INTERSECT
RESEARCH FASTER

Intersect Australia

Research Digital Skills Training Program

2022 Report

Authors

Dr Anastasios Papaioannou[1], Dr Katherine Howard[1,2], Dr Jianzhou Zhao[1], Glen Charlton[1], Charlotte François[1,3], and Dr Jonathan Arthur[1]

[1] *Intersect Australia, Sydney, Australia*
[2] *University of South Australia, Adelaide, Australia*
[3] *University of Adelaide, Adelaide, Australia*

Keywords

data literacy, digital skills, education, training, researcher training, digital tools, digital technologies, data science, data management, surveys, machine learning, statistics, databases, high performance computing

Licence

Copyright

Table of Contents

1. Executive summary

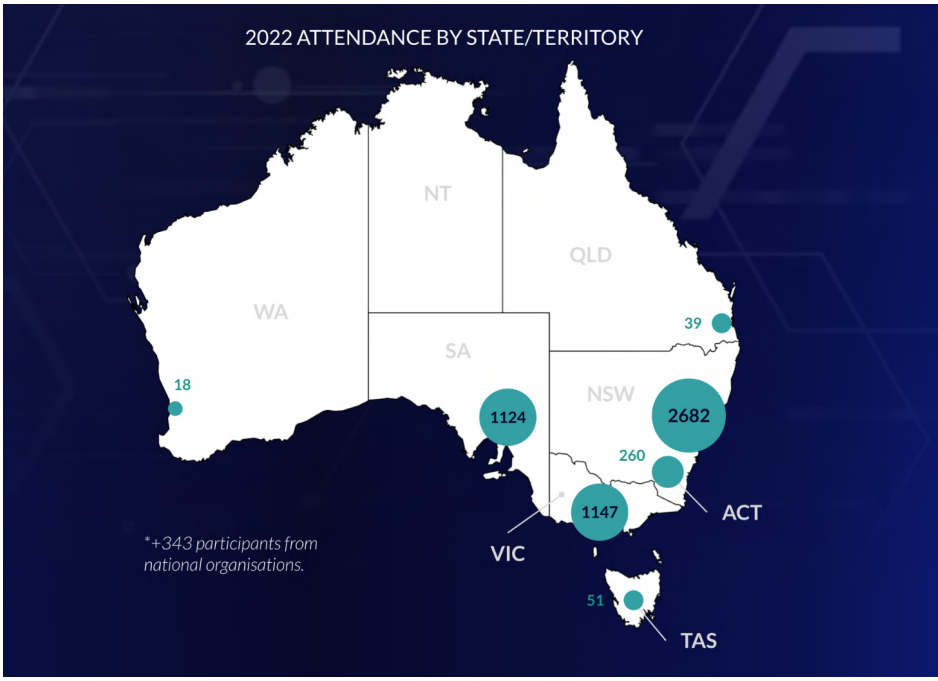
Intersect's mission is to help researchers to be more efficient and effective in their research; reducing the time to move from an idea to a tested solution. As a leading provider of digital skills training for researchers in the Australasian region, Intersect provides an extensive range of technology focused training to researchers and higher degree research (HDR) students across Australia. This training ranges from awareness to advanced levels; is delivered interactively either face-to-face or online; and covers categories such as Research Computing, Programming, Data Science, Data Analytics, Machine Learning (ML) & Artificial Intelligence (AI), Statistics, Data Visualisation, Data Collection, and Data Management. Intersect continually revises, updates, and expands its course catalogue, ensuring researchers always have access to the most relevant and useful research training.

Our hands-on, instructor-led, live, interactive training is delivered by over 25 highly experienced instructors and is targeted at enhancing the capabilities of researchers in digital tools and technologies. Our research and training expertise extends across various disciplines including, but not limited to: ICT, Data Science, Linguistics, Engineering, Statistics, Bioinformatics, Health & Medical Sciences, Materials Science, Sports Science, Spatial Analytics, Computational Chemistry, Numerical Modelling, Behavioural Science, and Social and Political Sciences.

In 2022, Intersect has trained a total number of 5,664 researchers in 300 courses (229 training days). The average attendance rate was approximately 76% based on the capacity (or registration number) of the courses. From those attendees, 4,419 or approximately 78% attended the courses Intersect scheduled at their local university while 1,245 or about 22% attended Intersect courses through the consortium of Intersect member universities.

Intersect members are located in NSW, VIC, SA, and ACT, where the majority of the training program takes place. However, the Intersect training courses are also attended by HDR students and researchers from other states and territories, such as Western Australia (WA), Tasmania (TAS), and Queensland (QLD), as well as in national organisations mainly due to the NCI-Intersect Training Partnership. The NCI-Intersect training partnership includes running a series of training courses aimed at users who may be new to NCI's high-performance computing environment, as well as experienced users looking for a refresher.

In 2022, most attendees (2,682) are from New South Wales (NSW) universities, organisations, and government departments constituting about 47% of the total attendees. 1,147 and 1,124 HDR students and researchers from Victoria (VIC) and South Australia (SA), respectively, attended Intersect training courses in 2022 representing 20.7% and 20.4% of the total attendees. The fourth largest cohort was from national organisations with approximately 350 attendees, followed by ACT with 260 attendees. Smaller numbers of attendees were from other states and territories .



To evaluate the quality of the training delivery, Intersect asks attendees to complete a course evaluation survey at the end of each course[1], where a Net-Promoter Score (NPS)[2] is also measured. Intersect's NPS during this period was +70 based on 2,090 responses (~37% of attendees) which is considered outstanding. Furthermore, the average scores of the primary metrics for measuring the quality of the training delivery exceed 9.1 out of 10, which indicates that feedback from participants is excellent.

This report summarises the Research Digital Skills Training Program provided by Intersect Australia across Australia for the calendar year 2022 as well as useful insights on historical trends.

[1] In our course evaluation survey, a scale of 0 (worst) to 10 (best) is used.
[2] The Net Promoter Score (NPS) has been widely adopted and is a metric that captures satisfaction (0-10 scale). In the NPS question, attendees are asked how likely they are to recommend Intersect Training to others. NPS can have a value between -100 (lowest) and +100 (highest). A positive NPS is considered great, while achieving an NPS of +50 or higher is considered outstanding and seldom achieved commercially.

2. Attendance overview

The vast majority of Intersect training courses are scheduled between February and November as shown in Figure 2.1, with only a few being scheduled in early December, and even less in January due to the holiday period. The two busiest months were March and August during which 39 courses were delivered and around 740 attendees were trained, respectively. Interestingly, there is a drop in the courses delivered in July due to the semester break.

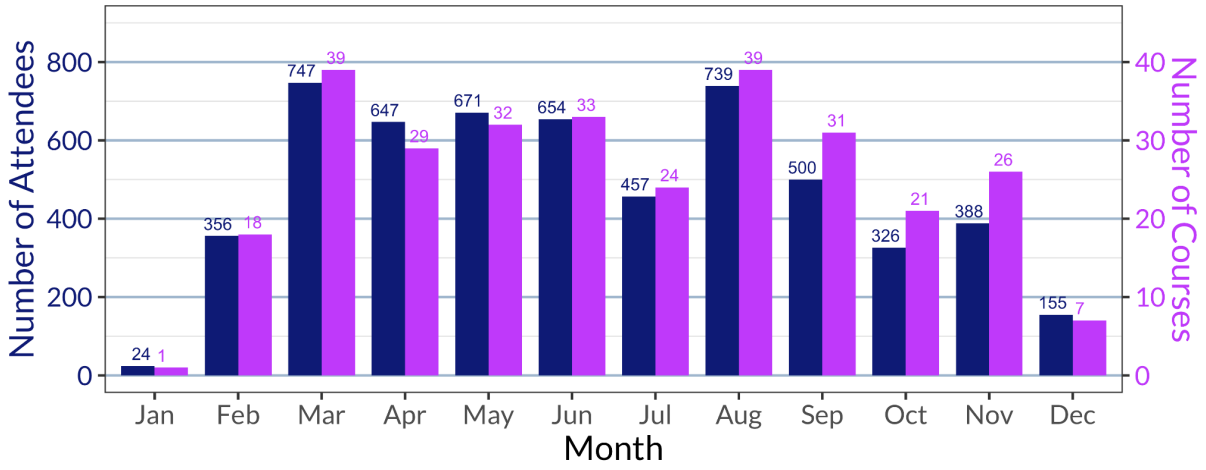


Figure 2.1: Number of attendees trained (left y axis) and number of courses delivered (right y axis) by month as part of the Intersect Research Digital Skills Training Program in 2022.

Figure 2.2 shows the distribution of courses delivered and number of attendees trained by the day of the week. Most Intersect courses were scheduled on Tuesday, followed by Wednesday and Thursday. A smaller number of courses was scheduled on Friday, and only a few courses were delivered on Monday, primarily due to internal and other commitments for the Intersect Digital Research Analysts, who are responsible for the scheduling of courses.

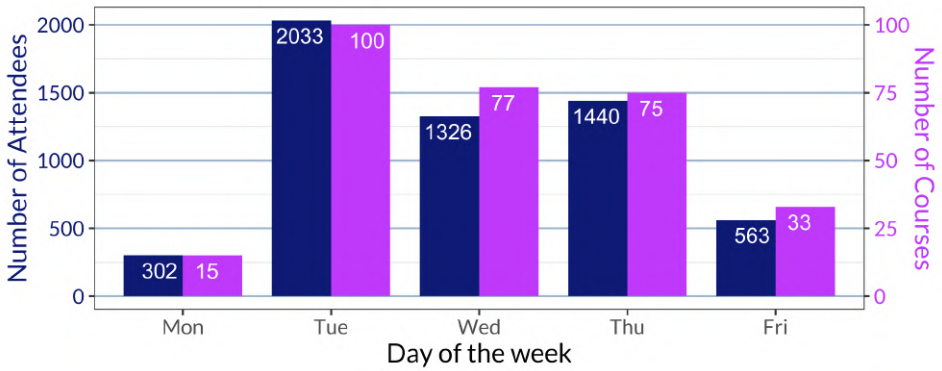


Figure 2.2: Number of attendees trained (left y axis) and number of courses delivered (right y axis) by day of the week as part of the Intersect Research Digital Skills Training Program in 2022.

There are two different types of courses that Intersect delivers: a) courses that are delivered by the Digital Research Analyst solo (usually with less than 15 attendees); b) courses that are delivered by 3 trainers (one lead instructor and two assistant trainers, usually with 15+ attendees). Intersect's trainer:trainee ratio is between

1:10 (one trainer per 10 attendees) and 1:15 (one trainer per 15 attendees). Therefore, the class size varies from as few as 8 attendees to 30+ attendees per course based on the type of the courses and the attendance rate. The distribution of courses delivered in 2022 by class size is shown in Figure 2.3.

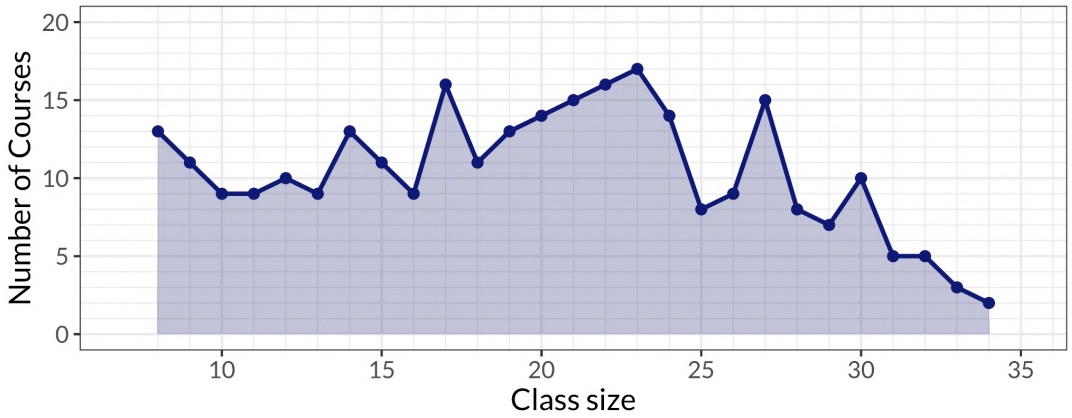


Figure 2.3: Distribution of number of courses by class size (number of people attended the course) for the Intersect Research Digital Skills Training Program run in 2022.

2.1 Attendance by Tool/Technology

Figure 2.1.1 shows the percentage of attendees that were trained in 2022 by Tool/Technology. The tools and technologies that are beyond the top ten with the highest attendance were aggregated and presented as "Other tool". R and Python programming courses were the most popular with more than 50% of attendees participating in these courses. NVivo, Excel, REDCap, Qualtrics, and SPSS courses were the next most popular courses with between 4% and 10% of attendees each. The courses were spread across the year and the majority of the courses were delivered in Quarter 2 and Quarter 3. The most popular course in Quarter 1 to 3 was R, followed by Python; this reversed in Quarter 4 with more attendees participating in Python courses. After introducing NVivo courses in 2022, they have become very popular (third most popular technology in 2022) especially among HDR students and researchers from Humanities, Arts and Social Sciences (HASS).

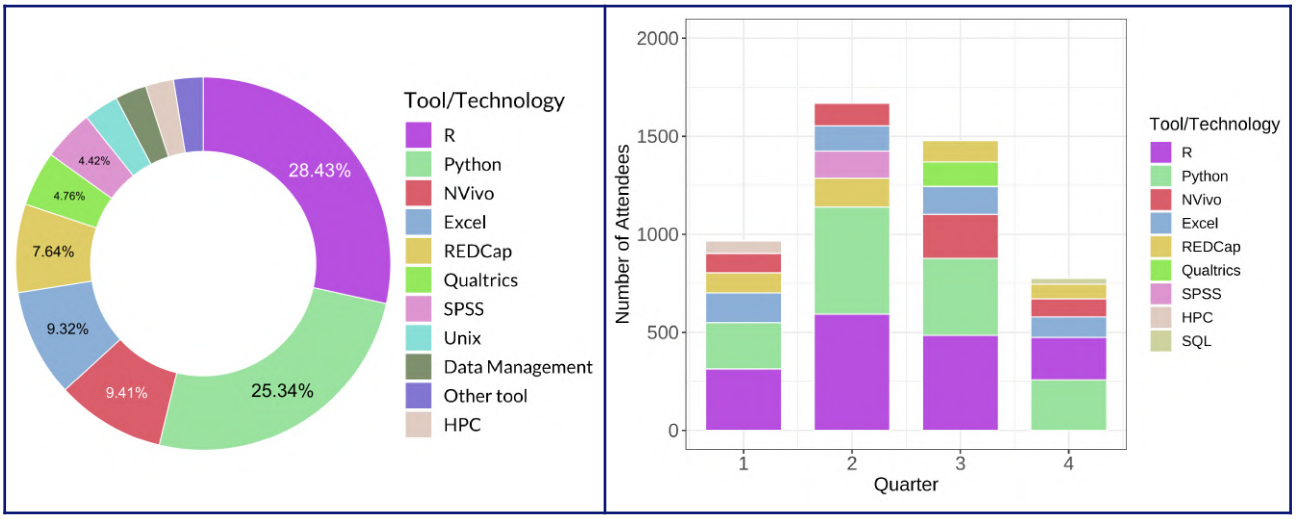


Figure 2.1.1: Left: Percentage of attendees by Tool/Technology; Right: Quarterly number of attendees by Tool/Technology (only top six tools/technologies per quarter are presented).

2.2 Attendance by Role/Position

Figure 2.2.1 (left) shows that Higher Degree Research (HDR) students (PhD) are the top consumer of Intersect's training program in 2022 making up 54% of the total attendees. The second highest consumer is Academics comprising approximately 14% of attendees followed by Professional (research-related) and Post-doc/Fellow with approximately 9% and 8%, respectively. PhD students were consistently the top consumer across all four quarters.

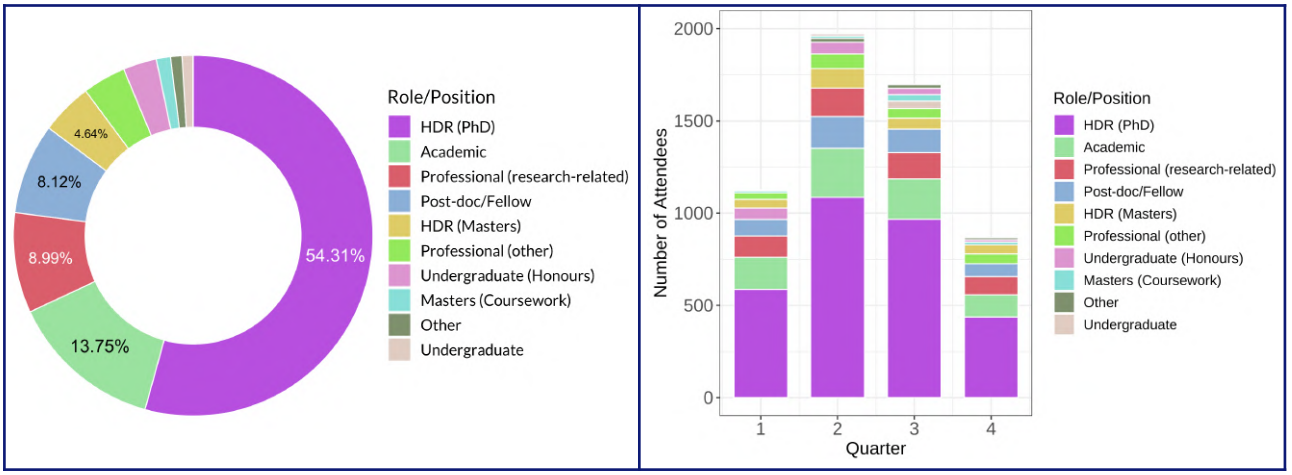


Figure 2.2.1: Left: Percentage of attendees by Role/Position; Right: Quarterly number of attendees by Role/Position.

Figure 2.2.2 shows the percentage of attendance by tool and technology and then Role. Interestingly, the distributions vary when comparing different roles with different tools/technologies. PhD students are the top consumers for all tools and technologies, except for SQL (Structured Query Language). Moreover, although PhD students are still the top consumer of the surveying tool courses (REDCap and Qualtrics), a much lower percentage in PhD students and a higher uptake of these courses in other roles such as Academic, Professional (research-related), and Undergraduate is observed. The highest percentage of attendance for Academics is observed in NVivo, Qualtrics, and REDCap courses, while for Post-doc/Fellow it is in Julia, HPC, and Unix. Regarding Professionals (research-related), the highest uptake is shown to be in SQL, REDCap, Julia, Git, and HPC.

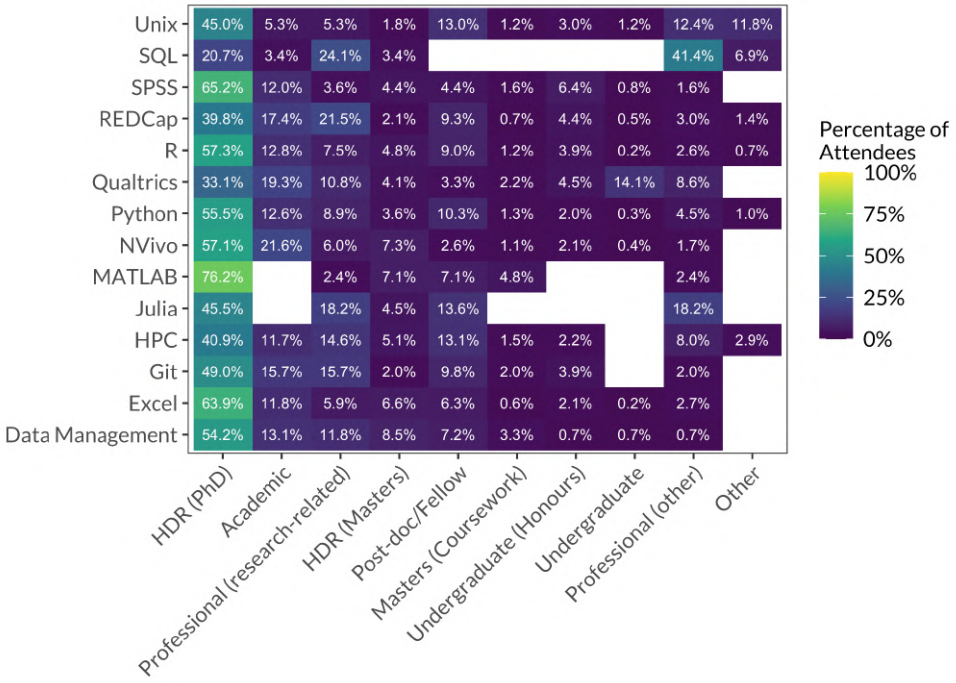


Figure 2.2.2: Percentage of attendees by tool/technology that is further split by Role.

2.3 Attendance by Faculty

Intersect consistently captures the Field of Research (FoR) code during course registration. As there is no controlled vocabulary or naming convention for university faculties, the FoR codes are mapped to generic faculty names to estimate attendance by faculty.

Approximately a third (~33%) of the training attendees are from the Faculty of Medicine and Health followed closely by the Faculty of Science, with 28% of the attendees coming from this faculty. In 2022, the Faculty of Engineering was the third biggest consumer surpassing the Faculty of Arts and Social Sciences which had been the third highest consumer for the previous two consecutive years, 2020 and 2021.

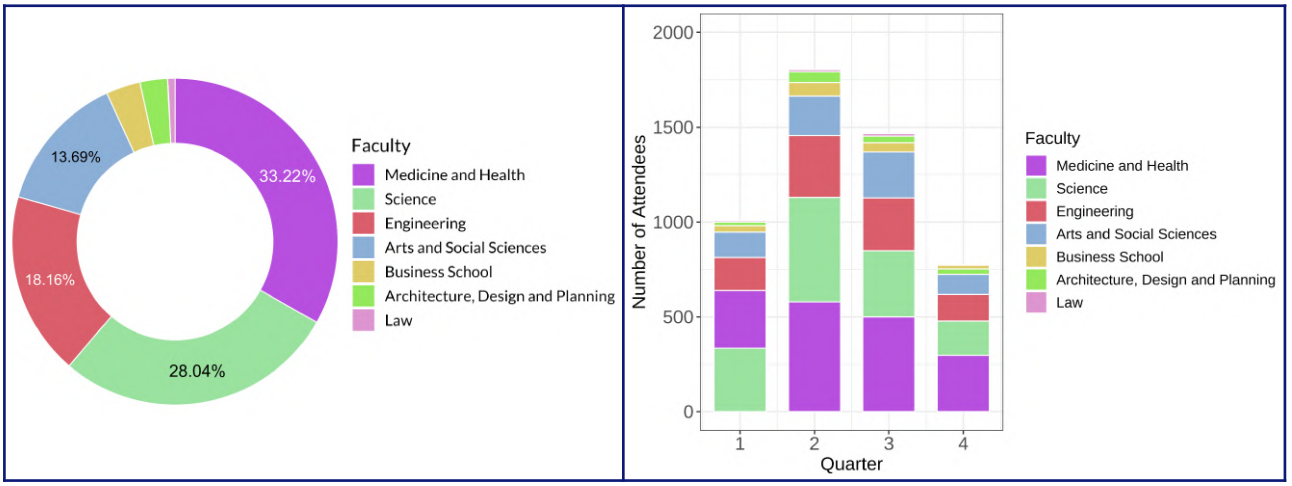


Figure 2.3.1: Left: Percentage of attendees by Faculty; Right: Quarterly number of attendees by Faculty.

Figure 2.3.2 shows the percentage of attendance by tool and technology and then by faculty. Regarding the programming courses, Python shows a different distribution compared to R with the vast majority of attendees being from the Faculty of Engineering and Faculty of Science while in R they are mostly from the Faculty of Medicine and Health and the Faculty of Science. The highest uptake for the MATLAB and Julia programming courses was observed in the Faculty of Engineering and Faculty of Science, respectively. As expected, the Faculty of Arts and Social Sciences is the largest consumer of the survey tool Qualtrics and the qualitative analysis tool NVivo. Attendees from the Faculty of Arts and Social Sciences have also shown an increased interest in attending SPSS courses. However only a small uptake is observed in the most popular programming courses such as Python and R, indicating that more work is needed to increase awareness on how programming and other digital tools can be integrated into HASS research. As for the other survey tool, REDCap, 75% of the attendees are from the Faculty of Medicine and Health.

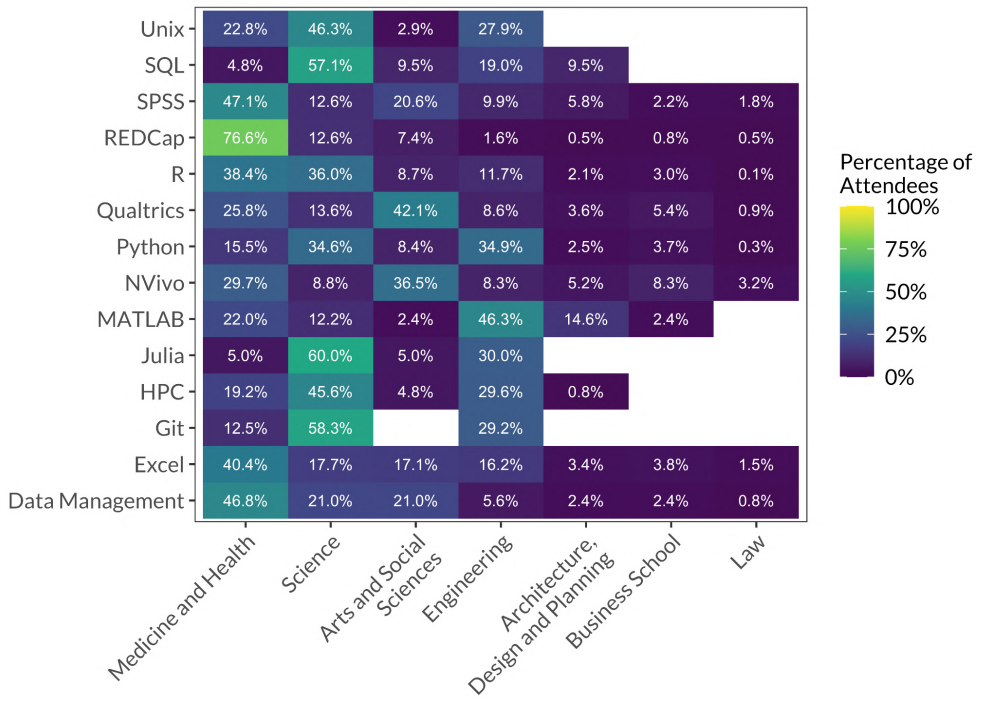


Figure 2.3.2: Percentage of attendees by tool/technology that is further split by Faculty.

Figure 2.3.3 shows the sankey diagram of the flow from Role/Position (top five) to Faculty and then to Tool/Technology (top ten). HDR (PhD) students is consistently the largest cohort among all faculties apart from the Business School where Academics are the largest cohort constituting about 54% of the total attendees in the school. Compared to the average percentages (see Figure 2.2.1 left), a higher number of Academics from the Faculty of Law, the Faculty of Arts and Social Sciences, and the Faculty of Architecture, Design, and Planning have attended Intersect courses with a value between 20% and 39% of the total attendees of each faculty. The highest percentages of Post-doc/Fellows and Professional (research-related) is seen in the Faculty of Science and the Faculty of Medicine and Health. As we mentioned, the largest consumer of Intersect training courses is the Faculty of Medicine and Health with about 33% of total attendees (see Figure 2.3.1 left). Participants from the Faculty of Medicine and Health attend mostly R courses (~34%), followed by REDCap (17.5%), Python (12.2%), and Excel (11.6%). HDR students and researchers from the Faculty of Science are predominantly interested in learning R and Python with a combined percentage of more than 70% of total attendees from this faculty attending R and Python courses. Notably, over 50% of attendees from the Faculty of Engineering participate in Python courses, and to a lesser extent R courses (~20%). Interestingly, the Faculty

of Arts and Social Sciences shows a more diverse interest in learning different tools and technologies, with the top five tools being from highest to lowest (24% - 11%), NVivo, R, Python, Qualtrics, and Excel.

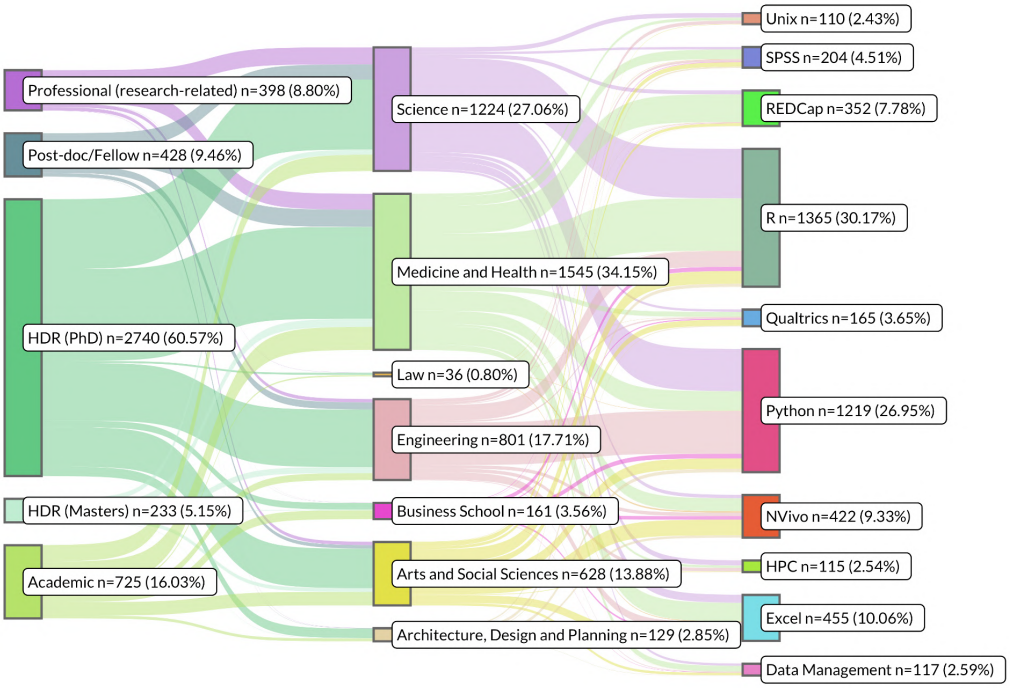


Figure 2.3.3: Sankey diagram depicting the flow from the top five Roles/Positions of attendees to the different Faculties and then to the ten most popular Tools/Technologies.

2.4 Attendance by Field of Research (FoR) code

Figure 2.4.1 shows the distribution of attendees based on the Field of Research (FoR) code. Note that the FoR codes that are not in the top ten highest number of attendees were aggregated and presented as "Other FoR Code" in Figure 2.4.1. Health Sciences is on top of the list with 21% of the attendees, while Engineering is the second FoR code with about 12% of the attendees. All the other FoR codes show a percentage between 4% and 7.5%.

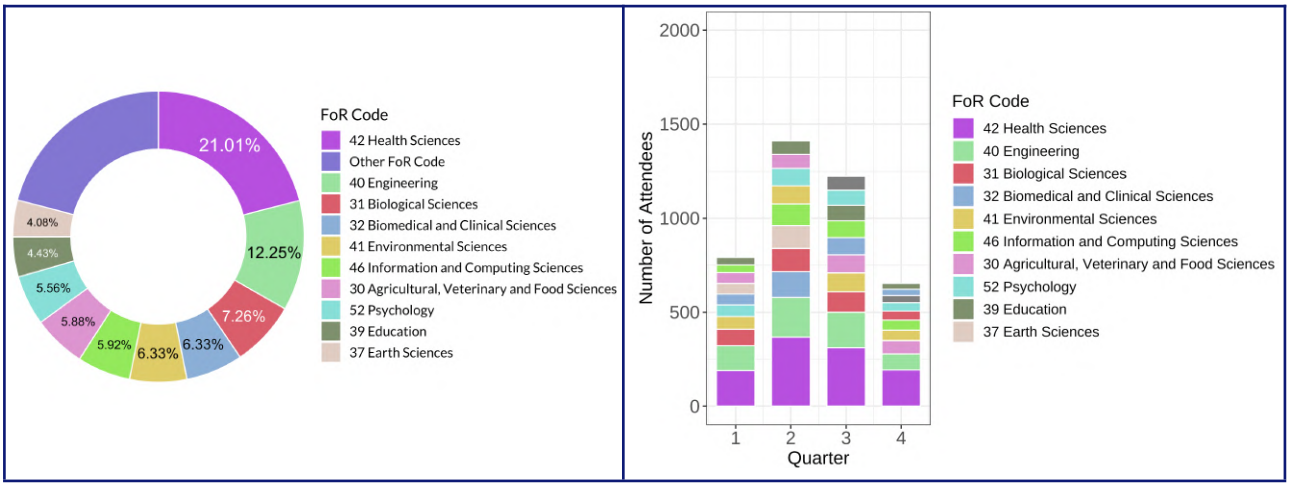


Figure 2.4.1: Left: Percentage of attendees by FoR Code; Right: Quarterly number of attendees by FoR Code.

Figure 2.4.2 shows the number of attendees by tool and technology then by FoR code. The comparison of distributions for the two most popular tools, R and Python, shows some interesting insights. The highest uptake of R courses is observed in Health Sciences, Biological Sciences, Biomedical and Clinical Sciences, Environmental Sciences, and Agricultural, Veterinary and Food Sciences. In the case of Python courses, the largest consumers are Engineering, Information and Computing Sciences, Earth Sciences, Health Sciences and Biological Sciences. It becomes apparent that researchers and HDR students from Medical and Health sciences as well Biology and Biomedical sciences prefer to learn R whereas Python is the preferred programming language for Engineering and IT researchers and HDR students. Regarding NVivo, Health Sciences is the top consumer followed by Education, Human Society and Commerce, Management, Tourism and Services. Regarding REDCap courses, the top consumer by a great extent is Health Sciences while Qualtrics courses are mostly attended by Health Sciences, Economics, Education and Psychology.

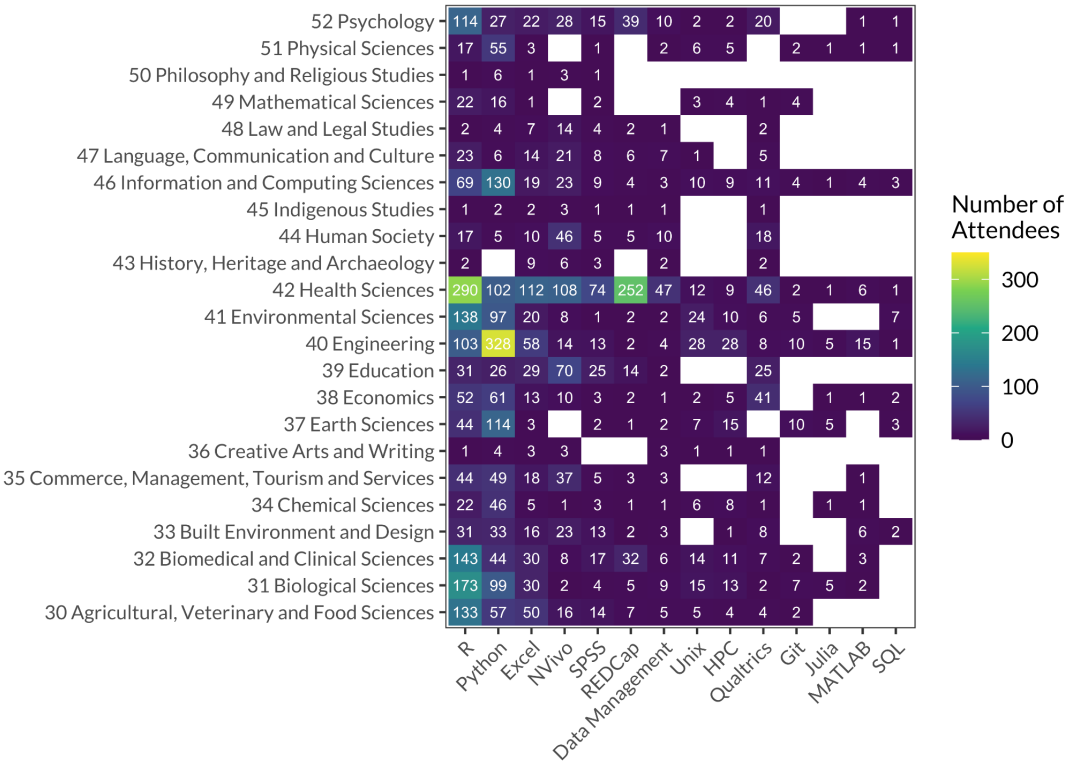


Figure 2.4.2: Number of attendees by tool/technology that is further split by FoR code.

2.5 Return to attend more Intersect courses

Figure 2.5.1 shows the distribution of the number of courses that the same participant has attended in 2022. Around 66% of the participants attended an Intersect training course once in 2022, which means one out of three participants returned to attend another Intersect training course. Nearly 19% have attended two courses, followed by approximately 8% who have attended three, and 3% have attended four courses. Approximately 4.5% of the total number of attendees have participated in the training program 5 times or more in courses delivered in 2022. Although we present only 2022 data in Figure 2.5.1, please note that it is also very common for the same person to attend various Intersect courses across several consecutive years (see Section 5).

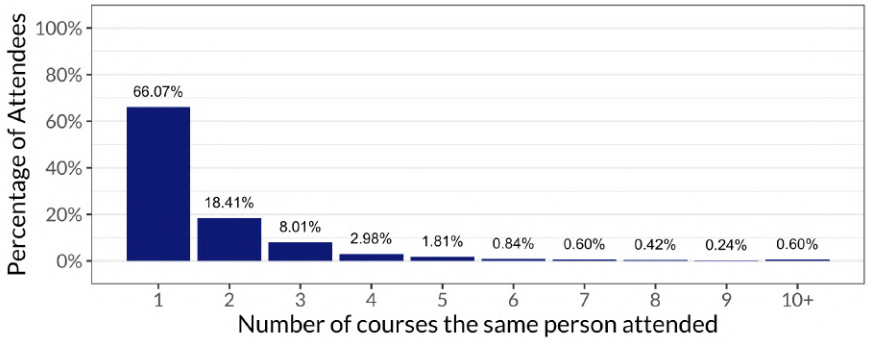


Figure 2.5.1: Number of courses that the same person attended in 2022.

2.6 Reason for attending

When participants register, they are asked to provide reasons for attending the course. Most participants answered that they are interested in these courses to learn skills that they can either apply to their work now (approx. 49%) or in the near future (approx. 31%). Approximately 12% of the attendees are learning these skills for better opportunities and employability in the future ("To learn skills that will help me get a job").

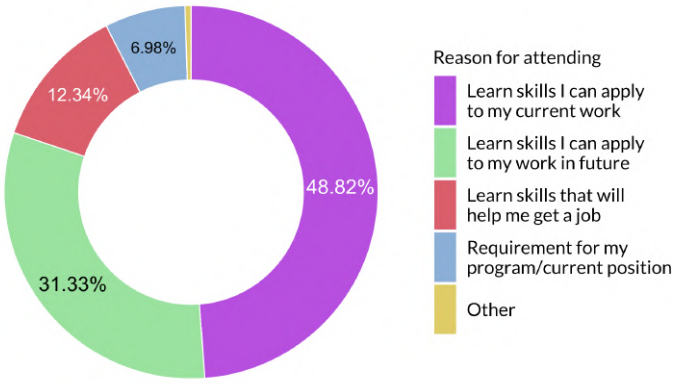


Figure 2.6.1: Reason for being interested in attending this training course.

Analysis on the reason for being interested in attending this training course by the Tool/Technology reveals a similar trend across all tools/technologies. The highest percentage for learning skills that can be applied to their work now is observed in Git, NVivo, REDCap, Qualtrics, SPSS, Excel and HPC. Regarding learning skills that can help them get a job, all programming courses (R, Python, MATLAB, Julia) show a high percentage compared to other tools/technologies. Another interesting case is REDCap, where the third most popular reason for learning this tool is because it is a requirement for their current position.

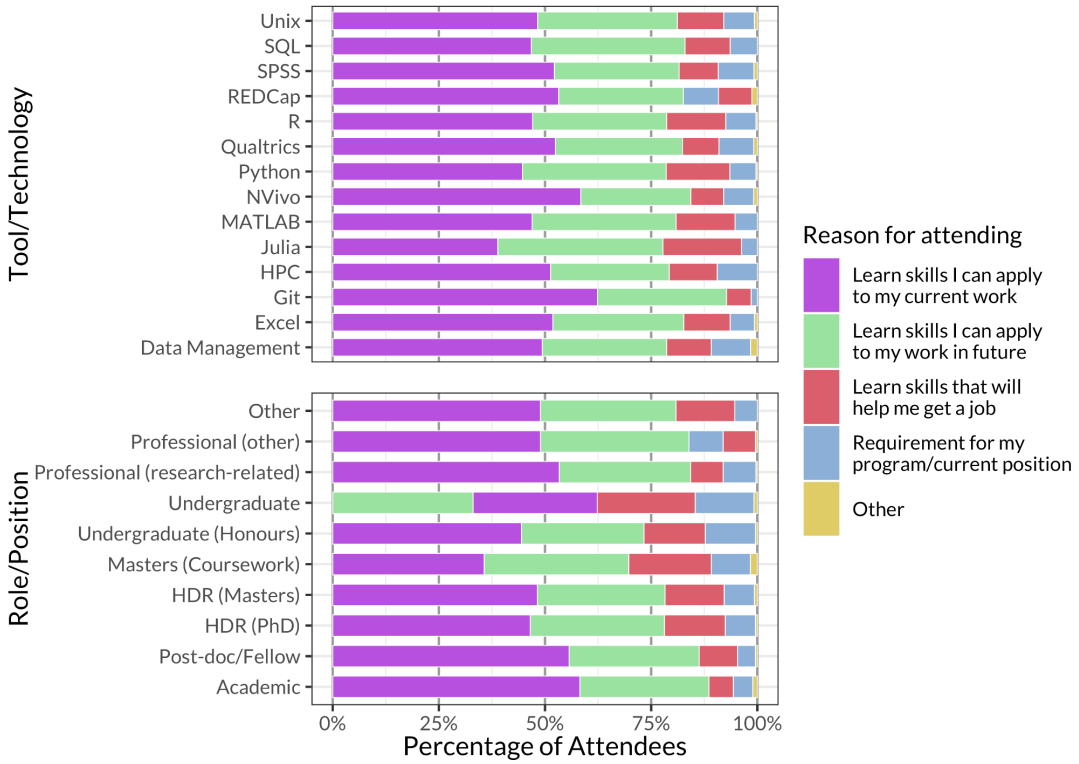


Figure 2.6.2: Top: Reason for being interested in attending Intersect training course by Tool/Technology; Bottom: Reason for being interested in attending Intersect training course by Role/Position.

Further analysis by Role/Position showed that the percentage of participants who learn skills for current and future work is higher among Academics and Post-docs compared to HDR Masters and PhD students where the percentage of those learning these skills for better opportunities and as a requirement for their program is higher. Interestingly, the small number of Undergraduate students who participated in Intersect courses attend the training mostly to learn skills that can be useful in the future. Undergraduate students together with Masters (Coursework) show the highest percentage for learning these skills to help them get a job.

Figure 2.6.3 shows the sankey diagram of the flow from Role/Position (top five) to the reason for attending the Intersect course and then to Tool/Technology (top ten). All top five roles consistently answered that the top reason for attending an Intersect training course is to learn skills that they can apply to their current work with a percentage that fluctuates between approximately 46% (HDR PhD and HDR Masters) and 58% (Academics). Attending the courses to learn skills that can help in their work in the future was consistently the second top reason among all roles with 30%-31% of responses. Interestingly, 14.3% and 13.9% of HDR PhD and Masters students responded that the reason for attending these training courses is also to gain skills that can help them get a job, and therefore they attend the courses to improve their employability. The percentage for this reason for Post-docs was 9% followed by Professional staff (research-related) and Academic with 7.7% and 5.7%, respectively. Analysing the reason for attending a course by Tool/Technology, a similar trend and percentages are observed for the top two reasons, *e.g.*, to learn skills they can apply to their current work now or in the future. However, Python and R courses stood out when analysing the response "Learn skills that will help me get a job" by tool with 15% and 13.9% of registrants mentioning that they want to learn Python and R to help get better job opportunities in the future. This also demonstrates how popular skills Python and R programming languages are when it comes to employability, and in particular among HDR students who are looking for opportunities to upskill themselves in emerging tools and technologies.

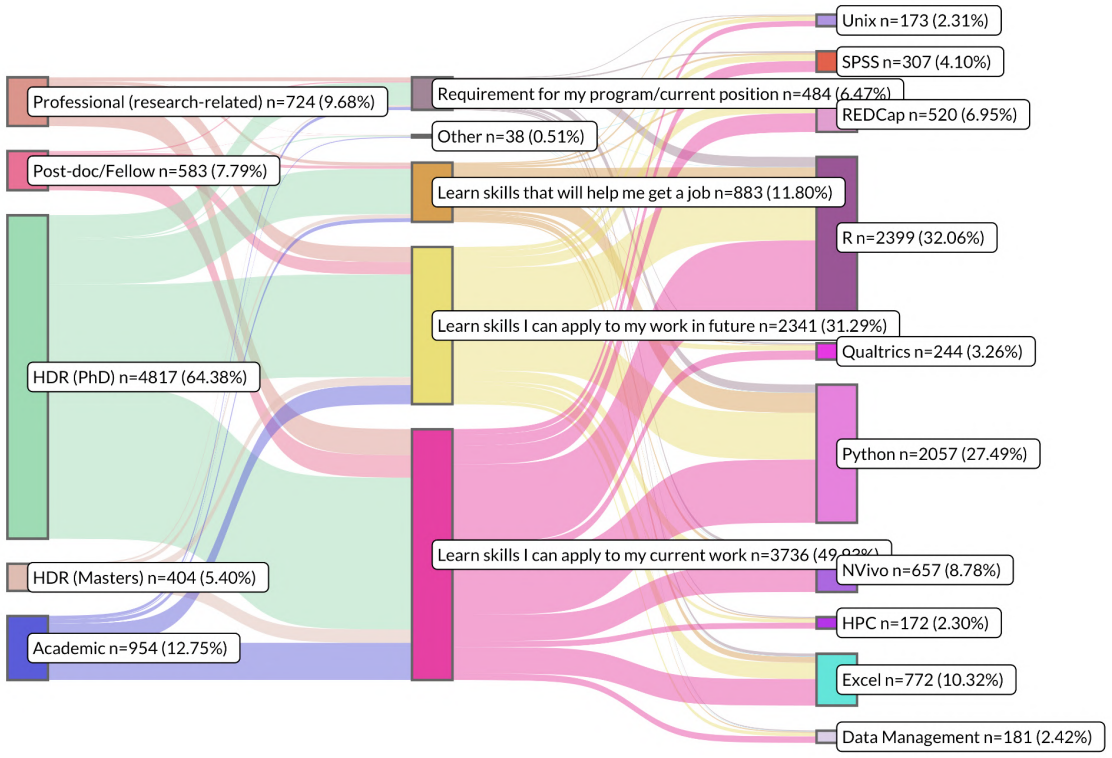


Figure 2.6.3: Sankey diagram depicting the flow from the top five Roles/Positions of registrants to the different reasons for attending a course and then to the ten most popular Tools/Technologies.

2.7 No shows

"No shows", *i.e.*, people who register for a course and then fail to attend, are a common issue faced by Intersect and other similar training providers. Intersect closely monitors the percentage of no shows and has applied various methods, in different use cases, to tackle the issue and minimise the impact.

Figure 2.7.1. shows the average percentage of no shows by month and day of the week. The highest average percentage of no shows was in June, July, and October during which more than 28% of registrants did not attend the training course without notifying the Intersect training team. The months with the lowest percentage of no shows are January and May, however Intersect only delivered one training course in January and therefore the sample is not adequate to draw any conclusions. April and December also show a lower percentage of no shows compared to other months of the year. Analysing the percentage of no shows by the day of the week on which the courses run, we notice that the no shows rate is similar across the days, fluctuating between 21% and 26%.

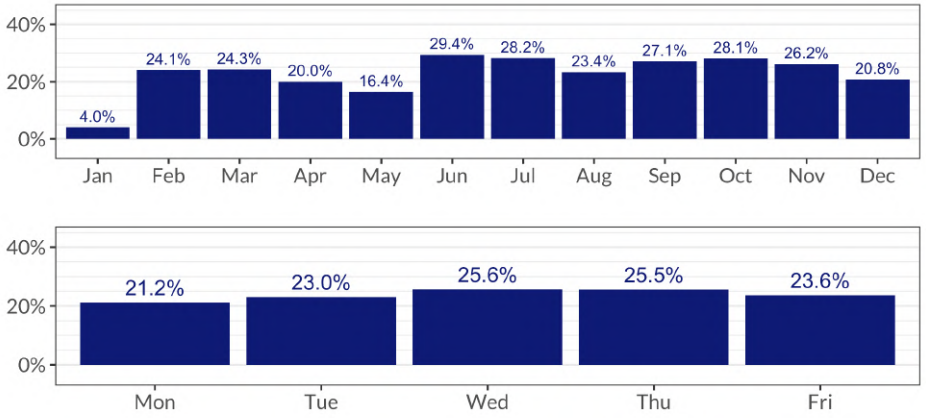


Figure 2.7.1: Top: Average percentage of no shows by the month the course was delivered; Bottom: Average percentage of no shows by the day of the week the course was delivered.

Some more useful insights can be derived by plotting the distribution of no shows by month and day of the week as shown in Figure 2.7.2. Monday and Tuesday are the days with the lowest median percentage of no shows, while Friday and Thursday are the highest ones. When comparing the distributions by month, the first 5 months of the year display lower median percentages of no shows compared to months after May. May is shown to have the lowest median percentage of no shows while October and June have the highest one.

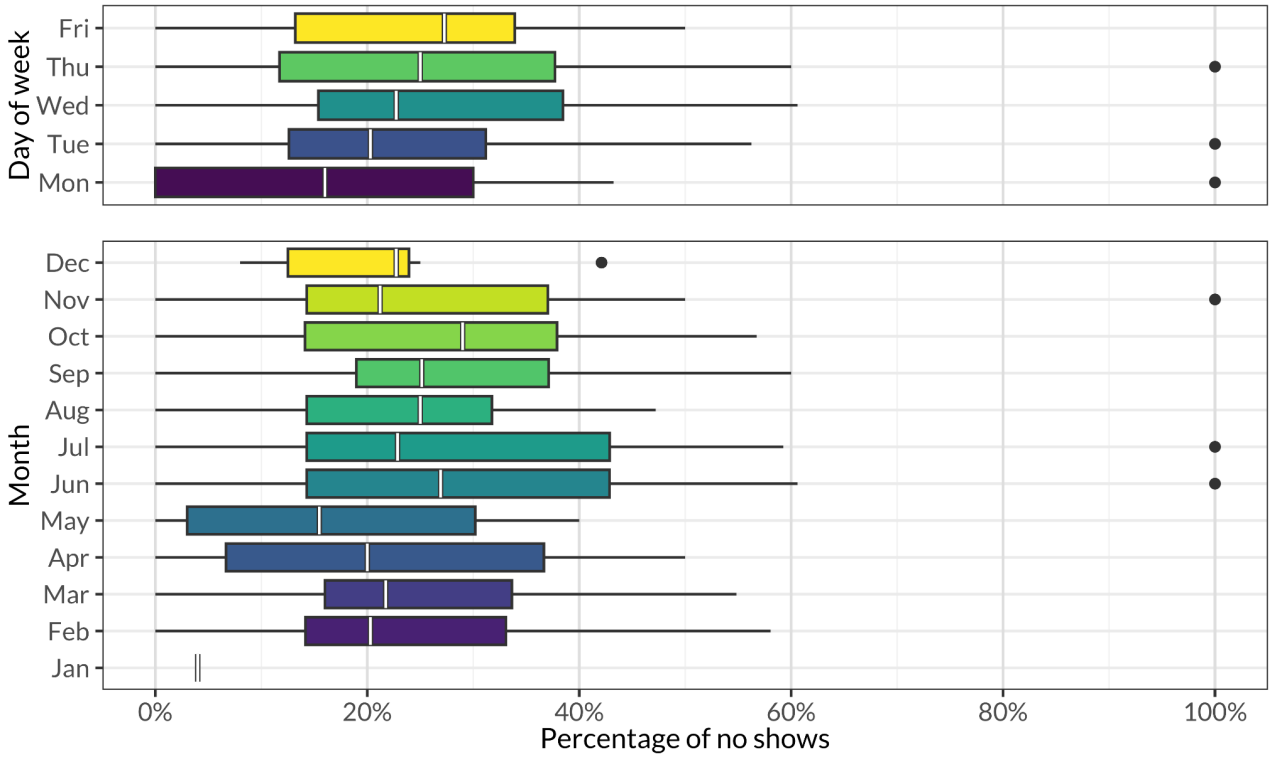


Figure 2.7.2: Top: Distribution of no shows by the day of the week the course was delivered; Bottom: Distribution of no shows by the month the course was delivered.

Further analysing the data about no shows by Role/Position and Faculty, we observed that the average highest no show rate is among Masters (Coursework) students with 36.8% of registrants failing to attend, followed by HDR (PhD), Post-doc/Fellow, and HDR (Masters) with approximately 29% no shows rate each. The lowest

average percentage is shown in Other and Professional staff. Participants from the Business School and Faculty of Engineering have the highest average percentage of no shows with approximately 32%, respectively, while participants from the Faculty of Law and the Faculty of Science display the lowest average percentage of no shows.

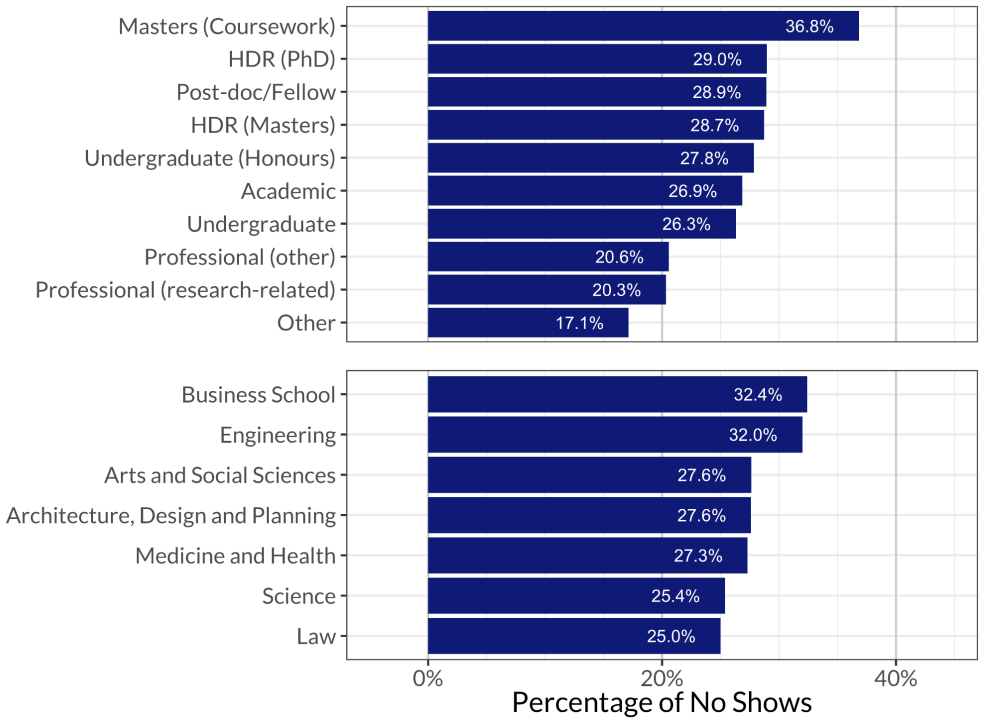


Figure 2.7.4: Top: Average percentage of no shows by Role/Position; Bottom: Average percentage of no shows by Faculty.

3. Evaluation

To evaluate the quality of the training delivery, Intersect instructors ask the attendees to fill in a course evaluation survey at the end of each course. In our course evaluation survey, a scale of 0 (worst) to 10 (best) is used. In addition, a Net Promoter Score, in which attendees are asked how likely they are to recommend Intersect Training courses to others, is also measured. A positive NPS is considered great, while achieving an NPS of +50 or higher is considered outstanding and seldom achieved commercially. The Intersect training program demonstrates high quality and is well received by the attendees. We have analysed all the survey responses in 2022 and the analysis is presented in the following subsections.

3.1 Evaluation by Tool/Technology

The Intersect training program demonstrates high quality and is well received by the attendees. In 2022, Intersect scored an NPS of +70 and all courses by tool/technology have independently achieved an average NPS of over +53, which is considered outstanding. This demonstrates that while we deliver digital skills training for researchers at scale, we still maintain, and where possible improve, the quality of delivery to provide the best possible learning experience to attendees.

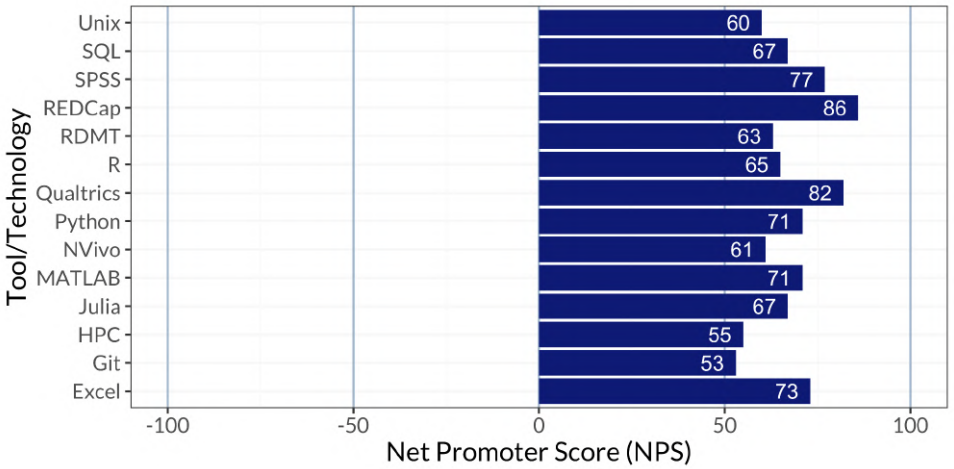


Figure 3.1.1: Average Net Promoter Score (NPS) by Tool/Technology. The course data was aggregated by Tool/Technology.

3.2 Evaluation by month and class size

The NPS score was further calculated by month (Figure 3.2.1) and class size (Figure 3.2.2). A very high NPS score is observed in all months, with all months consistently achieving an average NPS of more than 50.

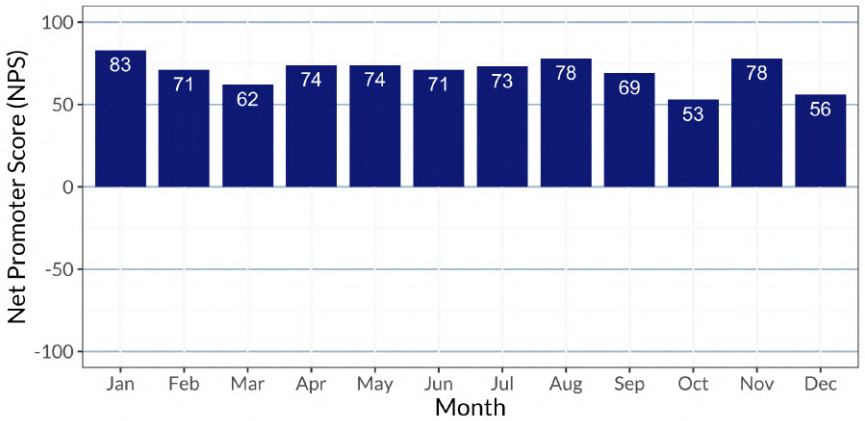


Figure 3.2.1: Average Net Promoter Score (NPS) by month.

When analysing the survey by class size, *e.g.*, the number of participants in a course, although there is a slight fluctuation on the NPS score, we notice a consistently high NPS of 50+ among all class sizes. It is worth noting that even classes in 2022 with more than 30 attendees achieved a consistently high NPS score (+50 and above), indicating that the quality of the course delivery does not drop in courses with higher attendance. It is important to note that analysis on historical data (based on over 10,000 responses) showed that there is a slight decrease in the NPS as we increase the number of participants in the training courses, and therefore the number of attendees per course should be carefully considered.

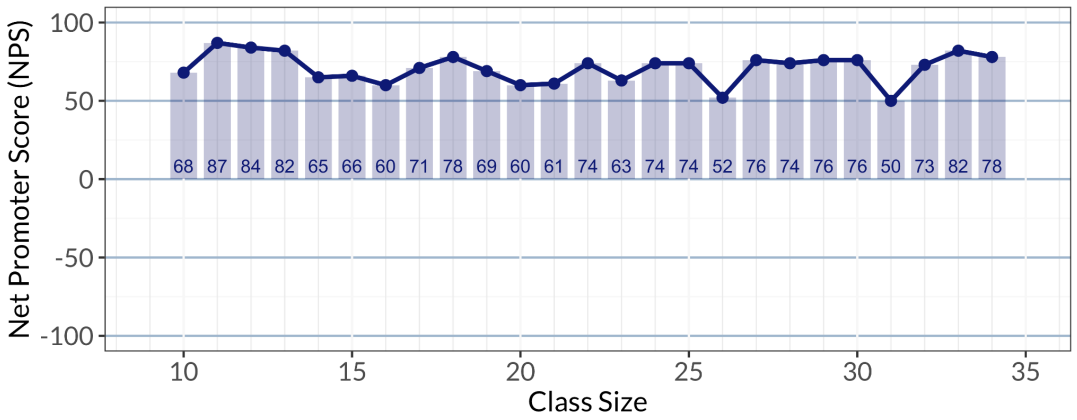


Figure 3.2.2: Average Net Promoter Score (NPS) by class size.

3.3 Metrics for measuring quality of delivery

Figure 3.3.1 shows the five primary metrics for measuring the quality of the training delivery per quarter, namely:

- **Clear answers**: "*Do you think that the instructors gave clear answers to questions?*"
- **Comfortable interacting**: "*Did you feel comfortable interacting with the instructors?*"
- **Good communicators**: "*Do you think that the instructors were good communicators?*"
- **Knowledgeable instructors**: "*Do you think that the instructors were knowledgable about the material being taught?*"
- **Welcoming atmosphere**: "*Did you feel that the training course atmosphere was welcoming?*"

All metrics are above 9.0 out of 10 in each quarter, which indicates that the training attendees appreciate the comfortable training environment, interactive teaching style, and excellent communication from knowledgeable instructors.

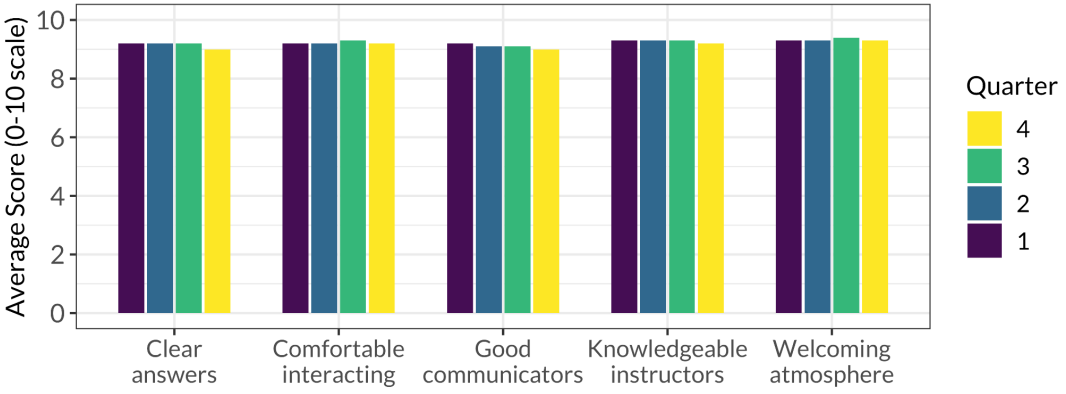


Figure 3.3.1: Primary metrics for measuring the quality of the training delivery per quarter.

3.4 Qualitative feedback and evaluation

Qualitative data collected from participants consisted of responses to three open-ended questions:

1. "*Which parts of the course did you find most useful? Why?*"
2. "*Which parts of the course did you find least useful? Why?*"
3. "*Do you have any other suggestions or feedback on this course or Intersect Training in general?*"

A high-level thematic analysis was undertaken on a sample of the 2,090 participants who responded to the Intersect course evaluation survey in 2022. Words and phrases representative of the emergent themes were then analysed in R to generate counts and quantify themes using the full data set. There were 1,500 comments for question one, 1,204 for question two, and 964 for question three. While this type of inductive analysis introduces a level of bias, it was an appropriate approach for an initial analysis. More detailed analysis on the full data set is continuing and will be presented in future reports.

Themes common to the first and the third questions included positive feedback on the style of delivery with 367 related comments. Comments relevant to this theme indicated that the course content was explained well by "really patient and helpful" trainers (Introduction to Machine Learning using Python: Classification participant #1) and in a "non-intimidating way" (Learn to Program: Python participant #1). Comments relating to the training being well organised and well run; the supportive, non-judgemental environment created; and the overall expertise and professionalism of the trainers were also common.

Comments that were categorised under the broader theme of the structure of delivery (281 comments) were again common to both the first and third questions. Participants appreciated the "fully interactive" nature of the training (Surveying with Qualtrics, participant #1), highlighting that this aspect is "what makes this course very useful and different to the readily available materials or content on the internet." (Data Manipulation and Visualisation in Python participant #1). For others, the "demonstration of steps and being able to follow along on my own computer" (Introduction to NVivo for Windows participant #1), "the real examples [and being] more hands on" (Unix Shell and Command Line Basics participant #1) and "the interactive practice" (Learn to Program: R, participant #1) helped participants in their learning. Having time to practise the newly learnt material and "[h]aving small exercises to scaffold the larger concepts'' (Learn to Program: R, participant #2) were also comments that featured prominently.

In addition to these common themes, each question elicited further responses pertinent to each question. These are discussed in more detail below.

What were the most useful aspects of the training?
In addition to the common themes noted above, participants highlighted specific tasks and/or concepts from their training to be the most useful: 32.5% (50 responses) of 'most useful' responses that are related to Excel courses mentioned Pivot tables; 19.8% (140 responses) of 'most useful' responses that are related to Python and R courses highlighted functions and loops while 6.9% (49 responses) nominated data visualisation in these courses; 9.1% (24 responses) of 'most useful' responses that are related to REDCap and Qualtrics courses found designing a survey particularly useful. Participants across all courses liked the hints, tips, and shortcuts given throughout the training; while others acknowledged that what was learnt will be useful in their future work and research. A number of attendees (78 or 5.2%) indicated that the entire course was useful.

What were the least useful aspects of the training?
Complementing the responses in the Most useful aspects that the entire course was useful, 508 (42.2%) participants indicated that there were no aspects of the training that were not useful. This suggests that the content and delivery of our training are appropriate for our audience.

Several comments emerged which could be grouped under Content, with 125 responses related to this broad theme. Participants noted that they had previously learnt some of the material - particularly the introductory

material on Day 1 of a 2-day course - and therefore was not especially useful for them. However, they were also cognisant that "it is important to start from basics to get everyone on the same page rather than starting from an advanced level" (Excel for Researchers participant #1), and that it was "good to get a refresher!" (Learn to program: R participant #3) as well as being "good to know that I have been doing it mostly correctly to date!" (Data Capture and Surveys with REDCap participant #1). Consequently, these participants noted that the pace of the training was slow (36 responses). However, these comments were often tempered with the understanding that it was "needed to make sure everyone was on track" (Excel for Researchers participant #2).

While some participants noted that for them, some of the material was a little too complex, they qualified this by highlighting that particular tasks/concepts were "not yet relevant to me" (Learn to program: Python participant #2), or that "it is not part of my role as of yet" (Data Capture and Surveys with REDCap participant #2), and therefore it was not essential that the concept was fully grasped.

The final theme to emerge from this question concerned the Context of the training, with participants suggesting that the data set used could be more relevant, and that "more interaction with [the data set] would have helped to understand what we were doing" (Learn to program: Python participant #3). Although data sets are selected for their broad understanding and general applicability, it is not always perceived this way by our participants. A good example of this was the weather data set used for Excel for researchers training, where one participant from the Humanities and Social Sciences (HASS) discipline advised that "this data did not resonate with me, and the examples we went through weren't super applicable to the kind of data I would be working with" (Excel for Researchers participant #2). However, this provides an opportunity for Intersect to make available multiple data sets and select one based on the representative disciplines of attendees.

Other suggestions or feedback
Reflecting some of the Least useful comments around Content, responses to this question related to the level at which the course was pitched. Some asked for "simpler data sets" to be used (Data Manipulation and Visualisation in Python participant #1), while others asked for "more complex examples" (Data Manipulation in R participant #1). This speaks to the challenge of running short-course training in that meeting each participant at their particular level - even in introductory courses - is not straightforward nor simplistic. However, as discussed earlier, participants did qualify their responses recognising the need to cater to all levels of knowledge and experience, as exemplified by the following comment from participant #2 in the Data Manipulation in R course: "It was perfect as one couldn't have a course meeting all expectations of everyone when the participant's [*sic*] experience varied."

Similarly, suggestions for training to be "more applicable to real-life research" (Learn to program: R participant #4) are challenging: what is 'real-life research' to a social scientist may look very different to what a biologist considers 'real-life research'. Likewise, requests for advanced training "in some specific sectors like finance" (Data Manipulation and Visualisation in Python participant #2) or "courses on climate data" (Data Manipulation and Visualisation in Python participant #3) are not always easily accommodated as it can be difficult to develop advanced courses in niche areas of research in a cost effective manner, given the complexity of the material and breadth of expertise required. The Intersect membership model aims to address this through placement of a Digital Research Analyst who can provide further support, advice, and contextualisation of the material taught. Specialist data science and analytics support is also available on a fee-for-service basis if needed.

Overall, the comments received, whether they related to "most useful" or "least useful" aspects, were overwhelmingly positive. Where they may be deemed critical, the author often acknowledged that they understood why the material was presented that way, such as with the pace and level of training described above, or they provide opportunities for Intersect to continue refine and improve their training to meet universities' and researchers' expectations.

4. Communication

Most participants found out about the Intersect training program and the courses scheduled through the "Faculty/School newsletter" (~22%). The second and third top options are through the "Research office/division email" and "University newsletter" with approximately 15.5%, respectively. The "University website" and "Your Supervisor" were another two popular ways of raising awareness about Intersect training courses with about 10%, respectively. This demonstrates the importance of universities adopting a proactive approach to facilitating the dissemination of information about training through their internal communication channels, as these vastly outweigh Intersect or third party communication channels.

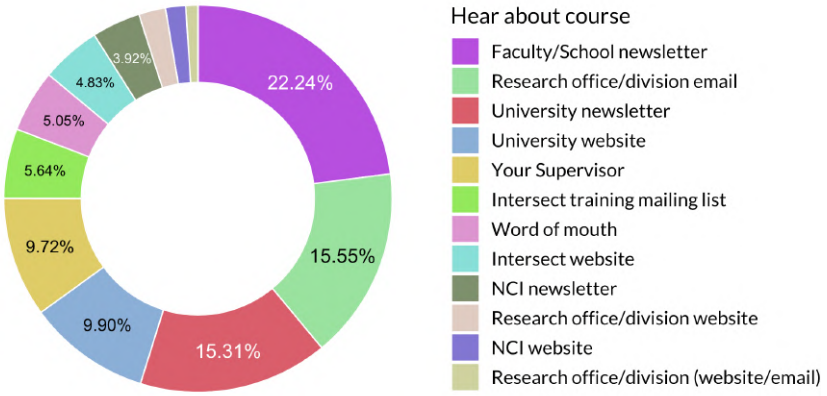


Figure 4.1: How did participants hear about the Intersect courses.

Figure 4.2 shows further analysis of the communication methods by Role/Position and by Faculty. HDR (PhD) students who are the biggest cohort attending Intersect training courses learn about Intersect training and the course scheduling through "Faculty/School newsletter" followed by "Research office/division email" and "University newsletter". These three communication channels also work best for Academic and Post-doc/Fellows who learn about Intersect training courses mostly through these channels. "Your supervisor" is a very popular communication channel among Undergraduate (Honours) but also among HDR students (PhD and Masters) as well as Professional staff (research-related and other).

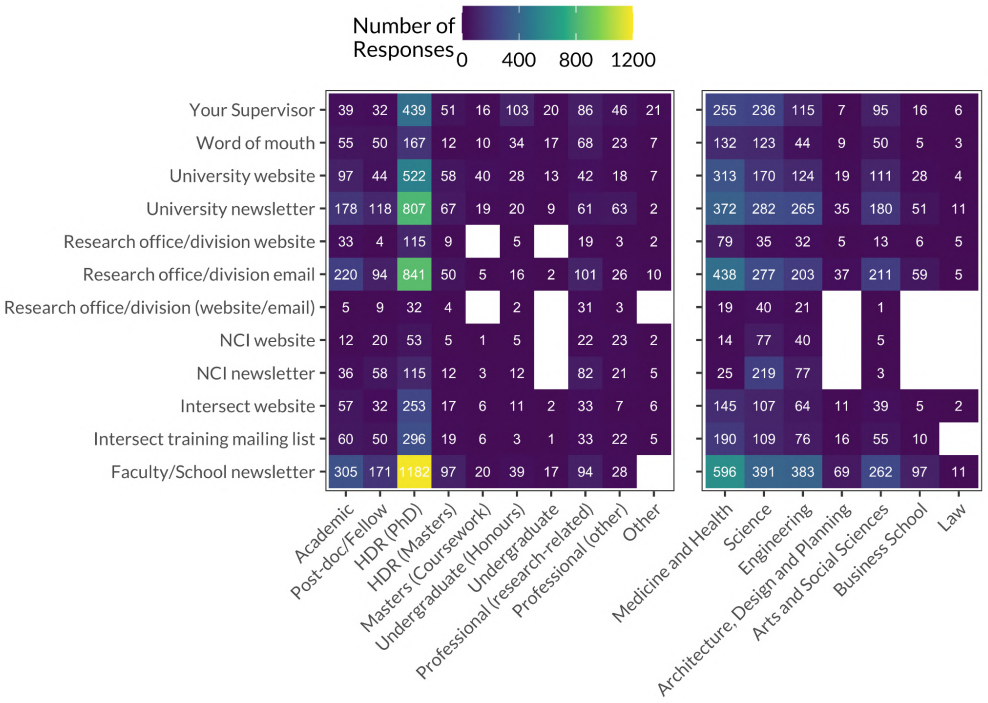


Figure 4.2: How did participants hear about the Intersect courses by Role/Position and by Faculty.

Participants from the Faculty of Medicine and Health hear about Intersect courses mostly via “Faculty/School newsletter” and to a lesser extent via “Research office/division email” and “University newsletter”. Participants from the Faculty of Engineering and the Faculty of Arts and Social Sciences hear about Intersect courses mostly through the same channels with the Faculty of Medicine. However, the Faculty of Science is observed to have a slightly different distribution of responses where researchers and HDR students from this faculty learn about Intersect training courses through five or more different communication channels.

5. Historical trends

Since the inception of Intersect’s Research Digital Skills Training Program in 2012, Intersect Australia has delivered over 2,200 courses and trained more than 32,000 researchers, HDR students, and professional staff in various universities, government departments, and organisations across Australia. Historically, the number of courses, as well as the training days, delivered annually consistently increased from 2012 to 2021, while 2022 shows a slight drop in both the number of courses and training days delivered (see Figure 5.1). In 2022, Intersect delivered 300 courses and 229 training days which indicates that Intersect delivered more than 6 courses on average every working week in 2022 (more than 1 course per working day).

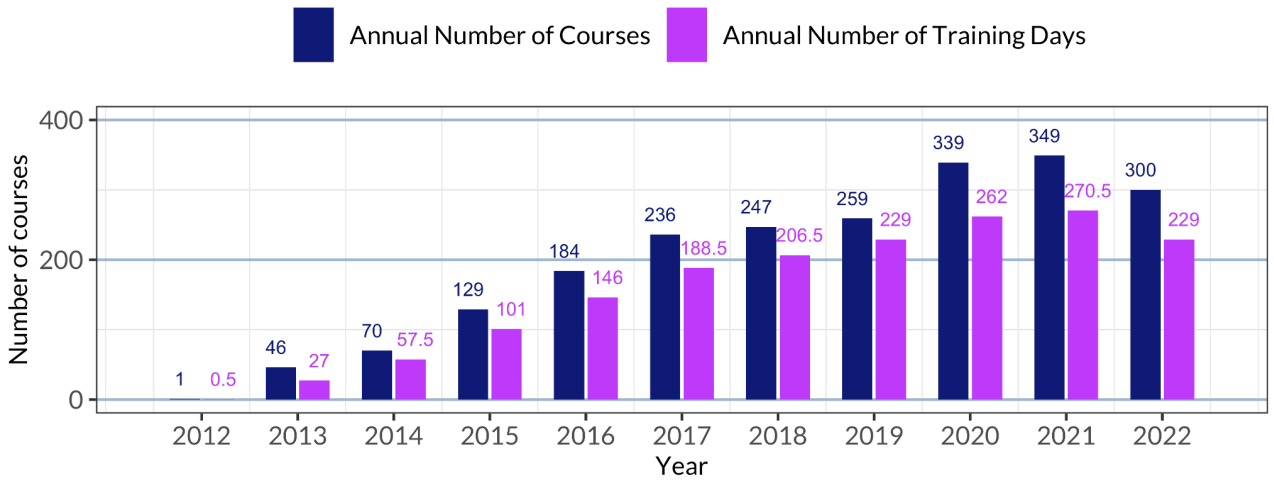


Figure 5.1: Total number of Intersect courses and training days delivered by Intersect Australia per year since 2012.

The number of researchers trained per annum has been consistently increasing from 2012 until 2021, while 2022 shows a slight decrease (higher number than 2020, but smaller than 2021) as shown in Figure 5.2 (bars and right y axis). The cumulative number of attendees per year shows an exponential increase until 2020 after which it changes to a linear increase (see line/points and left y axis in Figure 5.2). In 2022, Intersect reached a big milestone in the space of digital research technology training where over 30,000 participants have been successfully trained since the inception of our training program. This year, we anticipate reaching another huge milestone by successfully training over 20,000 *unique* researchers and HDR students across Australia, helping the Australian research workforce to upskill in emerging digital tools and technologies that are widely used in academia.

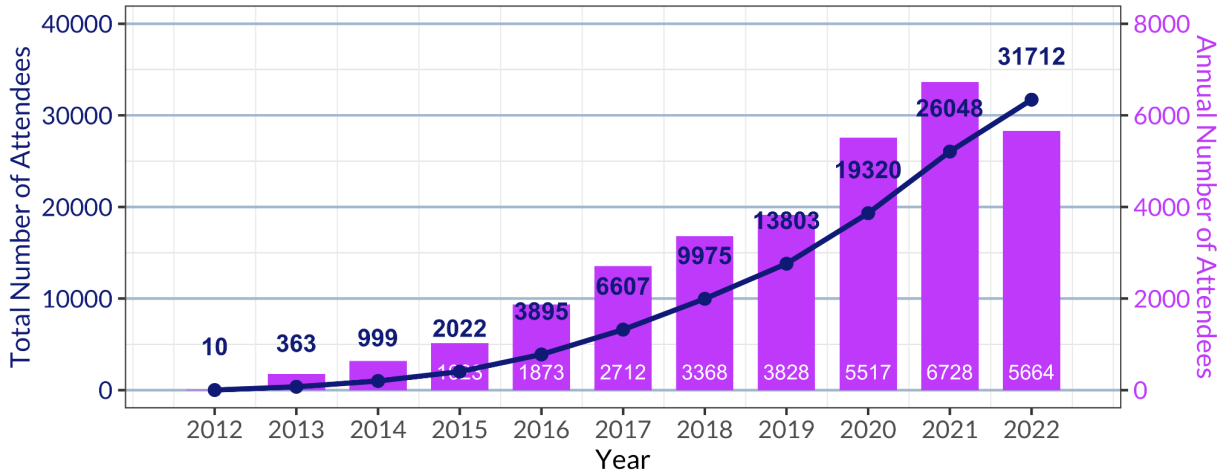


Figure 5.2: Annual and Total number of attendees since 2012.

Figure 5.3 shows the annual number of unique attendees and the cumulative number of unique attendees per year. At the end of 2022, Intersect has successfully trained over 17,500 unique HDR students and researchers across Australia, which makes Intersect one of the leading providers of research digital skills training in the Australasian region.

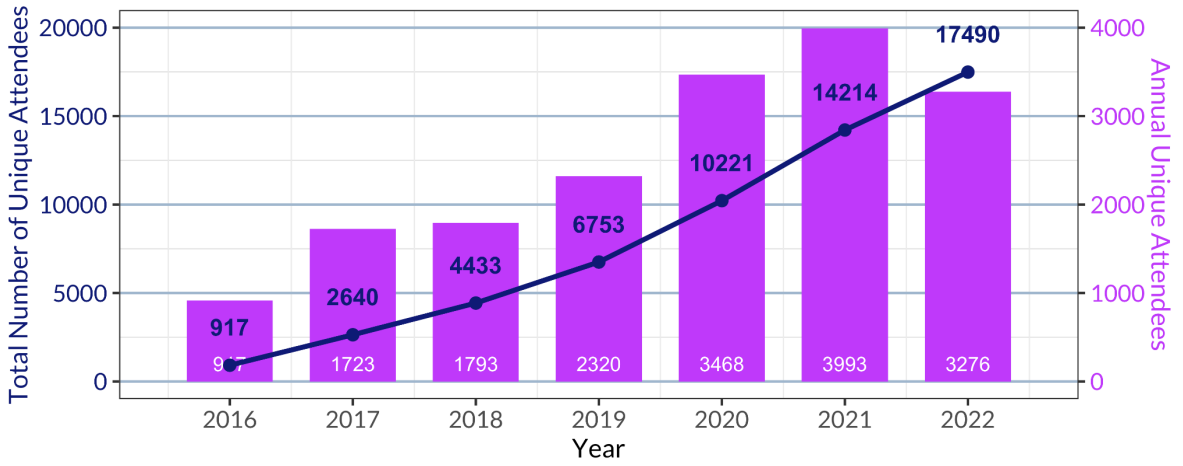


Figure 5.3: Annual and Total number of unique attendees since 2016.

Figure 5.4 compares the distribution of participants returning to attend more Intersect courses. The data presented in this figure include five time periods:

- Only 2018 attendees data
- Attendees data from 2018 to 2019
- Attendees data from 2018 to 2020
- Attendees data from 2018 to 2021
- Attendees data from 2018 to 2022

Analysing the 2018 attendees data only, 76.5% of participants attended only one Intersect course while 23.5% of participants returned to attend two or more Intersect courses. When taking into account a longer time period, *e.g.*, data from 2018 to 2022, the number of people who attended only a single Intersect course throughout these years drops to approximately 62% meaning that approximately two out of five participants return to attend more Intersect courses in the same or following years. Comparing 2018 only data with attendees data from 2018 to 2022, we observe a higher increase in the number of participants attending more Intersect courses while the percentage of participants attending a single course is decreasing significantly (see Table 5.1).

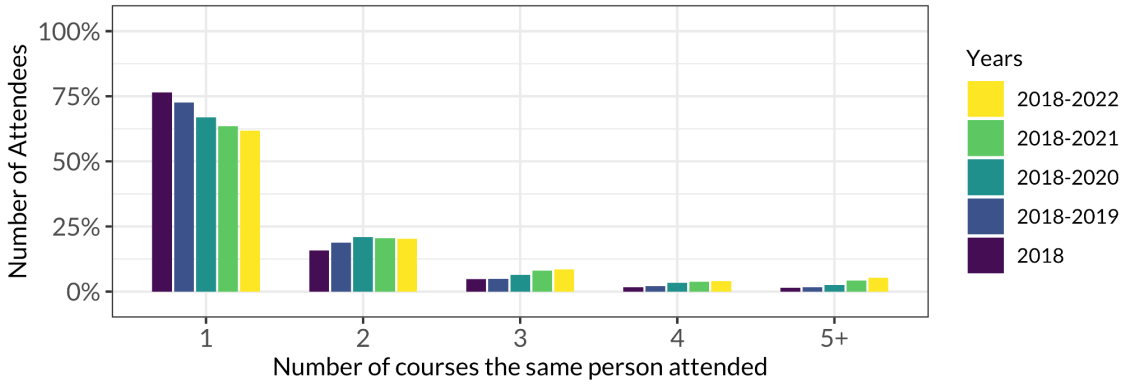


Figure 5.4: Number of courses that the same person attended. The data are presented in 2018 and then in consecutive years until 2022.

Table 5.1. Comparison of distributions of participants attending 1 to 5+ Intersect courses for 2018 only attendees data and attendees data for the 2018 to 2022 time period.

# of Courses	Only 2018 data	Data from 2018 to 2022	Difference
1	76.5%	61.9%	-19.3%
2	15.7%	20.3%	29.9%
3	4.8%	8.5%	81.0%
4	1.6%	4.0%	146.3%
5+	1.4%	5.3%	274.1%

Figure 5.5 shows the percentage of attendees by State/Territory in Australia. Although Intersect training was predominantly delivered in New South Wales (NSW) universities and organisations in 2016, in 2022 only 44.8% of attendees are researchers and HDR students from NSW universities and organisations. Victoria (VIC) is the second largest state in terms of number of attendees participating in Intersect training courses with 1149 attendees in 2022 or 20.7% of the total Intersect training attendees. University of South Australia joined as an Intersect member university in 2022 and together with the University of Adelaide have boosted Intersect training activities in South Australia (SA) with 1,130 SA attendees trained in 2022 which constitutes 20.4% of total attendees. The training partnership between Intersect and the National Computational Infrastructure (NCI) has also increased Intersect training activities in Australian wide organisations and universities as well as in other States and Territories across Australia.

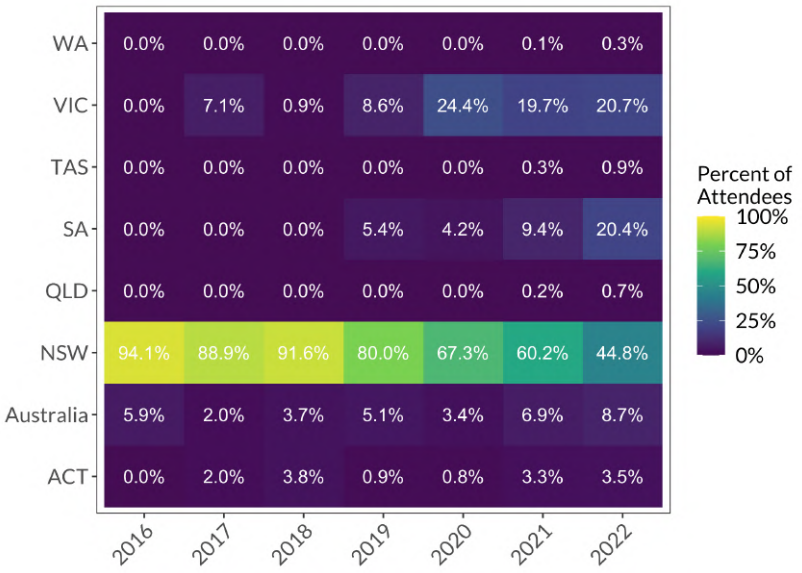


Figure 5.5: Percentage of attendees by state/territory since 2016.

Further insights can be derived by analysing the historical attendance data across several years and by Role/Position. Figure 5.6 presents the percentage of total attendees by Role/Position since 2017. The top consumer of Intersect training courses is consistently the HDR (PhD) students with the percentage fluctuating between 50% and 59%. Notably, although Academics constituted less than 3% in 2017 and 2018, a substantial increase was observed in 2019 with Academics constituting over 14% of attendees consistently since then. A slight increase was observed in 2020 and 2021 and we anticipate that this was most likely due to COVID-19, however this is only a hypothesis and is not based on any evidence or data. It is important to note that we restructured our Roles/Positions in our registration process in 2019 and therefore some inconsistencies occur in some of the Roles/Positions such as HDR (Masters) and Professional staff.

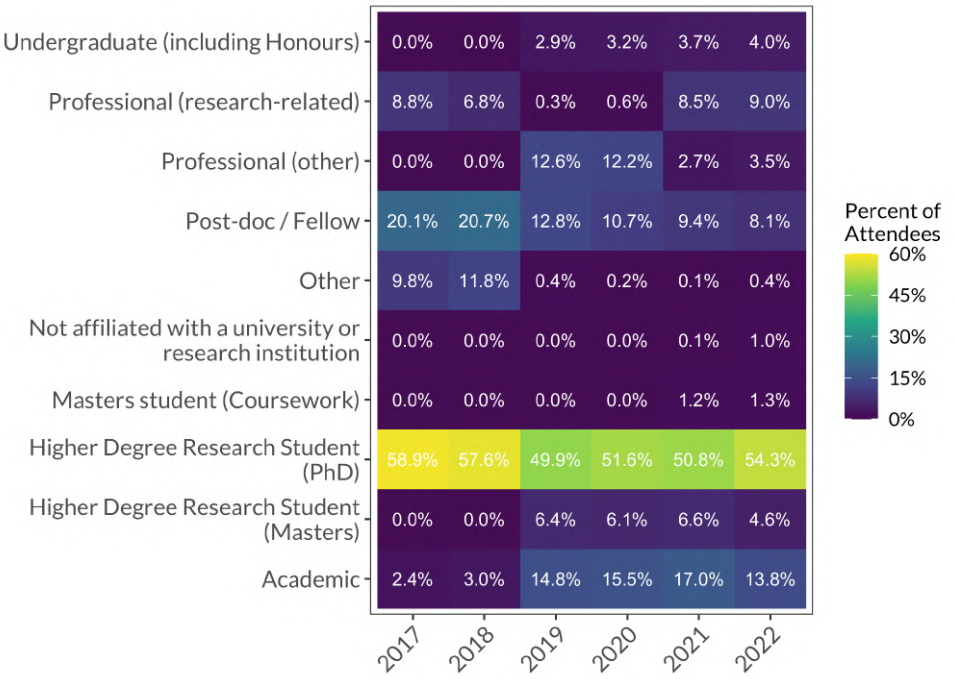


Figure 5.6: Percentage of attendees by Role/Position since 2017.

Figure 5.7 displays the percentage of attendees by Faculty since 2017. The Faculty of Medicine and Health is historically the top consumer of Intersect training with a percentage of attendees between approximately 26% and 35%. The Faculty of Science is following very closely as the second largest cohort of attendees (approx. 24% - 32%), with this faculty becoming the top consumer only in 2018. Although the third top consumer is historically the Faculty of Engineering, after 2019, a high rise of attendees from the Faculty of Arts and Social Sciences is observed surpassing the Faculty of Engineering during the pandemic (2020 - 2021) before dropping again to be the fourth top consumer in 2022. Since 2020, Intersect has invested time and effort to raise awareness of how digital tools and technologies can be beneficial and relevant to HASS researchers and HDR students and we are pleased to observe a higher uptake of Intersect training courses since then.

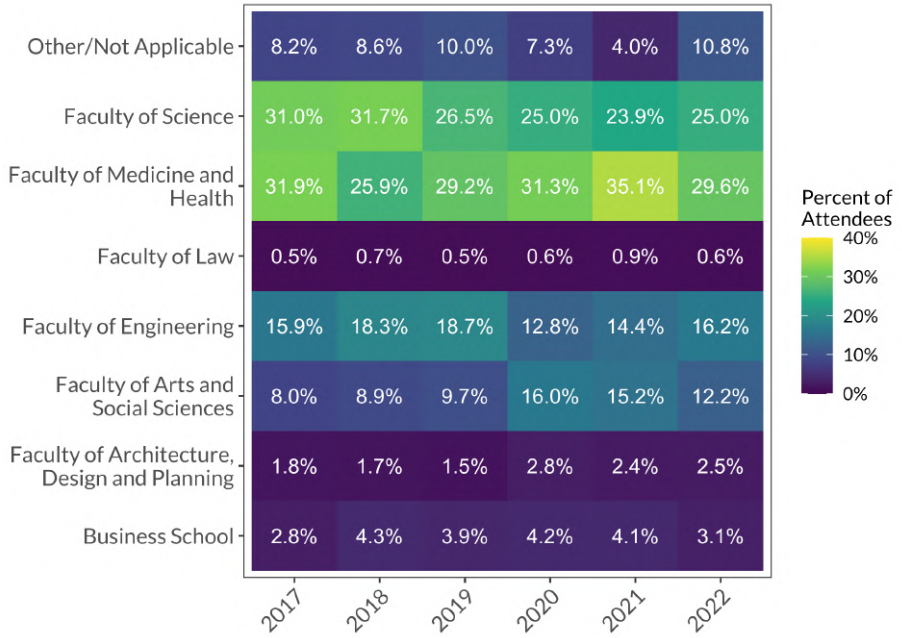


Figure 5.7: Percentage of attendees by Faculty since 2016.

Figure 5.8 shows the percentage of attendees per year split by the tool/technology. Until 2017, Excel courses were the most popular. However, in 2018, R and Python programming courses overtook Excel and became the most popular training courses Intersect delivers every year. Last year, despite Intersect introducing an intermediate Excel course, the percentage of attendance for Excel courses continued its year-on-year decrease, reaching less than 10% of total attendees for the first time in 2022. Since 2018, R programming courses continue to attract most attendees and be the most popular tool; very closely followed by Python which is consistently the second most popular tool/technology that Intersect teaches. Courses on surveying tools such as REDCap and Qualtrics started gaining popularity after 2018 and, in particular, in 2020 (the beginning of the pandemic), interest in these courses surged, reaching their highest ever percentage of attendance, with 12% and 11.7%, respectively. Since then, the percentage of attendance for REDCap courses has dropped but remains higher than in pre-pandemic years. In contrast,Qualtrics courses have dropped to lower percentages compared to the pre-pandemic years. Another interesting case is NVivo courses for qualitative analysis. It was introduced in 2020 and quickly became very popular. Notably, NVivo has surpassed Excel (the third most popular tool/technology in 2018) and became for the first time the third most popular tool taught by Intersect in 2022.

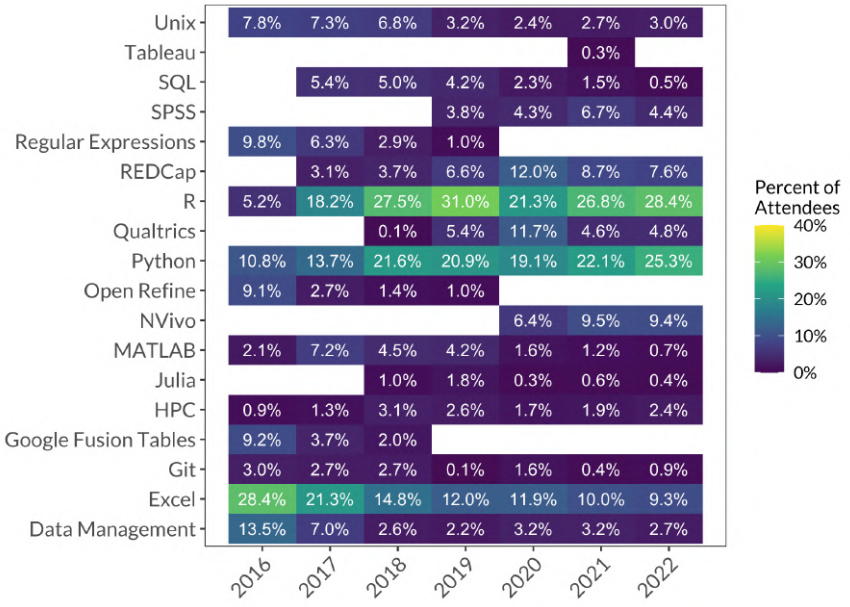


Figure 5.8: Percentage of attendees by Tool/Technology since 2016.

Intersect's annual NPS has increased rapidly over the years, from +46 in 2017, when NPS was first measured, to an NPS of above +70 since 2020. This demonstrates our commitment to the improvement of training quality and providing the best possible training experience to participants. Furthermore, Intersect managed to successfully move the entire course catalogue and delivery online after a tremendous effort from the Intersect team in the beginning of the pandemic in 2020. The consistently high NPS scores and metrics for measuring the quality of training suggest that moving our entire course catalogue and delivery online did not compromise the quality of delivery and training experience.

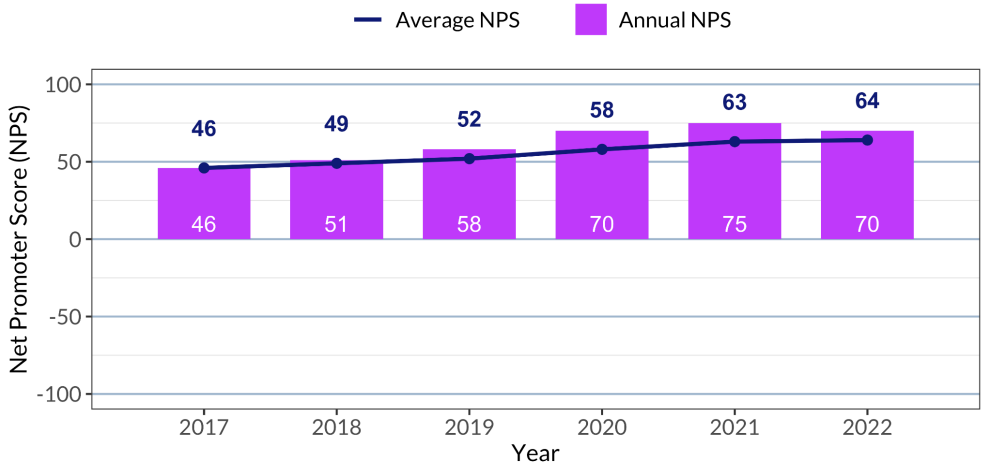


Figure 5.9: Average and Annual NPS per year since 2017.

As shown in Figure 5.10, all primary quality metrics consistently exceed on average 9.0 out of 10 since 2020, demonstrating a very high quality training delivery from very knowledgeable and experienced instructors in a training atmosphere that is welcoming participants to interact with the instructors and ask questions during the delivery.

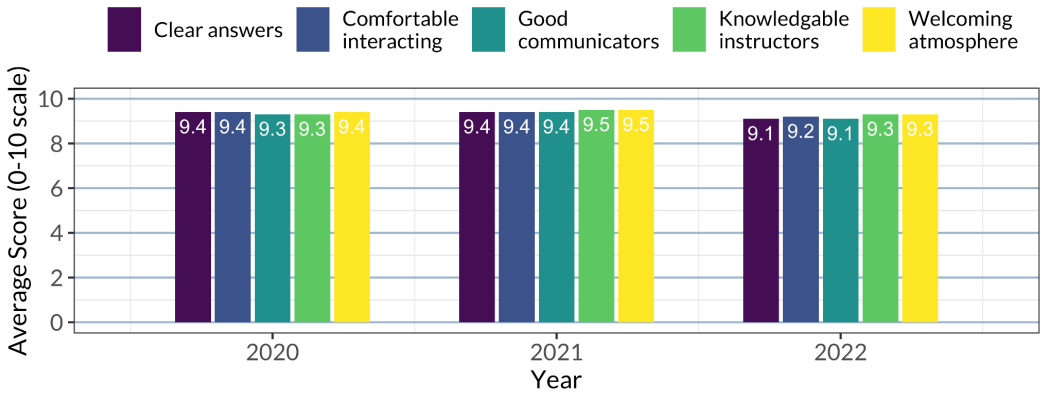


Figure 5.10: Primary metrics for measuring the quality of the training delivery per year since 2020.

Appendix A - Intersect training catalogue

Table 1: List of Intersect training courses as in March 2023.

#	Course Name	Course Code	Duration (Days)	Tool/Technology being taught
1	Excel for Researchers	EXCEL101	1	Excel
2	Beyond Basics: Conditionals and Visualisation in Excel	EXCEL201	0.5	Excel
3	Version Control with Git	GIT101	0.5	Git
4	Getting started with HPC using PBS Pro	HPC201	1	HPC
5	Getting started with HPC using Slurm	HPC202	1	HPC
6	Parallel Programming for HPC	HPC301	1	HPC
7	Learn to Program: Julia	JULIA101	0.5	Julia
8	Beyond the Basics: Julia	JULIA201	0.5	Julia
9	Learn to Program: MATLAB	MATLAB101	1	MATLAB
10	Getting started with NVivo for Windows	NVIVO101	0.5	NVivo
11	Getting Started with NVivo for Mac	NVIVO102	0.5	NVivo
12	Learn to Program: Python	PYTHON101	1	Python
13	Python for Research	PYTHON110	0.5	Python
14	Data Manipulation in Python	PYTHON201	0.5	Python
15	Data Visualisation in Python	PYTHON202	0.5	Python
16	Data Manipulation and Visualisation in Python	PYTHON203	1	Python
17	Introduction to Machine Learning using Python: Introduction & Linear Regression	PYTHON205	1	Python
18	Introduction to Machine Learning using Python: Classification	PYTHON206	1	Python
19	Introduction to Machine Learning using Python: SVM & Unsupervised Learning	PYTHON207	0.5	Python
20	Surveying with Qualtrics	QLTRICS101	0.5	Qualtrics
21	Learn to Program: R	R101	1	R
22	R for Social Scientists	R103	1	R
23	R for Research	R110	0.5	R
24	Data Manipulation in R	R201	0.5	R
25	Data Visualisation in R	R202	0.5	R
26	Data Manipulation and Visualisation in R	R203	1	R
27	Introduction to Machine Learning using R: Introduction & Linear Regression	R205	1	R
28	Introduction to Machine Learning using R: Classification	R206	1	R
29	Introduction to Machine Learning using R: SVM & Unsupervised Learning	R207	0.5	R
30	Exploring Chi-square and correlation in R	R210	0.5	R

31	Traversing t tests in R	R211	0.5	R
32	Exploring ANOVAs in R	R212	0.5	R
33	Research Data Management Techniques	RDMT001	0.5	RDMT
34	Data Capture and Surveys with REDCap	REDCAP101	0.5	REDCap
35	Longitudinal Trials with REDCap	REDCAP201	0.5	REDCap
36	Cleaning Data with Open Refine	REFINE101	0.5	Open Refine
37	Mastering text with Regular Expressions	REGEX101	0.5	Regexes
38	Regular Expressions on the Command Line	REGEX201	0.5	Regexes
39	Software Carpentry (Python)	SC101	2	Python, Git, Unix
40	Software Carpentry (R)	SC102	2	R, Git, Unix
41	Data Entry and Processing in SPSS	SPSS101	1	SPSS
42	Exploring Chi-Square and correlation in SPSS	SPSS102	0.5	SPSS
43	Databases and SQL	SQL101	0.5	SQL
44	Getting Started with Tableau for Data Analysis and Visualisation	TABLEAU101	0.5	Tableau
45	Unix Shell and Command Line Basics	UNIX101	0.5	Unix
46	Collecting Web Data	WEBDATA201	1	Python