Research Digital Skills Training Program

Intersect Australia - 2021 Report

Anastasios Papaioannou[1], Aidan Wilson[1,2], Ghulam Murtaza[1], Jianzhou Zhao[1], Marium Afzal Khan[1,3], Sam Ryan[1,4], Jonathan Arthur[1]

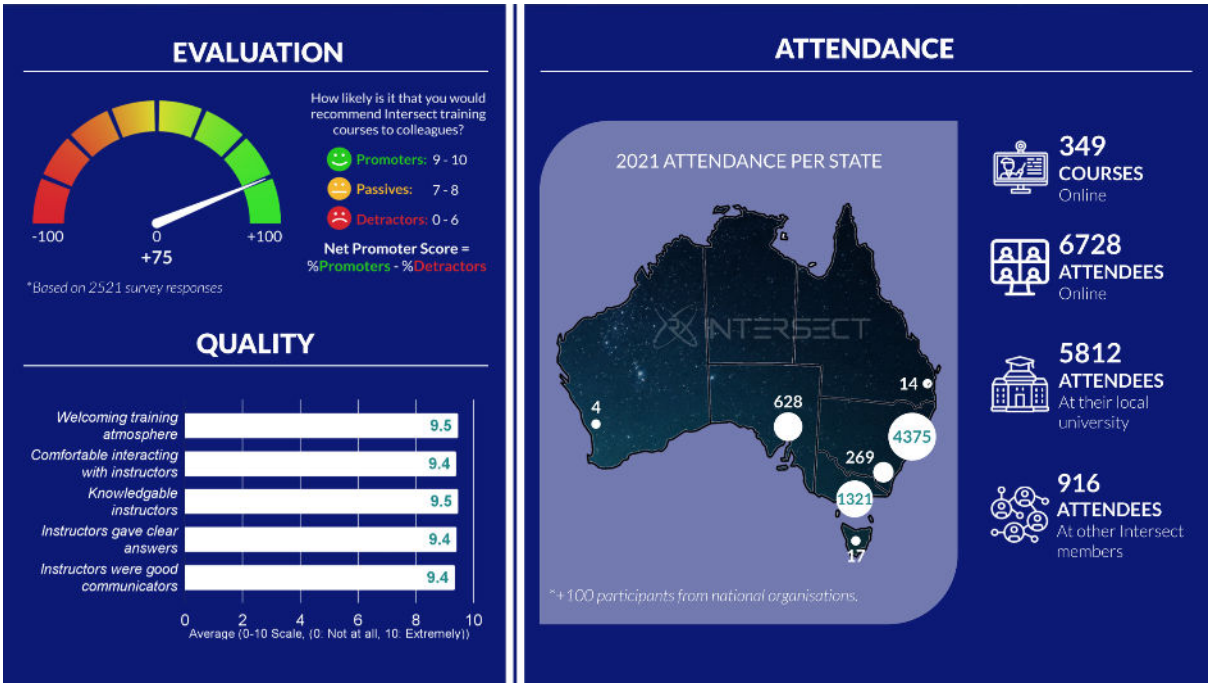
[1] Intersect Australia, Sydney, Australia; [2] Australian Catholic University, Sydney, Australia; [3] University of Adelaide, Adelaide, Australia; [4] La Trobe University, Melbourne, Australia

18 February 2022 | v1.0

This report summarises the research digital skills training program provided by Intersect Australia across Australia for the calendar year 2021 and insights on historical trends.

1. Training overview / Key highlights

In 2021, a total number of 6728 researchers were trained and 349 courses (270.5 training days) were delivered. The average attendance rate was approximately 77% based on the capacity (or registration number) of the courses. 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 +75 based on 2525 responses (~38% of attendees) which is considered outstanding. Furthermore, the average scores of the primary metrics for measuring the quality of the training delivery exceed 9.4 out of 10, which indicates that feedback from participants is excellent.



[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

Figure 2.1 shows the number of courses by Tool/Technology that were delivered in 2021. R and Python programming courses were the most popular, followed by REDCap, Excel, Nvivo, and SPSS. The courses were spread across the year with Q2 and Q3 being the quarters when the majority of the courses were delivered.

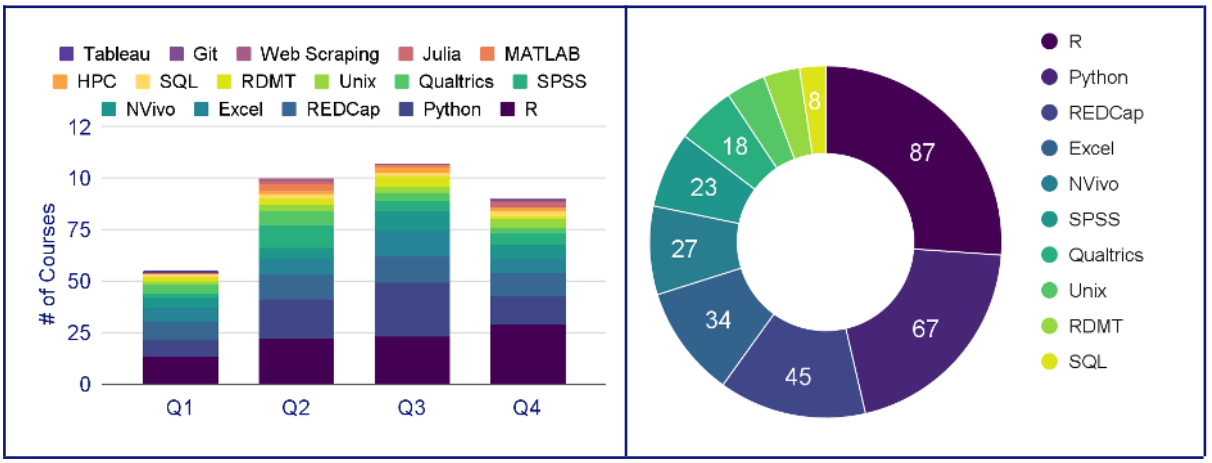


Figure 2.1: Left: Quarterly number of courses by Tool/Technology; Right: Total number of courses by Tool/Technology.

Figure 2.2. presents the distribution of course attendance. Attendance seems to be normally distributed with the peak found in the attendance between 15-20. In 2021, the average number of attendees per course was 19[3].

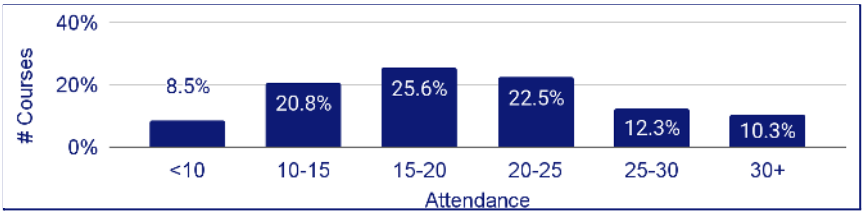


Figure 2.2: Distribution of course attendance.

About 14% of total participants attended a course at a university other than their local university, through Intersect's consortium of member universities, Intersect training partnerships with other organisations, and Intersect's Open Training.

According to Figure 2.3, R and Python courses had the highest number of attendees, 1,831 and 1,431 respectively. These combined constitute almost 50% of the total attendance in 2021. Excel, NVivo, REDCap, SPSS, and Qualtrics courses were also well attended with an attendance from 300 to 700.

[3] To accommodate the high demand in some member universities, Intersect delivers courses where the eResearch Analyst delivers the course solo, limiting the capacity to 15 attendees. When excluding courses where a limited number of attendees was applied, the average number of attendees per course increases to 21.2.

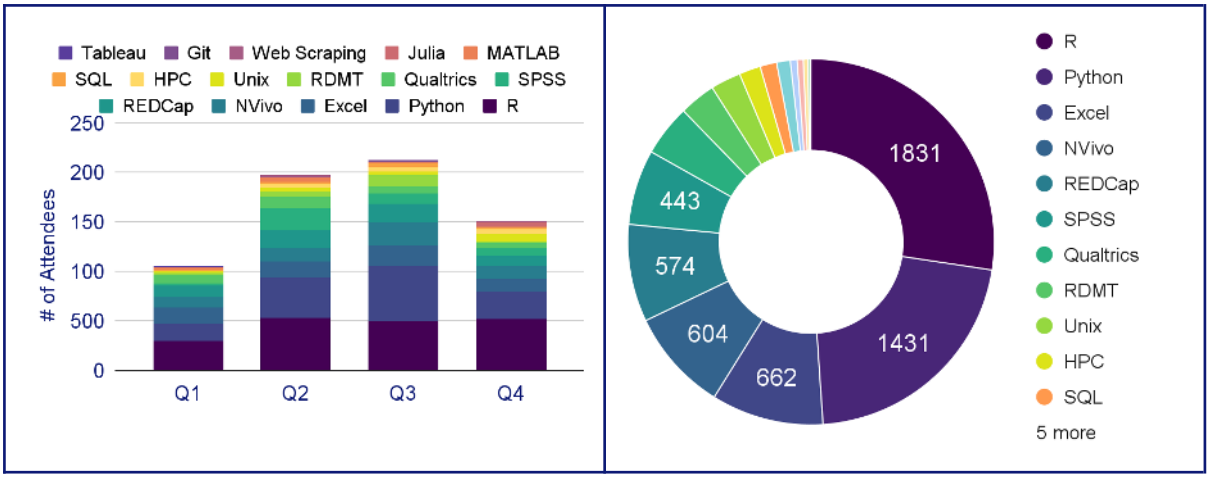


Figure 2.3: Left: Quarterly number of attendees by Tool/Technology; Right: Total number of attendees by Tool/Technology.

Figure 2.4 shows that Higher Degree Research students (PhD) are the top consumer of Intersect's training program with an attendance of 50.8% followed by Academic, Post-doc/Fellow, and Professional (research-related), with an attendance of between 8-17%.

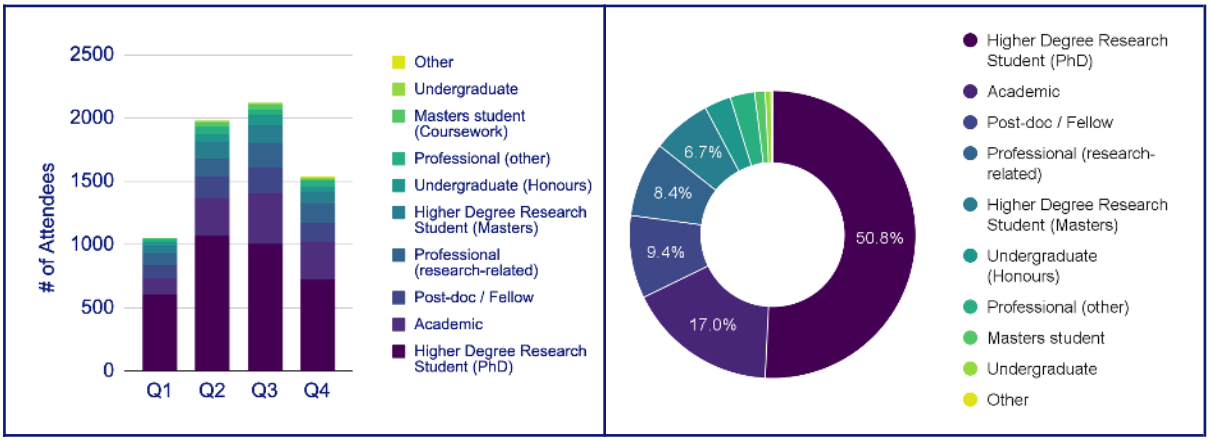


Figure 2.4: Left: Quarterly number of attendees by Role/Position; Right: Total number of attendees by Role/Position.

In addition, approximately a third (~37%) of the training attendees are from the Faculty of Medicine and Health[4] with the attendance from the Faculty of Science being approximately 25%. Interestingly, the third biggest consumer of the training program are researchers from the Faculty of Arts and Social Sciences. This demonstrates an increasing uptake of courses in digital tools and technologies in that faculty. The Faculty of Engineering also shows a high attendance constituting 15% of the total attendees.

[4] The Faculty information is not captured directly from participants but instead is mapped based on the Field of Research (FoR) code that is captured in the registration form.

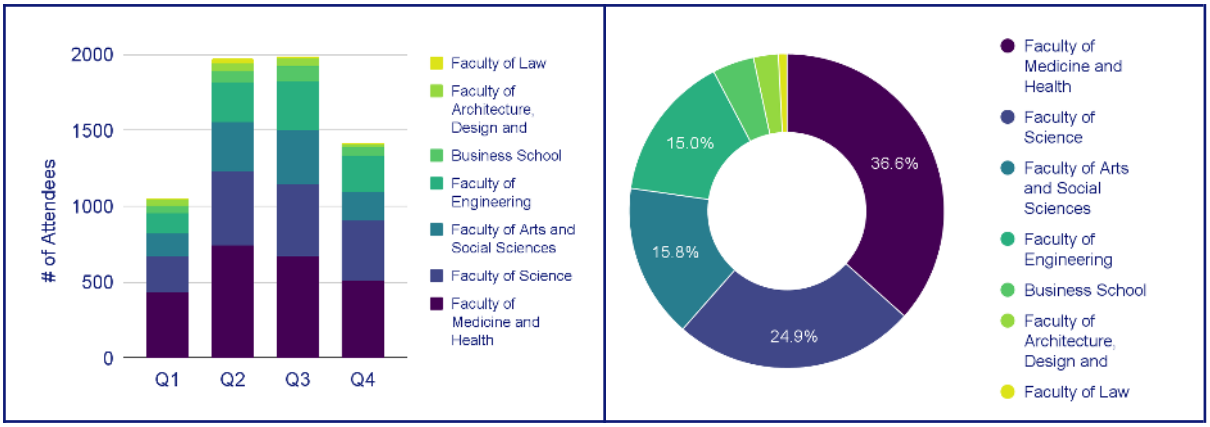


Figure 2.5: Left: Quarterly number of attendees by Faculty; Right: Total number of attendees by Faculty.

Analysing the data based on the Field of Research (FoR) code, Health Sciences is on top of the list with 1,644 attendees followed by Engineering, Biological Sciences, Biomedical and Clinical Sciences and Education as shown in the table. (The table only includes the top 10 FoR codes based on the attendance.)

#	FoR Code	# Attendees
1	42 Health Sciences	1644
2	40 Engineering	686
3	31 Biological Sciences	559
4	32 Biomedical and Clinical Sciences	464
5	39 Education	455

#	FoR Code	# Attendees
6	52 Psychology	339
7	41 Environmental Sciences	288
8	35 Commerce, Management, Tourism and Services	277
9	46 Information and Computing Sciences	276
10	30 Agricultural, Veterinary and Food Sciences	243

Figure 2.6 and 2.7 shows the attendance by tool and technology which is further split by Role and Faculty, respectively. Interestingly, the distributions vary when comparing different roles with different tools/technologies. Furthermore, it is worth noting that there is an increased uptake of NVivo and Qualtrics, as well as Tableau and Web Scraping courses, by the Faculty of Arts and Social Sciences, in comparison to the other faculties. Regarding surveying tools, REDCap attracted more attendees from the Faculty of Medicine and Health, as compared with Qualtrics where the attendance is almost equally distributed between the Faculty of Medicine and Health and the Faculty of Arts and Social Sciences. Regarding programming courses, Python showed a different distribution compared to R with the vast majority of attendees being from the Faculty of Science and Faculty of Engineering while in R they are drawn from the Faculty of Medicine and Health and the Faculty of Science.

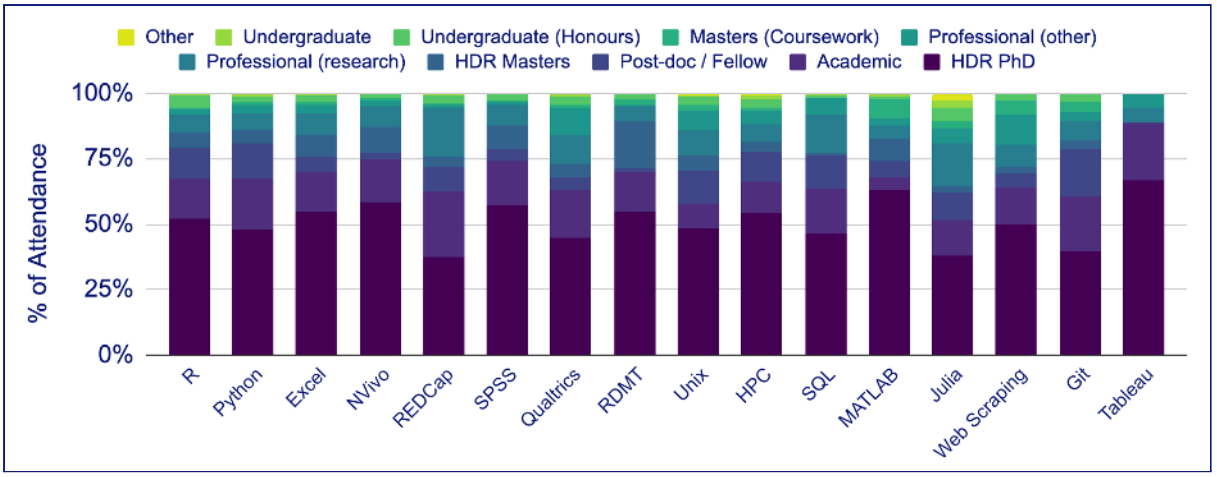


Figure 2.6: Tool/Technology attendance by Role/Position.

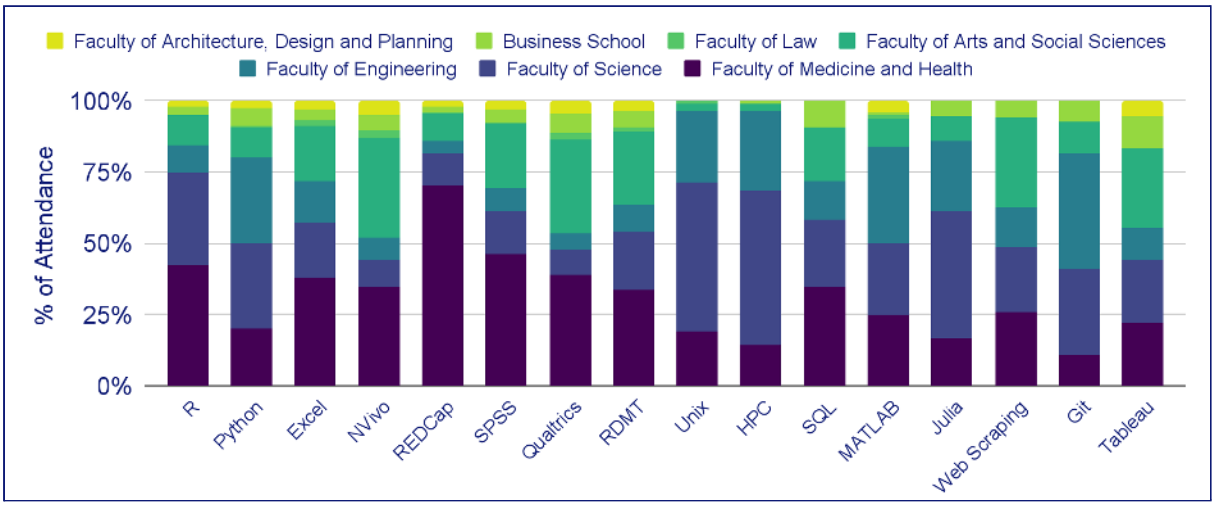


Figure 2.7: Tool/Technology attendance by Faculty.

When participants register, they are also asked to provide reasons for attending the course[5]. Most participants answered that they are interested in these courses to learn skills that they can either apply to their work now or in the near future. Notably, a considerable number of attendees are learning these skills for better opportunities and employability in the future ("To learn skills that will help me get a job").

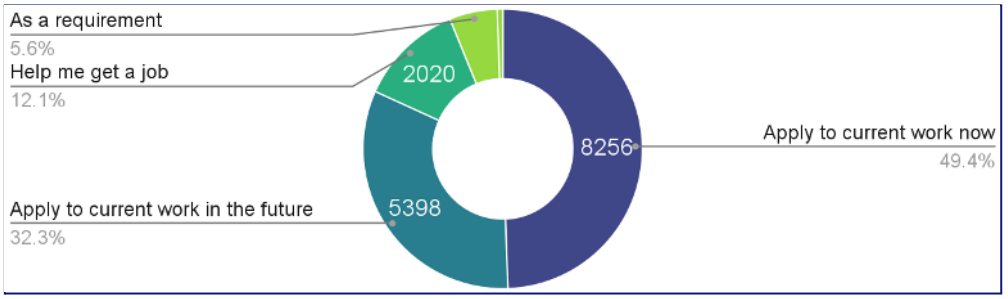


Figure 2.8: Reason for being interested in attending this training course.

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 students where the percentage of those learning these skills for better opportunities is higher. Analysis on the reason for

[5] The data presented in this reason for attending a course part of the analysis refer to registrants and not just to attendees. Note that this is a question that accepts multiple responses, and therefore the numbers add up to more than the total number of registrants in 2021.

being interested in attending this training course by the Faculty reveals a consistent trend across all faculties.

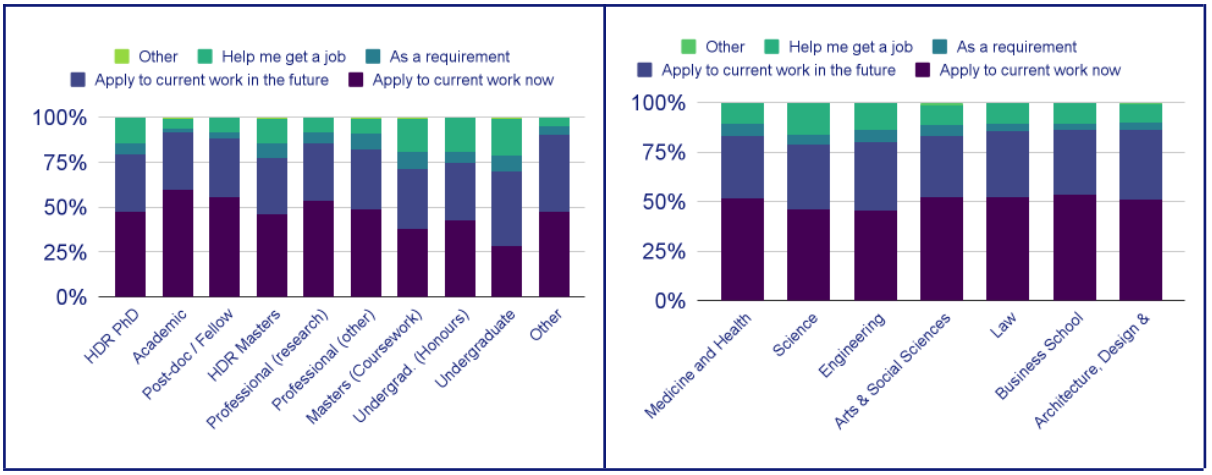


Figure 2.9: Left: Reason for being interested in attending this training course by Role/Position; Right: Reason for being interested in attending this training course by Faculty.

Figure 2.10 shows the percentage of researchers who attended 1, 2, 3, 4, or 5 or more Intersect courses in 2021. Interestingly, more than 40% of the researchers attended at least 2 Intersect courses in 2021, while researchers attending 3 or more courses is about 20% (1 out of 5).

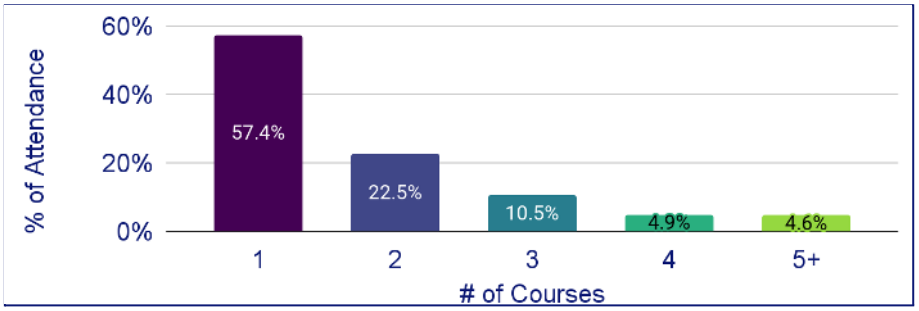


Figure 2.10: Percentage of attendance by the number of Intersect courses attended in 2021.

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. The Intersect training program demonstrates high quality and is well received by the attendees. Intersect scored an NPS of +75 in 2021, which is the historical high NPS at Intersect since the inception of the Intersect training program in 2012. All courses have achieved an NPS of over +60 apart from Tableau[6]. A positive NPS is considered great, while achieving an NPS of +50 or higher is considered outstanding and seldom achieved commercially.

[6] The Tableau course was delivered only once as a pilot course in 2021 (after we finalised the course development) in order to capture participants' feedback to improve the course content.

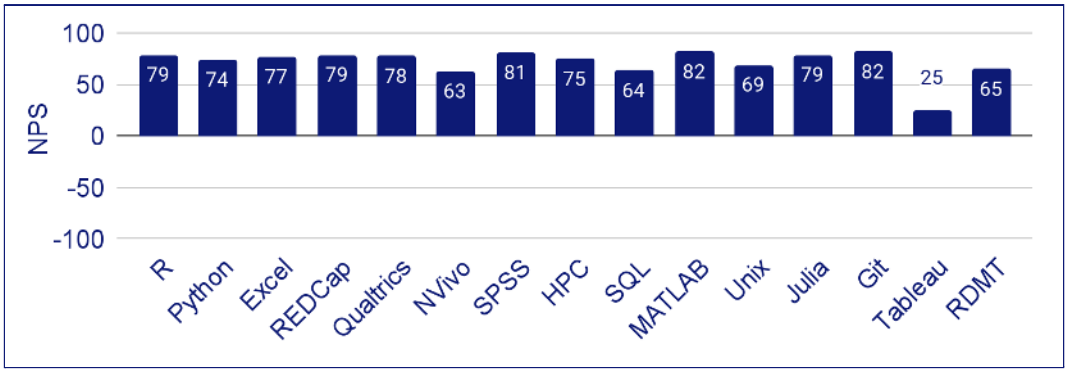


Figure 3.1: NPS per Tool/Technology.

The NPS score was further calculated by quarter and class size (Figure 3.2). A very high NPS score is observed in all quarters and all different classes. It is worth noting that even classes with 25 or 35 attendees achieved a consistently very high NPS score indicating that the quality of the course delivery does not drop in courses with higher attendance. However, further analysis that was carried out in historical data (based on over 7,500 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.

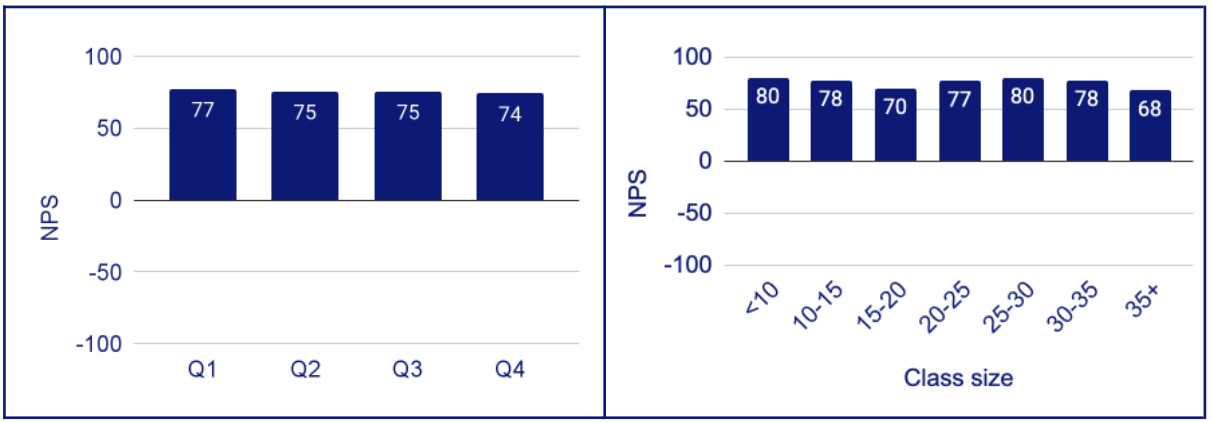


Figure 3.2: Left: NPS per quarter; Right: NPS by class size.

Figure 3.3 shows the five primary metrics for measuring the quality of the training delivery per quarter. All metrics are 9.3+ 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.

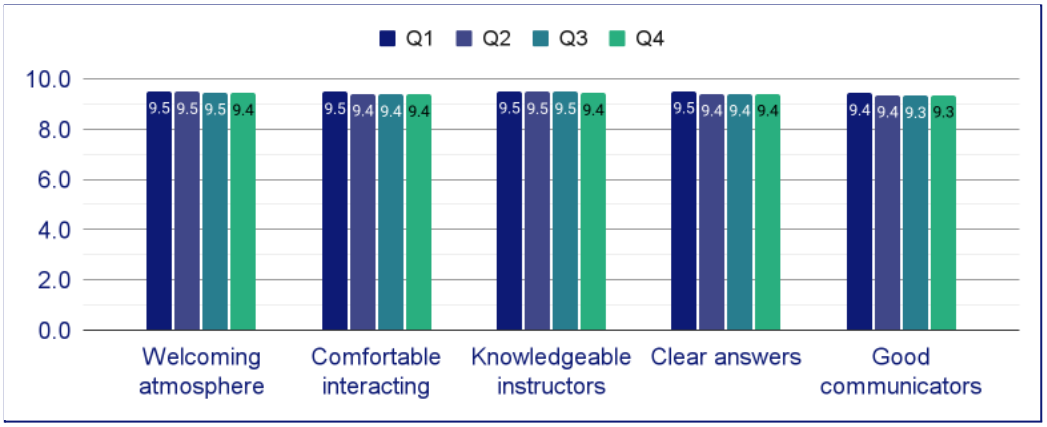


Figure 3.3: Primary metrics for measuring the quality of the training delivery per quarter.

As shown in Table 3.1, the primary quality metrics were further analysed by class size demonstrating that all five primary metrics for measuring the training quality have consistently achieved a very high average score (9.3+) in all class sizes.

Table 3.1: Average number of the primary training quality metrics by class size.

Class size	Welcoming atmosphere	Comfortable interacting	Knowledgeable instructors	Clear answers	Good communicators
<10	9.5	9.5	9.6	9.6	9.5
10-15	9.5	9.5	9.6	9.5	9.5
15-20	9.4	9.3	9.3	9.4	9.3
20-25	9.5	9.4	9.5	9.4	9.4
25-30	9.5	9.5	9.5	9.5	9.5
30-35	9.5	9.4	9.4	9.3	9.3
35+	9.5	9.5	9.5	9.4	9.5

Qualitative analysis was carried out in the three open-ended questions of the Intersect Course Evaluation survey:

- *"Which parts of the course did you find most useful?",*
- *"Which parts of the course did you find least useful?"*
- *"Do you have any other suggestions or feedback on this course or Intersect Training in general?"*

The most frequent words that appeared in the qualitative feedback are presented in Figure 3.4. In the least useful question (left chart), the words "time", "bit", and "basic" were commonly mentioned. This might indicate that some participants found parts of the courses to be more basic than expected for them. However, Intersect clearly mentions the course outline and learning objectives in the registration page. Furthermore, our trainers also ask participants what they think about the pace during the delivery of a course and adjust accordingly. In the most useful section (centre chart), the words "data", "learning", "functions", "examples", "step", and "coding" were among the most frequent words. This might indicate that the participants found the stepwise demonstration of learning and the problem-solving with examples and relevant datasets very helpful and useful in the training. In the "Other feedback" (right chart), frequent words include "training", "session", "helpful", "people" and "instructor", where further analysis showed that participants found the training/session helpful in one way or the other and also appreciated the instructors' delivery and assistance.

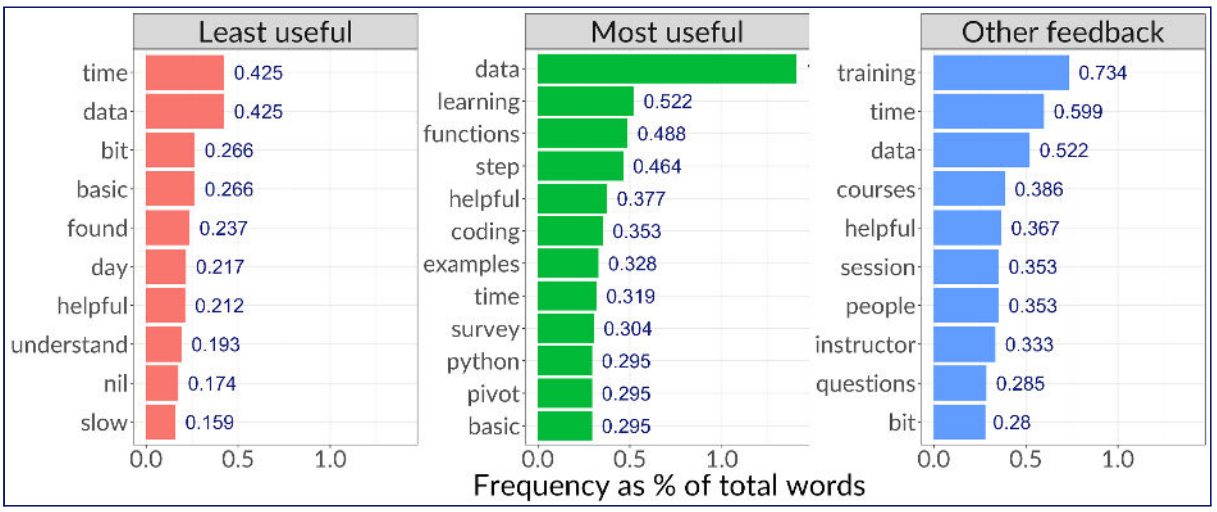


Figure 3.4: Most frequent words for each open text question in the Intersect Course Evaluation survey, namely, "Which parts of the course did you find most useful?", "Which parts of the course did you find least useful?" and "Do you have any other suggestions or feedback on this course or Intersect Training in general?".

Further qualitative analysis shows the most common bigrams (pairs of consecutive written words) of the top frequent words as found in each open-ended question (see Figure 3.5). Common bigrams of the "least useful" feedback (left chart) include "time spent" and "bit slow" which shows that participants provide constructive feedback on what parts of the training they think instructors should have spent less time on and also on what parts of the training instructors should have gone a bit faster. It is worth mentioning that Intersect is dedicated to constantly evaluating and improving the quality of delivery by taking into account all the constructive feedback provided by the researchers.

From the common bigrams of the "most useful" section (centre chart), "pivot-table(s)" was found to be the most useful thing for the participants of the Excel courses. Furthermore, "data-analysis", "data-cleaning", "data-visualisation", and "data-manipulation" were other commonly found bigrams in the feedback from the participants of the programming courses. "Step-instruction(s)" and "practical-examples" were other common bigrams observed in the "most useful" section. Based on the common bigrams of the "Other feedback" (right chart), the most common bigrams refer to the pace of the course (e.g. "bit-slow"), the various aspects of the training session(s), the instructors (e.g. "excellent-instructor", "answering-questions"), as well as some topics covered in the training courses (e.g. "data-set", "data-analysis", "research-data").

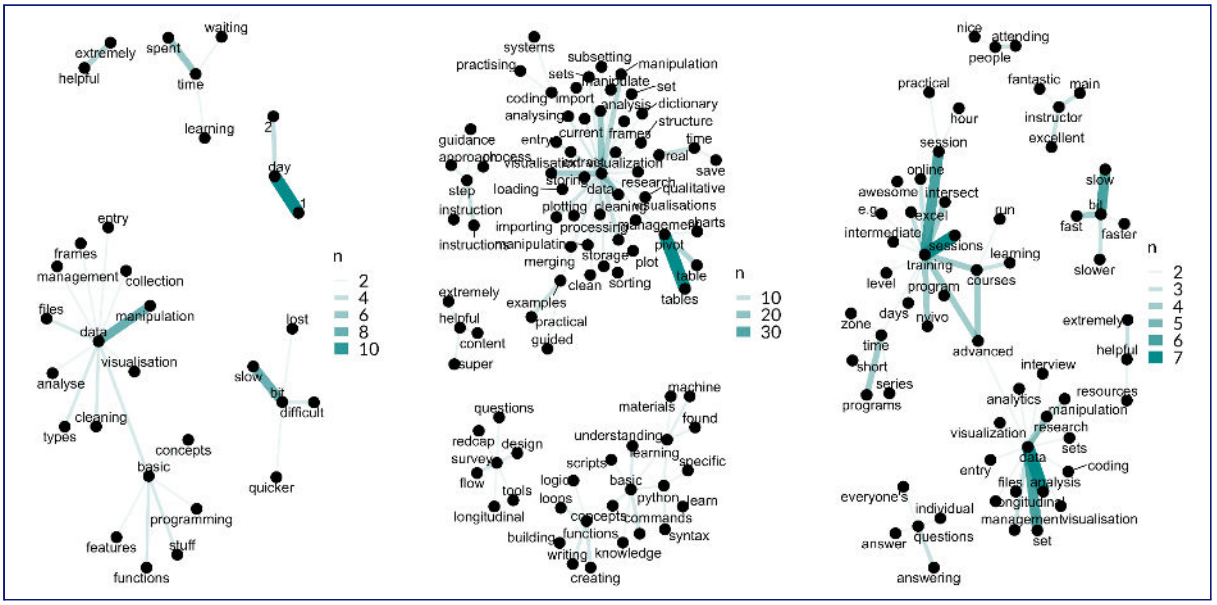


Figure 3.5: Most common bigrams (pairs of consecutive written words) of the top frequent words as shown in Figure 3.4. From left to right, plot for the "Which parts of the course did you find least useful?" open text question, the "Which parts of the course did you find most useful?" question and the "Do you have any other suggestions or feedback on this course or Intersect Training in general?" question.

The following shows some qualitative feedback from participants:

- *"The instructors gave a good overview of many different features of Qualtrics which I see will be valuable for research. Also, he stayed back in the breaktime to answer questions which was great."*
- *"For me the most useful part of the course was learning how to use the Syntax in an effective and efficient manner in SPSS. I look forward to apply what I have learned to improve my research management style."*
- *"Emphasising the importance of thinking about your structure, how you want to export and follow up and how it can impact the design of your survey was very helpful."*

- *"Amazing course, as a young researcher dreading using R and upset trying to even figure out how to open files in R - I have gained so much inspiration from the women teaching and running this course that I am ambitious to gain more skills in using R. Thank you for making something so complicated feel achievable."*
- *"Statistics with pandas and visualisation with matplotlib. Basically, most of the Python course is new to me and improve my current method of processing data in my research."*
- *"This was helped a lot by the excellent preparation instructions which were extensive and straightforward to follow. It meant I had all the important parts already stepped up and learned the most (on functionality and usability) in the time we had. I started with no knowledge, learned how to install everything from the email instructions (took approx an hour the night before), and learned how to use Nvivo to a point I am confident I can do some of the basics now (from approximately 2 hours training). Completely worth my time to attend."*

4. Communication

Most participants found out about the Intersect training program through the Faculty/School newsletter (~29%). The second top option is through the Research office/division (website/email) with ~18% followed by the University newsletter and University website. Notably, a considerable number of participants learn about these courses through their supervisors (~9%).

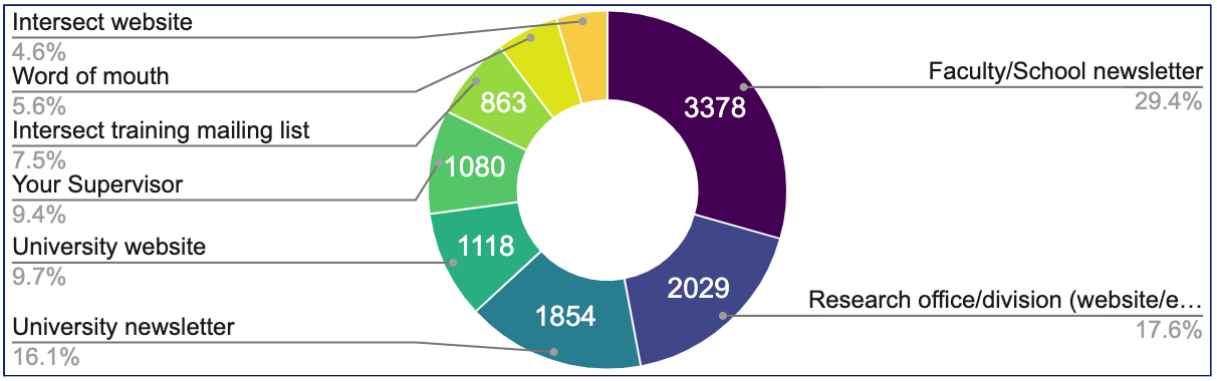


Figure 4.1: How did participants hear about the course?

Figure 4.2 shows further analysis of the communication methods by Role/Position. Faculty/School newsletter, Research office/division (website/email) and University newsletter were the major sources from which researchers obtained the training information. It is also interesting to note that HDR Masters students, Professional staff (other) and Undergraduate students heard about the courses from their supervisors.

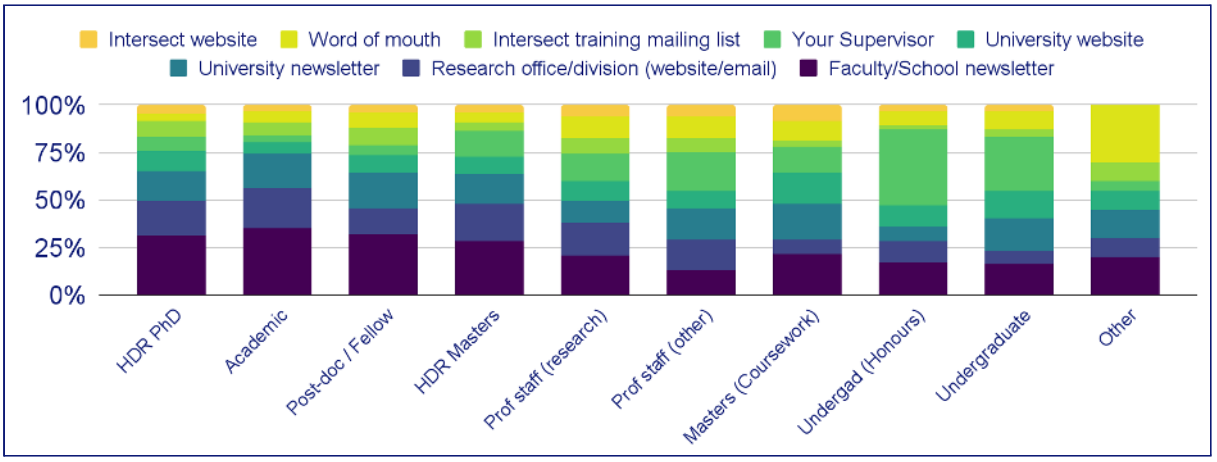


Figure 4.2: How did participants hear about the course by Role/Position.

4. Historical trends

Since the inception of Intersect's Research Digital Skills Training Program in 2012, Intersect Australia has delivered over 1,800 courses to more than 26,000 researchers and HDR students. Historically, the number of courses, as well as the training days delivered annually is constantly increasing (see Figure 4.1). In 2021, Intersect delivered 349 courses and 270.5 training days which indicates that Intersect delivered more than 6.5 courses on average every working week in 2021 (more than 1 course per working day).

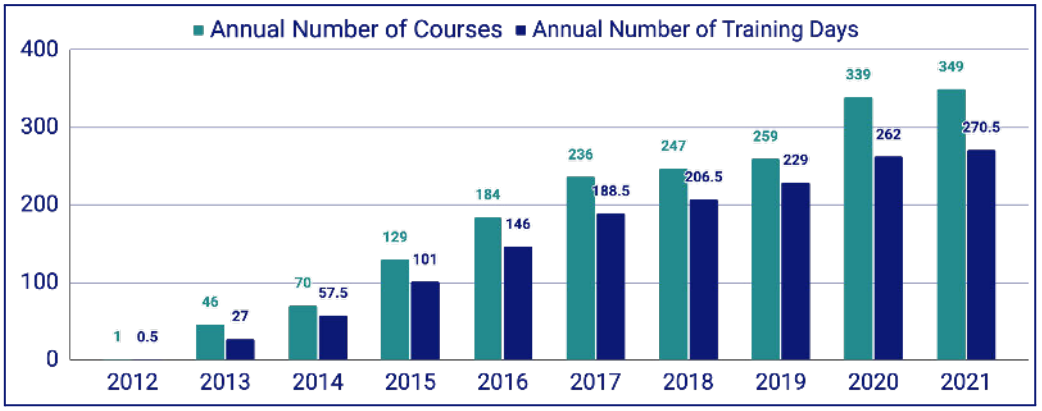


Figure 4.1: Total number of Intersect courses and training days delivered by Intersect Australia per year since 2012.

The number of researchers trained per annum has been consistently increasing since 2012 while the cumulative number of attendees per year shows an exponential increase (Figure 4.2). In 2021, Intersect reached a huge milestone in the space of digital research technology training where over 20,000 HDR students and researchers have been successfully trained since the inception of our training program. You can read more about this training milestone in an article published in June 2021.

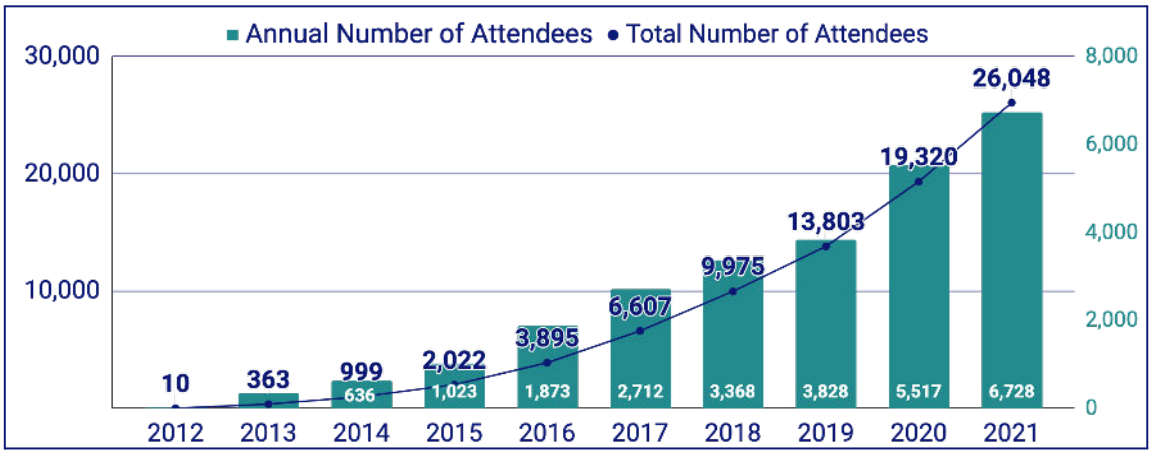


Figure 4.2: Annual and Total number of attendees since 2012.

Note that the 20,000 training milestone refers to the number of attendees and not the individual number of HDR students and researchers. Further analysis shows the annual number of unique attendees and the cumulative number of unique attendees per year as displayed in Figure 4.3. In 2021, Intersect has successfully trained over 12,500 unique HDR students and researchers across Australia, which makes Intersect one of the leading providers of research digital skills training in the Australasian region.

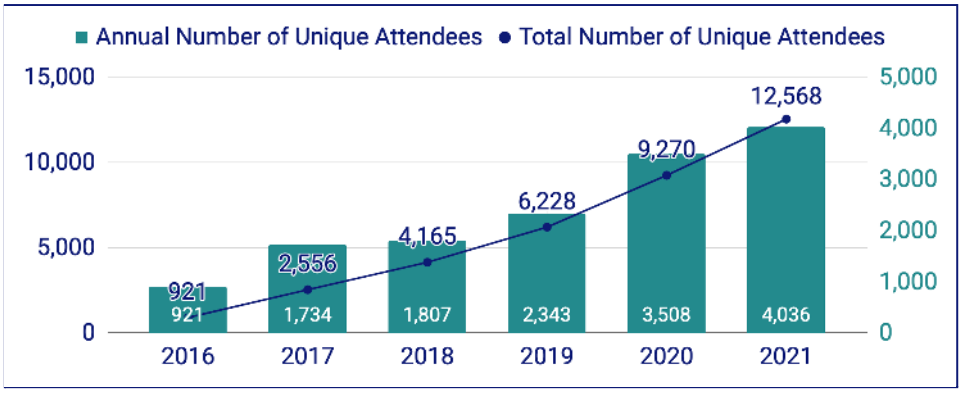


Figure 4.3: Annual and Total number of unique attendees since 2016.

Further analysis demonstrates the total number of attendees per year split by Role and Faculty[7] (see Figure 4.4). Every year, the top consumer of Intersect's training program is the HDR (PhD) students constituting at least 50% of the total attendees. Interestingly, the biggest increase in the number of attendees was observed in the Academics where ~70 attended Intersect training courses in 2018 compared to ~1,100 in 2021. Similarly, an substantial increase in the uptake of Intersect training courses was observed in the professional staff (research-related). This might indicate that a higher number of higher profile researchers and professional staff are undertaking these training courses to up-skill themselves due to the fact that digital tools and technologies are becoming an integral part of the research lifecycle. Furthermore, online delivery of training might have also contributed to this as it makes training more accessible to researchers and research professionals who are not based at main university campuses. Flexibility in terms of time commitment could be another factor as it is easier to commit two half-days online rather than a full day in person.

In addition, about a third of the training attendees are from the Faculty of Medicine and Health followed by the Faculty of Science, the Faculty of Arts & Social Sciences and the Faculty of Engineering. Notably, a constant increase in the absolute number of attendees is observed in all individual Faculties every year. However, the highest increase is in the Faculty of Medicine and Health and the Faculty of Arts & Social Sciences with a relative increase of 155% and 230%, respectively, when comparing the number of attendees between 2019 and 2021.

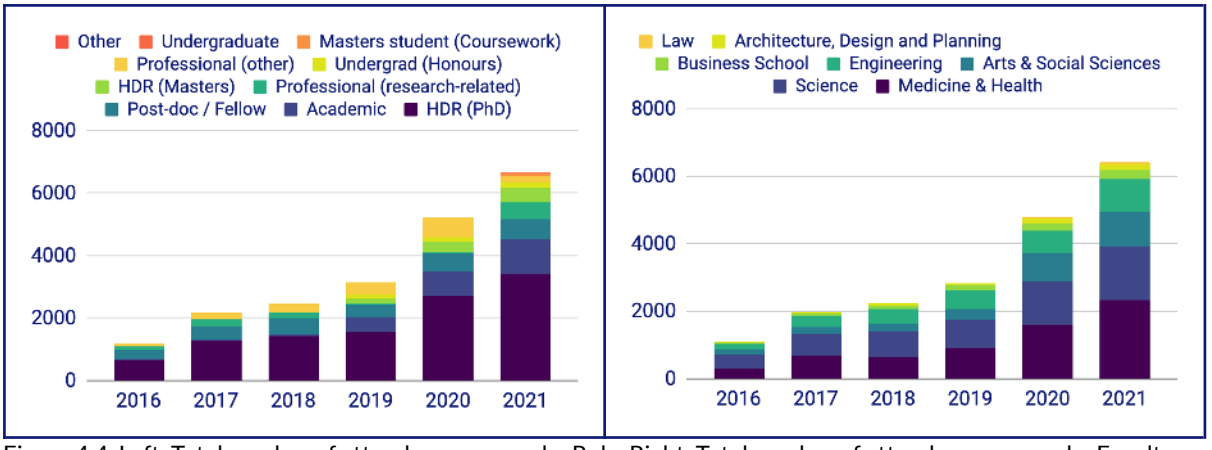


Figure 4.4: Left: Total number of attendees per year by Role. Right: Total number of attendees per year by Faculty.

Figure 4.5 displays the number 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, which continues to date. Surveying tools such as REDCap and Qualtrics courses are becoming increasingly popular, and in particular after the beginning of the pandemic in 2020. The introduction of NVivo courses in 2020 made

[7] The Faculty information is not captured directly from participants but instead is mapped based on the Field of Research (FoR) code that is captured in the registration form.

this tool very popular especially among researchers from the Faculty of Arts & Social Sciences, with NVivo becoming the fourth most popular tool taught in 2021.

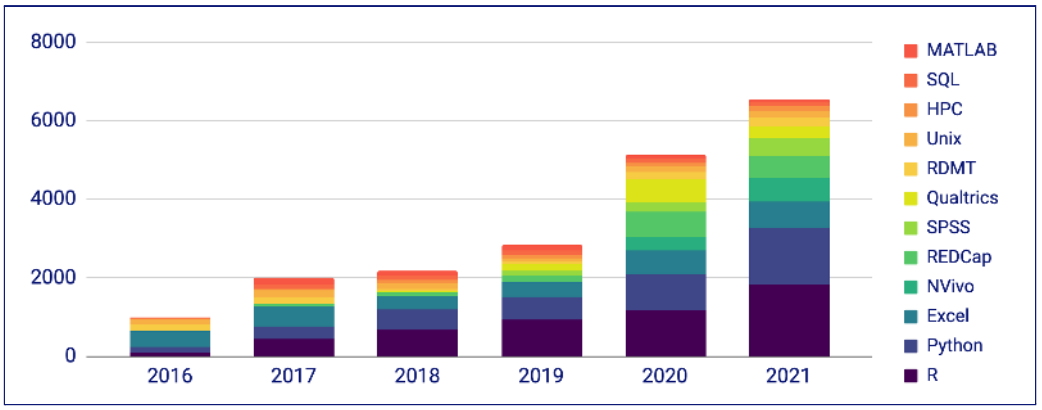


Figure 4.5: Total number of attendees per year by Tool/Technology.

Intersect operates a comprehensive, data-driven quality control process to ensure robust, quality delivery of training, including a course evaluation survey at the end of each course and the widely used industry metric of Net-Promoter Score (NPS). Intersect's annual NPS has increased rapidly over the years, from +46 in 2017 (when we started capturing NPS) to +75 in 2021. This demonstrates our commitment to the improvement of training quality and providing the best possible training experience to participants. Furthermore, this also suggests that moving our entire course catalogue and delivery online did not compromise the quality of delivery.

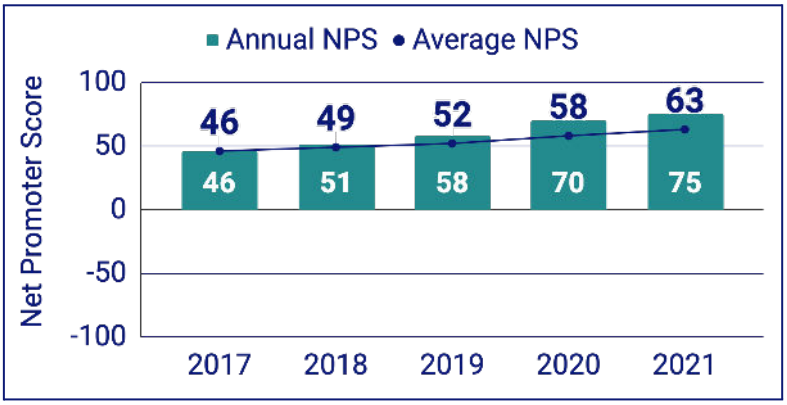


Figure 4.6: Average and Annual NPS per year since 2017.

In Intersect's course evaluation survey, five other primary quality metrics have been captured on a scale from 0 (worst) to 10 (best) since 2020. As shown in Figure 4.7, all primary quality metrics exceed 9.3 and 9.4 out of 10 in 2020 and 2021, respectively, demonstrating a very high quality training delivery from very knowledgeable and experienced instructors.

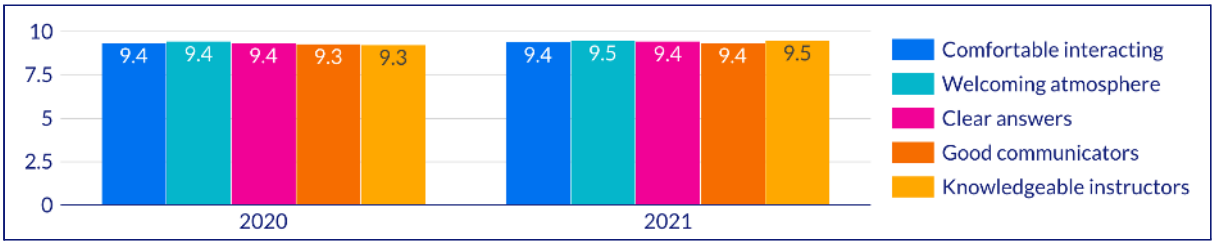


Figure 4.7: Primary metrics for measuring the quality of the training delivery per year since 2020.

Appendix 1

Table 1: List of Intersect training courses as in February 2022.

Course Name	Course Code	Duration (Days)	Tool/Technology being taught
Excel for Researchers	EXCEL101	1	Excel
Version Control with Git	GIT101	0.5	Git
Getting started with HPC using PBS Pro	HPC201	1	HPC
Getting started with HPC using Slurm	HPC202	1	HPC
Parallel Programming for HPC	HPC301	1	HPC
Learn to Program: Julia	JULIA101	0.5	Julia
Beyond the Basics: Julia	JULIA201	0.5	Julia
Learn to Program: MATLAB	MATLAB101	1	MATLAB
Getting started with NVivo for Windows	NVIVO101	0.5	NVivo
Getting Started with NVivo for Mac	NVIVO102	0.5	NVivo
Learn to Program: Python	PYTHON101	1	Python
Python for Research	PYTHON110	0.5	Python
Data Manipulation in Python	PYTHON201	0.5	Python
Data Visualisation in Python	PYTHON202	0.5	Python
Data Manipulation and Visualisation in Python	PYTHON203	1	Python
Introduction to Machine Learning using Python: Introduction & Linear Regression	PYTHON205	1	Python
Introduction to Machine Learning using Python: Classification	PYTHON206	1	Python
Introduction to Machine Learning using Python: SVM & Unsupervised Learning	PYTHON207	0.5	Python
Surveying with Qualtrics	QLTRICS101	0.5	Qualtrics
Learn to Program: R	R101	1	R
R for Social Scientists	R103	1	R
R for Research	R110	0.5	R
Data Manipulation in R	R201	0.5	R
Data Visualisation in R	R202	0.5	R
Data Manipulation and Visualisation in R	R203	1	R
Introduction to Machine Learning using R: Introduction & Linear Regression	R205	1	R
Introduction to Machine Learning using R: Classification	R206	1	R
Introduction to Machine Learning using R: SVM & Unsupervised Learning	R207	0.5	R
Exploring Chi-square and correlation in R	R210	0.5	R
Traversing t tests in R	R211	0.5	R
Exploring ANOVAs in R	R212	0.5	R
Research Data Management Techniques	RDMT001	0.5	RDMT

Data Capture and Surveys with REDCap	REDCAP101	0.5	REDCap
Longitudinal Trials with REDCap	REDCAP201	0.5	REDCap
Cleaning Data with Open Refine	REFINE101	0.5	Open Refine
Mastering text with Regular Expressions	REGEX101	0.5	Regexes
Regular Expressions on the Command Line	REGEX201	0.5	Regexes
Software Carpentry (Python)	SC101	2	Python, Git, Unix
Software Carpentry (R)	SC102	2	R, Git, Unix
Data Entry and Processing in SPSS	SPSS101	1	SPSS
Exploring Chi-Square and correlation in SPSS	SPSS102	0.5	SPSS
Databases and SQL	SQL101	0.5	SQL
Getting Started with Tableau for Data Analysis and Visualisation	TABLEAU101	0.5	Tableau
Unix Shell and Command Line Basics	UNIX101	0.5	Unix
Collecting Web Data	WEBDATA201	1	Python