FACULTY INTERVIEW

Dr. Cort Rudolph • St. Louis University • Associate Professor

Q: What struggles do you see with students as they learn R or what is a barrier to them even adopting it in the first place?

A: Honestly, I think the biggest barrier to learning R is learning SPSS initially. SPSS can serve as a crutch that people use as a benchmark for everything that they learn in R. For example, initially, some people really miss the comfort of a point-and-click GUI-interface. Thus, I think that one barrier stems from the fact that, initially, R is introduced from the command line. When I introduce R, I teach with RStudio and R Markdown from day one — this gives a bit of point-and-click like functionality to the process. Another barrier, is that people miss "seeing" their data (e.g., in a manipulatable spreadsheet-like representation). That, and the fact that variable transformations, like scale scores, "live" in one's code (and not in the data, proper) are difficult SPSS-induced habits to overcome.

Q: What impact have you seen institutional supports, such as course work in R, have on student adoption of R as their primary statistical software?

A: I teach a number of statistics and methods courses in R (i.e., univariate, multivariate, psychometrics, SEM, & meta-analysis). Speaking on behalf of my students, I think they would agree that the advantage of R as a true "statistical computing environment," is that it provides a (nearly) "one stop shop" for all stages of an analytic workflow (e.g., data manipulating, modelling, graphical representation, reporting). R and related tools (e.g., RStudio, GitHub) streamline the research process, and allow for the easy adoption of open science principles as well. That, and I think they have seen advantages on the job market (e.g., in terms of possessing these advanced skillsets).

Q: In your opinion, what are the pros and cons of teaching introductory graduate-level statistics courses in SPSS vs. R?

A: I think "easing" from SPSS to R is actually a big part of the barrier that many perceive with R. The problem with SPSS is that it gives the false impression that data analysis, broadly defined, is a point-and-click endeavour, with "one best way" to do any number of tasks (e.g., consider how t-tests are implemented in SPSS). It's not that R is really that difficult to use (by-and-large, the syntax itself is far simpler than SPSS, for example), it's that SPSS has oversimplified the process as a whole. Take the popular "PROCESS" macro, for example. If tasked with explaining exactly "what" this "process" is doing, I would wager that most SPSS users would not be able to explain the under-workings of these models. However, if tasked to specify such a model in R using a series of `lm` functions or via `lavaan,` one would by default much more clearly understand what's going on "under the hood," so to say.

Q: What are the realities of moving your research related analyses or course content to R and why is workload worth it?

A: The realities are that you will gain a tremendous amount of control over the entire workflow of your research and teaching process. However, like most things that are worthwhile, this takes time and discipline. There will be an initial amount of effort required to adopt R as your primary analysis framework, but really that workload is only a "one time thing" in many cases. Once you have a good workflow (i.e., a workable way of doing some "thing" or set of tasks, e.g., cleaning data or specifying a CFA), then you can always go back to that workflow the next time you need to use it. I always tell people that the advantage of R, is that "if it worked for you once, it will work again." For teaching, a distinct advantage of R is the ability to have a single R Markdown file that contains all of the embedded scripts and supporting content that can be "knit" together into a set of slides that can be easily shared with students.

Q: Overall, what is your "elevator pitch" for students considering learning R?

A: R is an in-demand skill: universities want people who can teach statistics and methods classes in R and companies want people who are likewise competent in R. Simply put, R is how people "do" statistics in 2020. R is the "lingua franca" of statistical computing, and should be adopted as everyone's primary language, as such. As a field, we are moving toward a more open science view of research, which will eventually require you to not only share your results, but also the data and code the led to such results. Because data and analysis code are packaged together, R facilitates the adoption of such open science practices directly in this way. Moreover, because it requires you to write down each step of your analysis workflow, R is a key tool for facilitating reproducible analyses. If you are trying to sell the idea of adopting R to your administration, remind them that all of these advantages are free!

Dr. Robert Partridge • Wayne State University (WSU) • Associate Professor

Q: What struggles do you see with students as they learn R or what is a barrier to them even adopting it in the first place?

A: The biggest thing I see as a barrier is that they feel intimidated by code or the prospect of coding. Despite most students growing up in a computer / software mediated environment the way users interact with software is quite distant from the coding world. All of the student's intuition about using software is challenged when they are faced with a blank coding console which is disquieting. Then the code itself is like an alien language and they have to learn not only the vocabulary (i.e., what the commands are) but also the syntactical structure (i.e., how the commands are organized) and there is not a comparative cognitive framework to which they can translate things. Like oh this command in R is the equivalent of this command in X—it has all been hidden from them. Students that have a history of programming or writing syntax in SAS, SPSS, STATA etc. pick things up much more quickly because it fits their schema about how the process works. The other challenge is that many of the support materials are written in computerese and so just trying to figure out what the manuals are saying is at first very frustrating. Until they start thinking in code and develop that schema about how code works the manuals and user-guides are just more gobbledygook. When you add all this to the time pressures that students and junior faculty are under all the incentives are to just avoid the learning struggles and go back to what they already know because it is faster - like a toddler; you can go more places once you master walking but for a while crawling is still the fastest way to get from point A to point B.

Q: What impact have you seen institutional supports, such as course work in R, have on student adoption of R as their primary statistical software?

A: Having institutional supports is essential and I have certainly seen student's comfort level and positive attitude toward R increase substantially by having those supports available. Students find themselves almost having no alternative but to use R at some point. After the first introductory seminars in Stats, the majority of analyses require a more specialized software than SPSS and these are more often than not very expensive and students have limited access to them. R being free

without the restrictions of a student demo version of some other software make R the only viable option for many students both in their research and for advanced classes. Students can usually work out the code for the specific analysis on their own, but the data management side is really challenging because that is the most foreign aspect of working with code for students, so it is the simple things like just calling in their data, organizing it, recoding variables, etc. that can be so frustrating for many students. Having formal seminars and developing a learning cohort of students that can support and learn from each other is the key to truly learning R as a primary research tool. Once student's see the power of what they can do with R and have a support system that helps them get there they can't imagine doing stats any other way.

"Piled Higher and Deeper" by Jorge Cham www.phdcomics.com

Q: In your opinion, what are the pros and cons of teaching introductory graduate-level statistics courses in SPSS vs. R?

A: First and foremost I think that statistics has to be understood as software independent on one level. I also believe that starting with SPSS and moving to R might be doing more harm than good. Using the language analogy, using SPSS conditions a style of thinking about analyses or set of schema about how to work with data that makes it harder to learn R or Python afterwards because you have to unlearn those intuitive was of understanding how the analysis is being carried out. If you start from the beginning with R it make take a bit longer to get rolling, but the learning becomes deeper and more ingrained because your schema is consistent from the beginning. I might be biased because I am very old and learned to run stats on a mainframe and none of us knew what we were doing at first but after a month or so it was old hat, so I know that even if it is uncomfortable at first, it can be done.

Q: What are the realities of moving your research related analyses or course content to R and why is workload worth it?

A: I think the key to moving to using R for my primary analyses in both my own research and in classes is nailing down the fundamentals of reading in the data, managing data sets, etc. I have been slowly adding R to my SEM and Longitudinal Modelling courses and the biggest challenge was just getting the data to a place to be analyzed so I mostly just showed example code. After spending the summer learning and practicing with the basics of R I found that fully integrating R with full examples into my lectures was much easier. I also like the fact that with R I can include all the members of my lab into the data analysis process. I don't have to have multiple copies of Mplus or SAS or SPSS and the analyses don't have to be done while physically in the lab. It really opens things up. I also feel like that adds more training value to the lab because they are getting trained in both substantive research and developing proficiency in R. I think the perception is that making the transition is really hard and time consuming, but I have found that it comes down to about 3-4 weeks of really devoting time learning the basics and after that it all comes much faster and the benefits really outweigh the difficulties.

Q: Overall, what is your "elevator pitch" for students considering learning R?

A: You will never have more time available to you to focus on learning R than you do right now and it is becoming essential in both the academic and private sector to know how to use it. If your department doesn't offer formal supports find an "R buddy", schedule a couple hours a week to learn, and take advantage of the enormous resources online— it can be overwhelming but reach out to the R community (they are more than happy to tell you about R) and ask for advice and find a tutorial, blog, video series, that works for how you learn (an option you don't always have in the classroom). Then roll up your sleeves and follow Ms Frizzle's advice to take chances, make mistakes, and get messy; you won't break the internet, I promise.

Dr. Robert Partridge is an Associate Professor in the Department of Psychology at WSU and the director of the Research, Design, and Analysis Unit at WSU. His research interests include the influence of urban poverty on early childhood development, adolescent risk behaviors, developmental psychobiology, emotion regulation, early childhood literacy, longitudinal methods, complex adaptive systems, and relational developmental systems theory. Dr. Partridge also has over 20 years of experience in community-based research, public health and prevention interventions, and evaluation research.

Case study

The authors identified three ways the IO program at Wayne State University is encouraging graduate students to learn and use R,

"Introduction to R" Summer Seminar

Bi-weekly workshops covering data management in R for psychology graduate students.

Integrate R syntax and output into statistics lectures/labs

Faculty review R output as part of their statistics lectures and allow students to complete their homework assignments in R.

WAYNE STATE UNIVERSITY
MIKE ILITCH
SCHOOL OF BUSINESS

June 1–6, 2020 in Detroit, MI
Over 10 short courses are being taught in R in 2020.

Caitlynn Sendra • WSU • SAP SuccessFactors • Doctoral Candidate

Q: Tell us about your recent experience navigating the applied job market in Fall 2019. Specifically, what were employers' expectations that you knew how to use R on day one?

A: Regarding my experience with my recent job search, it has become clear to me that, if you have a desire to do anything IO related that might require analytics, learning R is more of a requirement than a recommendation. Analytics jobs that were historically staffed by IO professionals are increasingly being staffed by data scientists. Every company I spoke to and every job posting I read was *selecting* for R, not training for R. I generally agree with the choice for companies to select on this. Training R can take quite a bit of time and requires a different pattern of thinking when compared to SPSS. After all, you aren't learning statistics, you are learning statistical *programming*, a very different skill.

Throughout the entirety of my job search, I saw a total of 3 relevant job postings that used SPSS, the rest all required R/Python and SQL. Regarding companies that *did* use SPSS, the only ones I saw were companies which were founded by IO psychologists. Unfortunately, given how niche our field is, these companies are few and far between, and job seekers cannot rely on these companies having job openings at the time of their search.

Q: How would you rate your proficiency in R prior to starting your job search?

A: Prior to my job search, I did have some experience with R, having had taken an undergraduate course in it, the CARMA short course, and in using it on a variety of smaller projects. Though, I was in no way an advanced or even intermediate R-user.

**Q:** How did employers' react to your limited proficiency in R?

A: Every in-person interview would ask me about my comfort level with R, and I was always honest in saying I am still learning and that I am more proficient with SPSS. Although this never elicited and overly negative response, the fact that this question was always brought up led me to believe that it was at least partially my lack of proficiency with R that led to me being passed over for a variety of roles.

Q: How did you learn R once it became clear that it was a skill you needed to land an analytics-focused IO job?

A: When I began my job search and realized that I was not competitive with my current skillset, I personally decided to use DataCamp to sharpen my R skills. This website effectively "gamifies" the daunting task of learning R and data science, making it both fun and enjoyable.

Q: How can graduate programs better equip their students with the R skills necessary to be competitive for analytics-focused IO jobs?

A: Graduate programs are doing a disservice to their students but insisting on the use of point and click programs like SPSS. If graduate programs do not want to instruct on the use of R, a good alternative would be to subsidize the cost of services such as Skillshare, DataCamp, Pluralsight, etc. which offer introductory courses in R and other useful programming languages (e.g., Python, SQL). The unfortunate reality is that, if students are not instructed in R and in popular statistical methodologies (namely, machine learning and AI), it is difficult for them to be competitive in the job market.
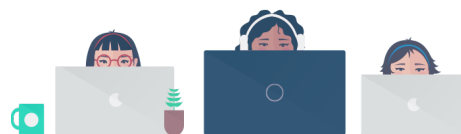
Dr. Cort Rudolph is an Associate Professor of IO Psychology at Saint Louis University. He received a B.A. from DePaul University, and a MA and PhD from WSU. Cort's research focuses on a variety of issues related to the aging workforce, including applications of lifespan development theories, wellbeing and work-longevity, and ageism/generationalism. He is an associate editor of the Journal of Vocational Behavior.

Caitlynn Sendra received her MA degree in IO Psychology in April of 2019 and is currently working towards her doctorate degree. She enjoys learning how to apply statistical methodologies to solve organizational problems and has been teaching herself R throughout her graduate career. She was on the job market from October to December of 2019, and is now a Human Capital Management Research Consultant at SAP SuccessFactors.