M: [talking about a current course] Doing a scatterplot and finding the best fit line in Excel, like I was supposed to. And I got scolded for that. Because I was like “I know how to run it in R” but Excel, it just wasn’t working. I ended up having to copy and paste it from Excel, like the spreadsheet into Word and then copy it from one Word document into a brand new Word document, and then copying it from there into Excel again. And then took all of the formatting changes out.

A: But your task that you were given was that you needed to use your data, fit the scatterplot, find the line of best fit (linear line), and write the equation for that line, based on Excel’s

M: Yes.

A: And you were scolded for doing it in R?

M: Yes, because it is my adviser’s course and she wanted me to…I don’t use Excel, I like only use R if I’m using anything. So, I was like “I don’t know how to do this in Excel” and then I just figured “I’ll just make sure the data, there’s nothing wrong with the data period” because it just wasn’t…I don’t know what Excel was doing with the formatting. But it’s because I have the newer version of Excel, I think. Because it only happened on two of our laptops in the entire class, and everyone else was fine. So, I’m like “I know how to run it in R” and then I can check if there’s like any missing values or a reason for the error. And then it ran and it was fine and I got the values. I’m like “awesome”, now I have to figure out what is wrong with the formatting. And I did the copy and paste without formatting, and it still formatted it in a way that my Excel spreadsheet kept only reading two numbers and saying “this is your scatterplot – two numbers”, when I had like 10,000. But I did use R in that class, even though I wasn’t supposed to.

A: And that was somebody that is your adviser?

M: Mhmm.

A: But I thought you were advised by Chris Guy?

M: No, I’m advised by Christine Verhill.

A: Okay.

M: Yeah, Chris Guy is on my committee and Kezia is on my committee. And then Kevin Cavman from fish and wildlife services.

A: [explanation of why I’m asking for code and interviews] Go ahead and walk me through your code. So this is the code that you sent me. You can give me a background on what you are doing here, what you are doing for this analysis.

M: Okay, so I was using a linear regression to see if…We’re using something called a fat meter which is supposed to be able to measure the lipid content and/or the whole-body energy content of a fish. And it’s been validated in some monids but not in any Sturgeon. And they have used linear regression to validate the readings they are get with whole body proximate analysis. So, they take the fish, chop it up, and then combust it to get whole body energy, and then you do lipid analyses and protein analyses. And then they relate that back to the fat meter measurements. So, I’m halfway done through the whole-body energy now, as of Thursday, cause I finally got in the lab and had time to do it. But this is just the lipids here. So, for the fat meter measurements with people who validated with the monids, they took measurements from different parts of the fish to see if it was more representative of whole body energy in the anterior, posterior, or central part. So, I did the same thing. My data is kind of funky, because this is my pilot data before I figured out what sites I wanted to measure on my fish. So, it’s not, I don’t have the same sample size across, like I should. So, my first dataset were 10 fish from the fish tech center, and they were all small. They were probably, they were smaller than 300mm probably. And so I could only do the central measurement on them. And it was probably the second or third time I had ever used the fat meter, so I hadn’t really refined my measuring abilities. And the second dataset, these anterior and posterior measurements are on bigger fish, and I got those fish from Garrison damn. So they’re probably around this size [gestures with hands], so I can fit 2 or 3 different fat meter readings on them. So, the anterior and the posterior, and then this is before we decided to do a central measurement on every fish, to see if we could relate that central measurement to proximate analysis readings. So, I ended up having 9 fish in this dataset and that’s because I removed...So, I had a fish that, all of my fish had 30% fat content or less, and then I had one fish that was 2% and I thought that was strange. And the guy that I work with at the tech center was like, “maybe it’s was really skinny fish.” And then I went through and calculated the condition factor, which is the ratio of length and weight, and it was only…It had the lowest condition factor, but only by a marginal difference between the next lowest, and the next lowest was like 25%. So, I was wondering if I made a mistake when I was doing my lipid content analyses. And so, now that I have my whole-body energy stuff, I can look at the lipids, and then you can add all of the proteins and carbs in to see if it makes up a whole fish. Because if I see that it doesn’t make up a whole fish, then I made an error in my lipid analysis, I think. So, anyway, I removed that because it was two standard deviations outside of…

A: The question that I have regarding this line right here, is that you have removed your outlier, so you have a new dataset without the outlier, but I don’t see where you removed the outlier. How did you do that?

M: In Excel, cause I know that there’s a subset…that I could subset it, and I forgot how to subset it. And I was trying to crunch this out for my proposal, so I just wanted to get the data out, so I just went into Excel and deleted that and then ran it again. So, I was working with two different datasets. And I also didn’t know…I have like a weird…and this is probably, I could just subset it and then it makes its own new dataset in R, but I wanted to be able to look at them in Excel, too. And like compare them to each other, if I was making changes to one dataset verses the other, yeah, I don’t know. I did it in Excel.

A : Okay, um, so go ahead and walk me through your code, like what you’re doing as we go through.

M: So, I’m doing the linear regression of the upper anterior measurement of my fat meter on the lipid content of the fish. And then I was checking what the numbers were looking with summary and then I plotted it. And then this is different code than what Kezia gave us to run that, but I started looking at different things than what we were looking at in class. And then I changed my mind about everything that I wanted, and then I just found something online that kind of looked like what I wanted. So, I put that code in there.

A: Do you know where you found it online?

M: Not Stack Exchange, um, and I bet if I Googled it, it would be one of the top three hits.

A: So you just Googled it? What did you Google? Did you Google plotting in R…

M: Yeah, linear regression plot or scatterplot. And then I said that I wanted that abline, cause I couldn’t find that in my notes. Because I know we learned it, but I just couldn’t find it in my notes, because I didn’t have my notes and I was looking at my old R script. And I couldn’t remember what lab we did it in, so then I was looking through my old lab scripts.

A: So, is this, was this for your own research or was this to get something that was exactly like what you would have gotten in Excel?

M: This was for my own research, I didn’t do any of this in Excel. The only thing I did was look at the data, well…I punched the data in in Excel, and then I imported it to this.

A: I was just thinking that, this is a scatterplot and there’s your linear regression line, and that’s exactly what you’d get out of Excel.

M: Yeah, and I wanted to look at the plot and then I just got the linear regression line out of that. Because I wasn’t sure there would be anything, and I didn’t know, like I was like “I don’t want it to plot something if there’s nothing”. Which I ended up doing further down and then there was nothing. I was like that’s not really a relationship, but I put it in my proposal and then my adviser was like “why did you put that in there?” And I’m like “because they need to see that there was no relationship”, and she was like “don’t’ put a line if there’s no relationship.” And I was like “yeah okay”, so I need to take the line out. Um, yeah, so I was just finding the best fit line and then…

A: Okay…then…

M: Yeah, and then I should have put a break, because this is now the posterior measurement. And these are all with the outlier in there, and then I ran them all again with the outlier removed.

A: Okay.

M: So, I wanted to see how it changed the line, and then where on the scatterplot the outlier was. And it was like, if all the fish are on this line here, the outlier is over in this incredibly wrong area. And then I re-ran everything the same way, just with the outlier removed.

A: What’s the reasoning behind, where’d you use this? [pointing to a line of code]

M: I don’t remember why I put that in there. I was like, “oh, there’s something”. That’s a confidence intervals, right?

A: Mhmm.

M: Yeah, so I was like “is this, does this make sense?”, so I put the confidence intervals in there, and then I never did anything with them, and I don’t even remember what they ended up being.

A: Mhmm.

M: But that ended up not being, it was non-significant relationship, but it wasn’t, it was like 0.06. And then there’s only 9 fish. So, it’s kinda like, “okay, well that’s hopeful that we’ll get more fish in”…

M: Oh! I did this because, the way that Kezia had us write our results was like “I’m 95% confident that” and then we give the upper bound and the lower bound. And when I did that I was like “wow, so the relationship was like between...like I was confident that there was either no relationship or there was a relationship”, was essentially what I was saying in one sentence, because the confidence interval intervals were so big.

A: And it contains 0?

M: Yeah, so I was like “um, I’m just going to say that it’s not significant, but that it potentially will be after.” So, that’s why I had that in there. and then I didn’t use it, because it was telling them that it’s maybe zero. “I’m confident that it’s maybe 0 or 1,” and I didn’t want to say that.

A: So, the code that you found on the internet, did it have this format? Like, with the data, plot y ~ x? And then this piece, but this piece is different. It’s not, like not, most people don’t use that piece.

M: Oh, that’s a Kezia piece.

A: Okay.

M: Yeah. All that is stuff that I remembered learning in Kezia’s class. It was just that I forgot how to plot it.

A: This was really just the with this,

M: Yeah.

A: Okay. So, um, I do notice that you are using a with here, so with your data, plot these guys. But then here when you are fitting your model, you say fit the model of data $ something. So, you have…they’re slightly different from each other.

M: Yeah, and so that [pointing to $ code] is what I think we learn in Kezia’s class. I think we did talk about this, “oh if you don’t want to do $ and call the data every time, you can do it like that, where you call the data first and then do it after.” I think we learned both, but I did this mostly when we were in class, because that’s what I remembered the most. And then I forgot everything, because I didn’t get to work with data for a long time.

A: Okay.

M: Um, and then I started doing that, because I found that online and I was like “okay”. And it wasn’t getting an error, I don’t know why I changed, but it wasn’t getting an error…

A: No, that’s fine…

M: I just ran with it…

A: I was just noticing that you were doing it one way, like you were using the with when you were plotting and you were using $ when you were fitting your models.

M: Yeah.

A: And I was wondering if you preferred one or the other, or like why you were using one method for some and one method for the other.

M: I’m trying to remember if there was...so my dataset wasn’t just this, there was also like length, weight, and then other variables in there. And I had a hard time, I don’t know if it’s cause I was being stupid and I had a hard time calling in the beginning and I wasn’t getting what I wanted. So, I’m like “okay I’ll just bump it down” and then do the with to see if I can get the plot that I wanted. And then I did that instead.

A: Okay.

M: And then it worked, so then I just copied and pasted everything and kept working with that.

A: So, then you’re doing the middle measurements of your fish?

M: Yeah, so these are the middle measurements and this dataset was…I think this was actually 12 fish, because I had 2 fish that were big enough that they had three individual measurements, so they had a central measurement in addition to their posterior and anterior. In addition to all of the small fish, and that was like not significant at all.

A: So, these are different fish than these fish here?

M: Yes. And you’re calling…

A: …the name of the dataset you’re using here is the same as the…

M: So, all of the fish are in the same dataset, but only some of them had the middle measurement. So, these are all of the fish that had the middle measurement.

A: Oh, how did you subset out the ones that had the middle measurement?

M: Um, that was just in the way that I recorded it when I was making the measurements, I had different columns. So, I have like the anterior measurement is one column, the middle measurement another, and the posterior the other.

A: Okay, I’m just thinking from the code standpoint, if I were to run this line of code here, this lm of these guys, and that should be the same as this. Like, this code right here is the same thing as this code right here, but one is middle, and one is anterior.

M: Yeah.

A: How are you separating the data for this linear model, that are using the anterior, those 9 fish, verses these 12 fish that have the same exact model. Like, how does R know that this dataset is 12 fish and that dataset is the other 9?

M: I don’t know if I’m right, but I thought this is how it worked…um, because it’s calling that column that only has 12 values, for those 12 fish.

A: Okay, so you have the whole…so these 9 fish are…

M: Oh, woah, does that not make sense?

A: …in these 9 fish are in these 12 fish? So, you have 12 fish total, and then you have this column that’s like “p\_sum”, right? Is that the middle measurement?

M: Yeah.

A: So, then out of those 12 only 9 have numbers there?

M: So, it’s um, let me get my…I brought my laptop, I can show this to you. Let’s see…um, so this, these are okay, this is kind of ugly and hard to understand. So, I have my mass, my fork length, the lipid content of all the fish that I analyzed. So, I had 20 fish total. And then the upper anterior measurements were only on these bigger fish that were from Garrison dam, so they only have the 10 values there. And then I was looking at logging them and seeing if that made a difference, and I couldn’t remember how to do that in R, so I just logged them all in here and made a new column in here. And see that’s the…I knew that there was a way to like add a new column in with how you wanted the data changed from, and I remember doing that and thinking it was really hard. It was one of our first labs, and I was like “okay, I’m just going to do it in Excel because it’s faster.” And then I have my middle measurements, so this is just the first 10 fish have only the middle measurement, because they were too small for other measurements. And then I have these two middle measurements here that I couldn’t remember, because I didn’t take good notes, on whether or not they were the middle overlapping measurement or they were just a central measurement. At the time we didn’t have middle overlapping measurement, so I think it was just the middle measurement, like separate completely from the anterior and posterior measurements. And I have this here because I was trying to work through my mind if it was middle overlapping or not. And then this is the upper…what was I doing here…why is this piece UAB here..? Okay, I’m not sure why this is here, yeah, I’m going to ignore that because I don’t know what that is. But this was the upper posterior measurements for these same fish down here, so they have two measurements across, aside from these ones that have the 2 middle measurements, too. And then I logged all of them, and then I don’t know why I logged them like…oh, I was logging these middle measure, no…what is this...J2…oh, um, so I took, I had multiple measurements on my fish, on these fish, because I was practicing with the fat meter. So, I was taking the average, I think, of what…something in here is an average. I need to go back through and figure this out, because it doesn’t make any sense. And then I calculated the condition factor for the fish, where I removed the outlier from. And I only removed the outlier from the second group. Um, so, I just calculated the condition factor for that second group. And this fish is the outlier at basically 0.8 and this one is 0.82, but it’s lipid content was 16.5%, where this lipid content was 3%. So, I pulled that guy out as the outlier. But, yeah, so I just have empty spaces for where they didn’t have measurements for.

A: Okay. So then, these fish IDs is a PS for certain sites and a GD for other sites?

M: Yeah. Um, so I made the poor decision of doing PS, like pallid sturgeon, but they’re all pallid sturgeon. So, I changed that when I went to Garrison Damn. So, I was like, okay Garrison Damn, these are the 10 fish we took from Garrison. And the new fish that I have are TU, so thermal unit, cause they are my thermal unit study. So, they have their own designation. I’m trying to do it as every site or every experiment has their own designation.

A: So, all of these models, these linear regressions are on separate, none of them are on fish that were both here and there?

M: Uh, except this one. So, these two fish are in the middle one. So, there’s 12 fish there in the middle measurement.

A: Oh, okay.

M: I ran it as this column for the middle measurement.

A: Yes, so then here, when you’re calling that column you get 12 values. Because it removes the ones that don’t have values.

M: Yep.

A: Yep, okay. Um, so then when you took the log, how did you, was it, is that in the literature that the log is pretty common for those measurements or did you decide yourself that you wanted to take the log for those measurements?

M: In the literature they did it. For that one study that “validated” it for the somonids, they took the log. And I didn’t include it in my data, because when I ran it, really briefly, I just ran it really quick, I don’t know if it’s even in there.

A: Here’s the log p sum…

M: Okay, it was um, it wasn’t more significant. And I didn’t want to like say, “oh look I logged these data because of a reason that I didn’t have.” So, I thought, “okay, I’ll just start with my 9 fish,” and then once it gets bigger, maybe, I’ll have more of a reason to keep the data logged, or to not log it. And then figure it out then. Um, so I just wanted to give them my preliminary data. Which that was all kind of rushed, and not well thought through.

A: Okay, um, and these were the means and standard deviations for…

M: the lipid content. And that’s when I was like, “oh, is this two standard deviations outside of the mean?” and I said yes.

A: For your outlier?

M: Yeah.

A: To decide if your outlier was an outlier?

M: Yeah. But, um, I talked to Christine and she said that I couldn’t do that. Because we’re trying to study fish that are pushed, like really unhealthy. So, really unhealthy fish are going to be outside the two standard deviations. But then I said well I don’t think this lipid content makes any sense, considering that it’s over 4 times smaller than the lipid content here, but it’s a marginal difference in condition factor. But condition factor doesn’t mean anything, like that’s the whole point of my study. Is like we’re looking at condition factor to say that it’s representative of health, but is it even representative of health? Does it even show the differences in lipid content here? But I just think that 3.5% sounds like nothing, so I’m going to look at the energy. And then I think that once I look at the energy and the protein, the fish would have to be all carbs and moisture to get that value. Which, I have the moisture content and I just haven’t looked at that yet. I need to look at that and see if it all adds up to 100%, then that fish just had 3% fat and the I’ll put it back into the dataset.

A: That is an interesting check that you have for your each of your measurements, that you can look at the sum of all of them and you know what they should be. Is it within a margin of error?

M: Yeah, it’s within a margin of error, and we’re also not specifically going to be analyzing samples for carbohydrates, because carbohydrates are such a minimal amount of fish composition. So, and I don’t have it, but I have it on my phone, so I’m going to look at it really quick, so I can tell you it…I forgot to bring my flash drive to download the data that I took when I was last at the tech center, so I had to take a picture of it, because I was scared that if the tech center computer crashed I would all my data.

A: That’s a valid fear.

M: So, that fish hasn’t been analyzed yet, but with the other fish…so, I need to do another calculation, because…so, these are percent [pointing to data]. So, lipid is percent of the entire fish composition, and then we calculated the energy and calories per gram. So, we calculated energy in calories per gram, but then we convert that to percent of whole body composition. And then from there I’ll have the two values plus I already got moisture, I just wasn’t paying attention to the moisture when I was looking at that, so that would be in here…yeah, so the moisture is in here. So, actually it should just be moisture and lipid should kind of add up together, cause it’s an inverse relationship between lipids and moisture…[calculations]…no, that doesn’t add up to 100…um

A: Do the measurement devices that you have, have margin of errors for the device?

M: Yes, and this was just the, we burnt it, and we took the weight before we burnt it and the weight after, and saw how much came off of it. We also, this is another question that I have to ask the guy that I am working with for this. We freeze dried the samples after, so we…for the moisture content, we took them and stuck them in a furnace for 3 hours or something. And then for the fish, like the homogenization of the fish before we did the lipid content, I freeze dried the samples. So, this isn’t actually the same sample as the moisture, so if I did a poor job at homogenizing my sample that I then burnt for the moisture content, then that would explain why the readings different or why they wouldn’t add up. Because these are, I’m like pretty sure these are much more homogenized than this, because this was like I had run these through a meat grinder and there were chunks of bones and fins and stuff, and I just like put some of it on a plate and stuck it in the oven. Whereas these were run through the meat grinder, and then they were freeze dried, and then they were ground through a coffee grinder after that, so they’re more homogenized. So, that might be something. I actually don’t know why we did this, I think this is just a, if we had a really weird value, we were supposed to see that it was…so, let’s see…yeah, I have to talk to Gibson why we even ran moisture. Cause we weren’t planning on using this, he just had me run it, and didn’t really tell me, he just told me there was an inverse relationship between moistures and lipids. In which case…I think this one is 3.52…and this one is the highest value, but I don’t know if it’s so much higher than the rest…but, that’s maybe what this is for, is to say “yep, this guy’s way higher than the rest”, meaning that this we’d expect to be way lower than the rest, so I need to ask…

A: So, when you read your data into R, cause you don’t have a data savement up here, do you use, how do you read your data into R?

M: Um, I do…somehow I think that it will be different…oh, I do import dataset.

A: So, you do import dataset, I guess that’s just…okay, so you import your dataset from Excel.

M: Mhmm.

A: And then you find the data that you want, and then do you select the sheet that you want here? From your dataset?

M: So, for these ones I had compiled them into its own sheet.

A: Oh, good so it’s just one sheet.

M: Yeah, this one has like 10 sheets.

A: So, when you…in order to run this analysis, is the only thing that you did with your data to run this analysis was to go into Excel and remove the outlier?

M: Yeah.

A: Okay, so you didn’t have to do any sort of reformatting?

M: No, everything worked. And I probably, I worked around my data, I didn’t work my data to work into that. I was just like “what way can I run it that it is going pull what I want, and I don’t’ have to change it?” I think I’m still bad at figuring out what I need to change. Actually, I think I changed the titles to things that I needed to write in there and not have to remember if there was a space or not, so I left fork length, because it didn’t matter because I wasn’t running that. But I put all of these, they were all different that what, um…actually, they’re not all different from them as on my handwritten datasheet, but I’ve made them easier to put into R.

A: Okay. So, in general, like, the skills that you are using here, when you’re running your code here, where would you attribute learning most of these things from?

M: I’d say to Kezia, but I also think that I don’t do it the way that, like I’m sure that I was better at making sure I did everything right when I was in her class, because we like had checks and I remembered all of the checks, and I was like “ah, okay.” And now I’m kind of like “oh God, I can’t remember anything” so I’m just going to put it in and see what happens, and “oh, that kinda makes sense, so I’ll do it again.” Everything I know in R is because of Kezia, but I also have lost a lot of that…

A: So when you put it in and you run it and it doesn’t work out, where do you go next?

M: Um, I’ll go to my notes and then old labs and old homework files and then see if I ran something similar, then I’m like “okay that broke and that’s where it broke because I did something different there.” And then I’ll change the code to look more like the code that I had in Kezia’s class, and if I can’t find it in there because I saved my file under homework 6, but really it was homework 7, and I’m looking for homework 7 and it doesn’t exist, then I’ll go into Google and then run a search in Google for that.

A: Okay, have you taken any other classes that have helped you learn skills that you feel that you are using in your data analysis?

M: No. I think it’s, I’m reading that Ecological Statistics by Gotelli and then that kind of is like “oh yeah, I remember learning that with Kezia” and then I feel like it’s refreshing me a little bit, but I haven’t…I was planning on taking 437 next semester, but…because I’m doing principal component analysis, I’m supposed to be. Um, but Christine didn’t want me taking it, because she thinks it’s going to be too much of a course load. So, instead…

A: Is that Sampling? What’s 437?

M: 437 is Intro to Multivariate Analysis and it’s with...

A: Jenny.

M: Yeah. It’s with Jenny and I didn’t tell her I dropped the class yet, because I haven’t yet, but Christine really doesn’t want me, she thinks it’s just going to be more than what I need to know to learn principal component analysis. Um, but I really like learning things from like the beginning, and I don’t mind the theory because I think that’s the way that it makes sense. Um, but she has her decision made and wants me to do an independent study kind of more instead. So, like I will learn principal component analysis on my own with the help of Kezia, which I haven’t even talked to Kezia about this, this was like all in the last two days that I said “oh, I’m going to take 10 credits” and Christine was like “nope, too much” and we’ll just do independent study for this and then we’ll use Dave Roberts and Kezia, and then you’ll teach me PCA, and that will be the goal of the independent study. And I was like “okay.” Um, so that’s the next class that I’m going to be taking is the independent study…

A: What are your other credits that you’re taking this semester?

M: I’m taking ictehology and then um, what’s my other class…oh, the communications course in Ecology. And then on top of that I wanted to do my thesis credits because my sampling will be done in March. My experiment’s done in March. So, I just do a huge sampling thing and then get to start writing, and I wanted to have thesis credits so I could potentially finish by the end of fall. Um, and I didn’t want to load like 20 credits on my fall semester, cause that didn’t make sense in terms of like money. Because I don’t get comped. I only get 10 credits as a TA covered. So, I wasn’t going to be like “yeah, I’m going to save these 3 credits and then max out and then have to pay my tuition because I wanted to get finished in one semester.” So, they’re just thesis credits and that independent study is supposed to be 3 credits.

A: So, what have you taken so far in your course work?

M: I took stat 511 and human dimensions in my first semester, and then I took habitat management and Christine’s Ecophysiology course last semester.

A: And that’s, cause you started in the spring?

M: Yeah.

A: Okay, um, are there other more quantitative classes that you expect to be taking?

M: Um, in the fall I’m taking…I have my program of study with me.

A: I’m impressed.

M: Well, I was supposed to turn in, but Diane left on like Tuesday last week and I didn’t get my last signature til Tuesday. So, I’m planning on taking spring, no, yeah spring would be Christine’s independent study class and then thesis credits and then Tom’s class, which I think I totally…and then fall 2019 would be foundations of ecology and then 7 thesis credits.

A: Is foundations of ecology, is that the WILD class? Which one is that?

M: Foundations of ecology is the BIOE 554.

A: Who’s that taught by?

M: I don’t know, I can’t remember. It’s whoever is teaching it this last semester, but I don’t know.

A: But that’s within the ecology department?

M: Yeah, and I don’t think that they do anything quantitative. The only quantitative class that they had in the ecology department for the fisheries focus was Chris Guy’s class. And I took the human dimensions class instead, because my friends were saying that it was helpful to have data for Chris’ class, and if you didn’t have data that it was kind of hard to visualize anything. Cause then you’re using data that’s not yours and then you’re kinda just…um, so I skipped that class. But then it’s not offered until 2020, so I’m not going to stick around for that.

A: Interesting.

M: But, honestly I wish I was taking more quantitative courses, because I think, like I learned a lot in Kezia’s class and it would have been really awesome to like continue with the math. So like, I feel like “oh, yeah, this is what I learned in there” and then build off of it and then being able to use it. Versus like, “okay you took it” and then now you go to analyze your data and not really know what you’re doing. And I realize that I don’t think a lot of people take stats classes, but then they’re sitting there analyzing data, and I’m like “that is so hard”, like you don’t, like, you’re just sitting there like “what statistical analysis should I run?” and then you don’t even know what they are, except reading in the literature. Which, is helpful, but then also like want to run a factor analysis instead of PCA, but I don’t know what either of those even mean, so how do I figure that out? So, I’m kind of bummed about not taking 437, but I’m just trying to graduate…

A: No, I understand. In terms of the resources that you use, do you have people that you talk to about your code?

M: So, not my code, because I usually can figure out, if something’s broken I’ll figure it out. But, like, I talk to Kezia about PCA and like “oh, yeah, that makes sense for your project.” So, it’s like, “okay.” So, I talked to her about figuring out that that makes sense, but that’s it.

A: Okay.

M: Which, cause Christine doesn’t have like an extensive stats background. Like she does stats, but not like the multivariate stuff, so she’s been talking to Dave a lot and then telling me what Dave says.

A: Okay.

M: Um, but, yeah.

A: Okay, in terms of like R, future R help, who do you expect you would go to?

M: I’d probably try and figure it out and then I also realize there’s a math help center downstairs and I was wondering if they can help with R…

A: [shakes head]

M: No, okay. Yeah, good like I didn’t bring in my R and go “principal component analysis”.

A: Yeah, they only do help for physical courses and they don’t do help beyond stat 217.

M: Well, there was a kid in my writing group, like with the writing center, and he’s in the stats department. And I was like, “oh maybe he’ll help me”.

A: Who’s in your writing group?

M: Bright.

A: Oh, yeah.

M: And then he stopped coming to the writing group, so I was like…

A: My writing group also went slightly defunct.

M: You can come to ours next semester!

A: Is yours the writing center?

M: Yeah, and were getting combined…you were with Erin, right?

A: Yeah.

M: Were you with another girl?

A: I was with two other environmental scientists and they both kind of went caput.

M: Cause I know that…

A: Yes, I was with a woman and a man.

M: Okay, cause Erin said that there was….

[writing group]

M: But, I call my dad.

A: Oh!

M: So, my dad is a hydrologist who works for Esri. And so he’s a phd, and I was like “have you done principal component analysis?” and he was like “yeah.” And so I’m like “great.” And I was talking to him on the phone with it and then I realized that I didn’t really know what I was trying to measure, because he’s like “you can’t do that.” He’s like, it kind of makes sense if my variables were this instead, and then he explained my variables to me. And I was like “okay.” And then I went and called Kezia and was like “these are my variables,” and she was like, “yep, that’s right.” And I was like “thank you” to my dad, because he figured that out for me.

A: That’s an interesting resource that I’ve never…like I’ve never experience that…

M: Yeah, my dad and then, um, one of the kids that I raced with on the cycling team in undergrad, he is working in a cancer research center. And I was like “have you done principal component analysis?” and he was like “yes.” And then so I talked to him on the phone about it, too. So, that was my intro to that, plus reading principal component analysis in a book and watching some video on YouTube. And then I went to Kezia and told her what I thought it was and she was like “yeah, you got it” and I’m like “no I don’t, I’m just saying everything that everybody said to me.” But, that doesn’t make sense, because I haven’t sat and learned the theory behind it. I feel like I’m going to struggle, cause I think that it’s really beneficial to sit in a class and like understand from the beginning why we’re doing it, versus like, oh it makes sense to do it, so I’m just going to do it.

A: I understand, that’s kind of how my brain works, too.

M: I mean I’ll do it.

A: There are point that I can understand that like, when especially with a high credit load that that would make sense. If your adviser thinks that its’ going to be a really difficult class and it’s going to take more of your time that what you necessarily have and, at a certain point advisers believe that part of us having higher education degrees is proving that we can teach ourselves stuff.

M: Yeah.

A: Which is great that you have all of these resources and you’re using them, and then you have somebody to talk to that says like “yes, you’re kind of getting these pieces” versus “no, you’re not understanding these pieces of PCA.” Are you expecting to do PCA in R?

M: Yes. And Kezia gave me some R code. She was like, “try using these functions”…functions?

A: Yeah.

M: ...I can’t remember anything…um, and then I was like “okay.” And then I haven’t done anything yet, because I’m just trying to…I’m like, so the government is shutting down on Monday.

[talking about stock piling food for fish]

M: Okay, I’ll learn principal component analysis later, because right now I have all of this stuff because the government will shut down...

[talking about fish dying]

A: In terms of your timeline, moving forward, so this code are you going to be modifying this code?

M: Yes.

A: For the larger group of fish that you have?

M: Yeah, I have 135 fish, well actually I have more than that because I have some shovelnose sturgeon, and a few before I died…

[sampling and analyzing fish]

A: In the spring you’ll be measuring your fish and collecting data on your fish?

M: Yeah.

A: Okay, then are you hoping to run an analysis or run something similar to this on those fish?

M: Yeah, so it depends. I’m also taking blood and something else, and the blood I haven’t learned…

[talking about training her tech]

M: Once everything gets done, I’ll probably just fix the code so that I can import the data and run it again, and then have a refreshed . As I go I’ll be working on that. Hopefully, by the end of June all of my samples will be analyzed.

A: And then you’d be starting back on all of this?

M: That and principal component analysis.

A: So, if I wanted to talk to you again, when you’re doing more of this sort of coding stuff, we can probably hold off until early summer?

M: Early summer or late spring, just depending on my time.

A: I can check in with you and see how you’re doing.

M: I’ll tell you if my fish start dying before then.