t necessarily start to think about, and this is a bad example of it, but I didn’t start to think about optimizing my code or I didn’t even know what object-oriented programming was until I started working with Geoff. So I took a break in between undergrad and grad school and forgot a lot of that coding and then once I got to Geoff’s lab he had me do Swirl stats, which was like to learn basics in R. So it was kind of a refresher on “oh yeah, this is how you code, this is how you write a function.” I found that to be fairly helpful. And then taking Geoff’s class in Environmental Data Management really reinforced that, as I was taking stats too. I mean, so that reinforced everything as well, getting to practice.

A: So, I guess I remember talking with Geoff about what his class, or atleast what the R in his class was going to look like, and sitting down with him and talking through some labs. But, I’m not 100% sure, and I know what you saw in 511 as far as what Kezia was doing, but where would you attribute the majority of your function writing capabilities to?

L: Function writing…uh…yeah, that’s hard.

A: Like where did you see it first? In Swirl?

L: No, I would say in Matlab, in undergrad.

A: Okay. Then in your graduate coursework, where have you seen it?

L: Um, I’ve seen it in Environmental Data Management for sure, and then in we did some in Stats? And then in Rob’s class.

A: And then, I guess, the other big picture things are looping and we did that in 511, but did you see it other places as well? Have you seen it in other courses?

L: Yeah, in Matlab again, I certainly wrote quite a few loops. And then in Data Management we started out with loops and then talked about if you’re doing independent, if like each thing that you want to do is independent of the other thing, that you can use apply statements and that allows for you to parallelize things to make them faster.

A: Yeah, I definitely remember near the end of the semester you were using a lot of apply statements, and was that a push from your lab?

L: Mhmm.

A: Okay.

L: Uh, yeah, Geoff says that everything in life comes back to lapply.

A: That could be true.

L: I don’t know. Sometimes you just need a for-loop, you can’t use an apply statement.

A: That is true. But I guess for me because that’s not the majority of individuals’ experiences, like they don’t have those capabilities and I’m wondering as far as your learning of those types of skills, what was the most helpful, what resouces did you have available to you or did you find to be useful to you when converting yourself from the looper in Matlab to the apply statement user?

L: What resources…um…I think the most helpful thing for me is just seeing examples of how to use it and also the help page in R. I still look up the help page for lapply sometimes, because I forget.

A: Did you feel like when you were struggling did you, could you go to Geoff and ask him?

L: Yeah.

A: Okay, was he your primary contact in that process or would you instead go to your labmates first and then go to Geoff?

L: Um, yeah, I would probably go to my labmates first.

A: Okay, do you have people that you, outside of your lab that you have a good working relationship with, or is it primarily the people that are in your lab?

L: In terms of like R?

A: Yeah, in terms of helping eachother or people that you can go to.

L: I guess I would say that Rob Payne, who is on my committee, and who was a post-doc in our lab.

A: Something that was brought up to me that I hadn’t thought about was that not every graduate student has a lab environment, and so when people say that they feel supported by their lab members, I can’t infer that onto most people in these fields, because not everyone has that type of environment they can use as a resource. So, it is good to know if there are people outside of your lab that you find to be helpful, or if it is mostly that environment that is supporting you in your apply endevors.

L: Yeah, it would say that it is mostly that environment. Geoff actually had us rewrite lapply, and that was a useful exercise in understanding how

A: Was that in that class?

L: Mhmm.

A: Okay.

L: And it also helped that two of my labmates were taking that class with me, so we were all working on the same assignments and stuff.

A: How many people are in your lab?

L: Well, it is dwindling. There’s me and Sam and Katie and Byron, technically, but Byron is tryig to finish up. He has a full-time job and lives in CA, so I don’t know if I would count Byron. Oh yeah, and there is Anne Marie, who was a post-doc and now she is a research professor. So me and Katie are really the only…and Sam is finishing up and working remote from Madison, so me and Katie now are the only two full-time in the lab grad students plus Anne Marie who is a research professor.

A: So and you said that Rob was another person that you..

L: Yeah, he is definitely gotten me more in to like object oriented programming, since his background is in JAVA and stuff, so he is really in to like writing like classes and stantiating objects. So that has been kind of interesting. There are these things called R6 classes in R that they have some different properties than S3 and S4, that make them easier for, actually I think they are environments. I’ve been trying to learn about that in working with Rob.

A: What was, I guess maybe a question that I have is how did the beginnings of your work in Geoff’s lab look? So you did the Swirl stats, and was that the primary background facilitation and then you started your work with them? Or was there other building blocks that you did in order to acquire the skills that were needed for you to start working in the context that you are working in there?

L: Yeah, definitely taking those two classes, Data Management and stats, were super helpful for me to start working.

A: Did you know coming in to…I guess it’s pretty obvious once you start working, but did you know going in to Geoff’s lab that it was a computationally intensive lab?

L: Yeah, I had an idea. During my interview with him I remember him asking “How do you feel about computers?” and I had visited and saw that they were running a lot of models, so I did have an idea.

A: So you had an idea that that was important in terms of the future of research in environmental science before you even came?

L: Yeah, I think so.

A: Has your

L: But certainly, I’ve come to appreciate it more.

A: Yeah, how have your beliefs changed or been shaped more in the that direction over the couse of your. Do you believe that computing is necessary…

L: Yeah.

A: In order to implement statistics?

L: I mean, yeah, I think unless you have a super simple linear model that you can come up with analytical solutions, numerical solutions are necessary when you have these crazy models where these is just no way that you are going to be able to characterize your estimates without computing.

A: How do you envision these skills that you’re learning applying to your future?

L: Oh gosh, I don’t know. That is a tough question because I don’t necessarily know what my future holds or what I want to do after grad school. But, I certainly really like coding and analyzing data.

A: Do you think that this will help you get a job?

L: Yes.

A: These skills?

L: Yeah, it think it will.

A: And it opens the door to the different types of jobs that you can apply to?

L: I hope so!

A: Are you interested in staying for your PhD?

L: I am on the PhD track.

A: I guess, in terms of, you said that you used Swirl to learn stuff before and the help files are very useful in going through processes, but do you have go-to resources that you use other than Geoff, the help files, and Google…

L: Yeah, I spend a lot of time on Google and Stack Exchange.

A: Have you taken any online classes or tutorials? Things like DataCamp and other things that exist and are free and are intensive   
L: Not really, I’ve spent some time on Hadley Wickam’s pages. I think has some good resources. In terms of other tutorials, I tried to learn Python one week with DataCamp but then I got distracted.   
A: I did the same thing. When you are tackling a new computational challenge, do you have a process you go through? Do you say, “oh well this looks very similar to something we did in this class, I’m going to go look at my course notes or I’m going to bring out this textbook”? Or “I’m going to look at my code and see”? Do you have a process that you walk through and then you’re like “well, that’s not quite working for me, now I’m going to Google it”? Or, “now I’m not quite sure what’s going on with this Google page, I’m going to go ask Geoff”?  
L: Yeah, I would say that is accurate, but it is kind of an iterative process of making sure that I have enough information to start, and then starting, and once I hit a road block finding more information, and then moving forward, and then kind of bouncing back and forth like that.   
A: Something that I did want to ask you, is that given your position in a computationally intensive lab and the skills that you have acquired and the classes that you are taking, have you found that you provide assistance to your fellow graduate students? Are you someone that other people go to because of what you know about R?  
L: I don’t think so, no. I feel like everybody in my lab and most of the people that I interact with are also good at R. And like Katie, I feel that Katie gets a lot of that. She is the person that people go to if they want R help.

A: I wonder at what point, when you will become the new Katie.

L: I don’t know, maybe when Katie leaves.

A: I know that she’s had that experience and I was wondering if you had had similar experiences, but are you planning on taking more stat classes?

L: I would like to take 512, but I can’t seem to get it to it. Yeah, I would like to take 512, that’s the next step.

A: Have you considered the stats certificate?

L: I have, I think I would like to go that route. It’s just 4 classes?

A: Yeah, you do 511, 512, and then you do I think it’s 2 more on top of it. But, I think it was offered this semester, which would mean that it would be offered again, but Time Series could be good for you. If you are dealing with mostly time-series style data.

L: Yeah, there is a ton of autocorrelation in all of my data.

A: Yeah, but I think that might be more useful to you than Sampling or something like that. But they probably won’t let you in if you haven’t taken 512.

L: I think you can take them concurrently.

A: You probably could. It might be a bit intense of a workload.   
L: Yeah, that’s true.

A: But if you have nothing else on your plate for courses, that could be okay.

L: Yeah, I was thinking of trying to take 512 this summer.