



Paper TT01

Experiences of Training Teams in R

Alanah Jonas, GSK, London, United Kingdom

Abstract:

Several years ago, GSK made the decision, like other pharma companies to begin to transition away from SAS and towards R. In this presentation, I will discuss the learnings from the various ways in which R training has been given at GSK during this transition period. Some of these methods include large in-person and virtual classes, creating documentation and online learning courses, and small tailored study team sessions as teams are about to start programming in R. I will discuss the issues that people have faced from the different methods, the benefits of each of the methods and the ways we have adapted training to help people and teams to thrive.

Introduction:

In this paper I will explore the evolution of R training methods over the past few years. During my time at GSK, I have had first-hand experience in both delivering and participating in various R training methods. Overall, the most important thing to note is that it is necessary to have a multi-faceted approach for a successful training program to maximize participants' learning and success. The report will detail the different methods of R training and the journey of delivering R training over the years, highlighting the learnings from each method and the adaptations we have made along the way. I will also talk through how we have supplemented our core training program with additional methods to give participants more opportunities to learn and practice programming in R.

Training Methods:

In this section I will detail the core R training methods we have used at GSK. These being In-person classes, Virtual classes, Documentation guides and AccelerateR- Small, Tailored Sessions.

Initially when R adoption started out, we began by delivering large in person classes across various sites, where there would be an instructor and additional helper for assistance with exercises. This training would run as full days training. I will detail the pros and cons of this method below.

1. In-Person:



Pros:

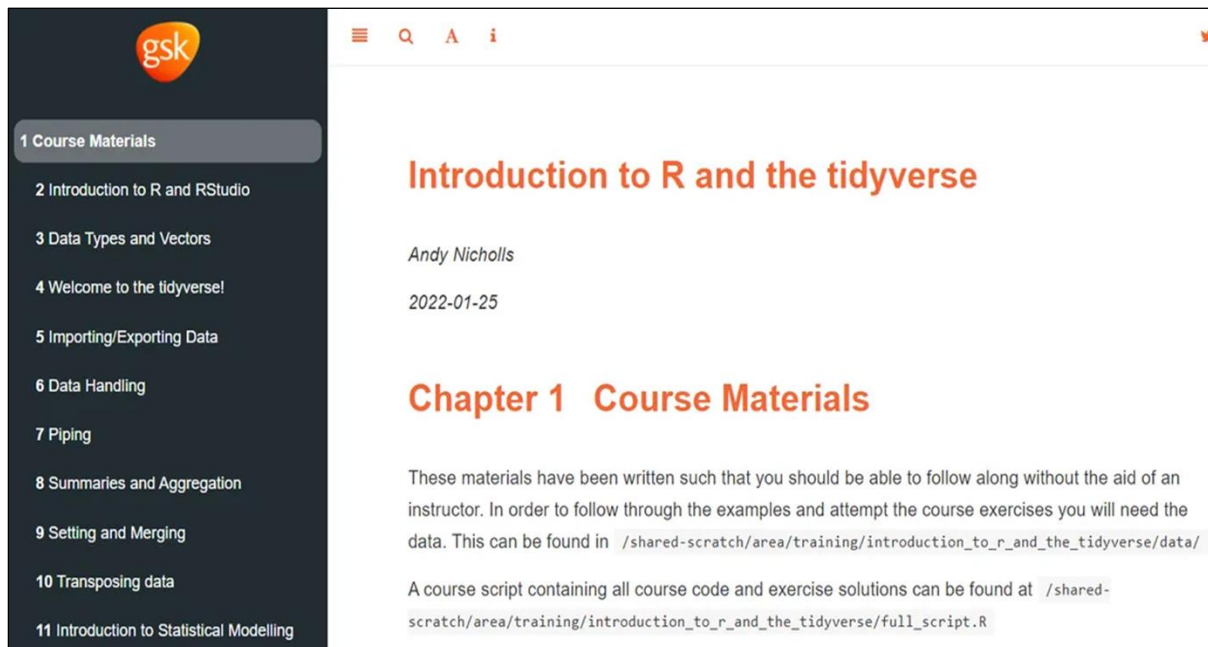
- **Focused Learning:** Participants can dedicate their attention to the training and hands-on exercises, as they are away from their normal work and feel more encouraged to focus due to being in a classroom.
- **Immediate Assistance:** Instructors and assistants are present in the room to provide help and address questions on the spot, to help with exercises or any questions on the materials.

Cons:

- **Pandemic Challenges:** On-site attendance has decreased since the pandemic and we have seen a significant increase in remote working. Further during the pandemic, we could not run in person classes, so we had to adapt.
- **Global Coordination:** As GSK is a large global company with a Biostats presence across the globe we need to conduct these classes across many sites. To run classes everywhere would be logistically hard to arrange, expensive and time consuming. Making it not possible to run a course in every location.
- **Timed Difficulties:** Participants and trainers will have a lot of other work and life commitments, so it is impossible to run the in-person sessions at a time to suit everyone's availability.
- **Knowledge Erosion:** If participants are not extensively using R, they will most likely forget the acquired knowledge, so it's hard to run the courses at a time for everyone when they are going to have the opportunities to program in R.

To run the in-person training classes R markdown documentation guides were developed on topics such as tidyverse and ggplot2. These guides also work well as an independent learning tool that participants can learn from and reference whenever they wish.

2. Documentation Guides:



Pros:

- **On-Demand Access:** The training materials are accessible at any time, providing great flexibility for learners.
- **Timely Reference:** The materials are available whenever a learner wants to learn or use R. They can reference the materials as a guide when they are programming in R.
- **Efficient Search:** It is easy to use the search functionality in these guidance documents to search for specific information or examples.
- **Guidance through Examples:** There are lots of examples within the documentation which provide an excellent starting point for learners.

Cons:

- **Lack of Interactivity:** By learning through these guides there is no real-time support from trainers when a learner requires assistance or has questions.
- **Engagement Challenge:** There is limited engagement and interaction with others compared to live sessions, by learning from guides.
- **Incomplete Coverage:** Not all topics will be covered in the documentation, a learner might be looking for a specific topic which may not be covered in the guidance document. If in a live session they could ask the trainer questions if you want to know something that has not

been covered, but also can reach out to trainers for support to or look for external resources on the topic.

- **Resource-Intensive:** Creating and maintaining the content demands significant time investment.

During the pandemic it was obviously no longer possible to run in-person training classes, so we adapted our r training program to be virtual. We split the training up into multiple sessions rather than being an entire day of training and had different trainers and assistants for each training block, to create less burden on the trainers. Going virtual allowed for more flexibility in attendance, for more sessions to be run and for the training to be attended by more people. This approach did have some great benefits as well as draw backs which I will discuss below.

3. Virtual:

Pros:

- **Scale and Reach:** Virtual training classes enable many participants to be trained simultaneously, there are no room capacity limits.
- **Pandemic Adaptability:** Virtual classes were a fantastic way to continue to train people throughout the pandemic to enable r adoption to continue to progress.
- **Geographical Flexibility:** Attendees can join from anywhere across the globe, dependent on suitable times, breaking location barriers.
- **Replay Capability:** Virtual Content could be recorded and watched later, so say a learner could not make one session they can watch the recorded one in their own time to catch up on what was missed. Or the whole series can be watched independently.

Cons:

- **Engagement Challenges:** Trainers often found there was reduced interactive engagement and fewer questions from participants compared to in person classes.
- **Distractions Concern:** Difficulty in ensuring undivided focus due to multitasking, it was hard to know if participants were engaged with the sessions as you cannot see their faces, they could easily be distracted by other things on their laptops.
- **Exercise Participation:** Similarly, it is hard to ascertain if participants complete exercises, as you cannot go physically round the room and see if people have questions, participants will often use exercise time as a break rather than actively completing the exercises.
- **Resource Intensiveness:** There was a lot of work involved with organizing and splitting up the content with the many trainers.

- **Scheduling Conflicts:** Similarly, to in-person classes participants might well find it tough to attend sessions amidst other commitments, particularly as the course was spread across several sessions.
- **Adaptation Difficulty:** As we were running such large classes there was limited flexibility in adjusting the course to different team needs.

GSK was finding that transitions to R were slow, not all study teams adopt and transition to R at the same time. Often meaning retraining of R is required. Therefore, the above methods were not that successful in helping teams to adopt R. This then led to the creation of AccelerateR, where we would focus on just in time training for study teams with support and mentoring. Virtual small group training aligned to the team's knowledge and experience of R is delivered. These sessions are recorded, and documentation shared to be accessed when the study team need. Data scientist and programmers are available to provide mentorship and support, through methods such as drop-in sessions and code reviews.

4. AccelerateR - Small, Tailored Sessions:

Pros:

- **Timely Knowledge:** Conducted when teams begin working in R, ensuring the R knowledge team members have acquired will be applied to programming tasks.
- **Customized Training:** Can Adapt training to individual team needs based on prior knowledge and experience with R.
- **On-Demand Assistance:** Additionally on demand support is available for team members so they always have someone to reach out to for guidance.
- **Personalized Interaction:** Small group setting with team members creates a comfortable environment for learning and asking questions.
- **Replay Capability:** Training sessions are recorded so that if a learner could not make a session, they can watch the recorded one in their own time to catch up on what was missed. Or all the training can be rewatched.

Cons:

- **Organizational Load:** There is an increased demand on organizers/trainers due to frequent and unique sessions.

Supplementary Learning Methods:

In addition to these four core R training methods that have been used at GSK a variety of supplemental training methods are also used. These provide a wonderful way for learners to further develop their R skills and practice them on things like mini challenges and to interact and learn from

other R users outside their normal teams. The three methods that I will discuss are R Code Corners - Small bi-monthly R challenges, Hackathons and R Book Clubs.

First, we have R Code corners which are small bi-monthly challenges based on the algorithms and logic that are used in day-to-day programming. For example, some of the challenges have involved creating an ADAM dataset using admiral, creating graphics or displays. The challenges are designed to build up R expertise in Biostatistics bridging the gap between the R training methods discussed earlier and its application to clinical reporting.

1. R Code Corners - Small bi-monthly R challenges:

JUNE'S TRAILBLAZERS

Asha Chakma
Liam Hobby
Kedar P
Ajay K Shrestha
Bharathi Rajendran
Sreeguna Chinnapareddygar
Nikhil Menon

Bhavani Surakanti
Lionel Panneel
Oliver Dean

R CODE CORNERS

Alanah Jonas
Cristina Pollari
Ajay Kumar Dampalapally
Bishal Banerjee
Sujay J Doss
Loic François
Madhukar Gawade

Jason Borg
Scott Worrell
Kevin Hudson
G Gayatri

FINDINGS

- Use of appropriate {admiral} functions!
- ADVS dataset created and bonus questions attempted successfully!
- Using factors or `scale_x_discrete()` to maintain visit order

LEARNINGS

- Use of {admiral} functions for deriving ADT, ADY and BMI to create advs dataset
- Exploration of advs data by calculating mean BMI per treatment group and plotting the distribution of weight by visit for each treatment group

QUICK TIPS

- Please fill the required details in the GSK header (v1.0) like the program name with a .R extension, input and output parameters and change log
- Include key variables in the `derive_param_bmi()` function to avoid returning NA values.
- Use `exprs()` instead of `vars()` as `vars()` has been deprecated.
- Use factor levels in `ggplot` to arrange visits in chronological order.

Pros:

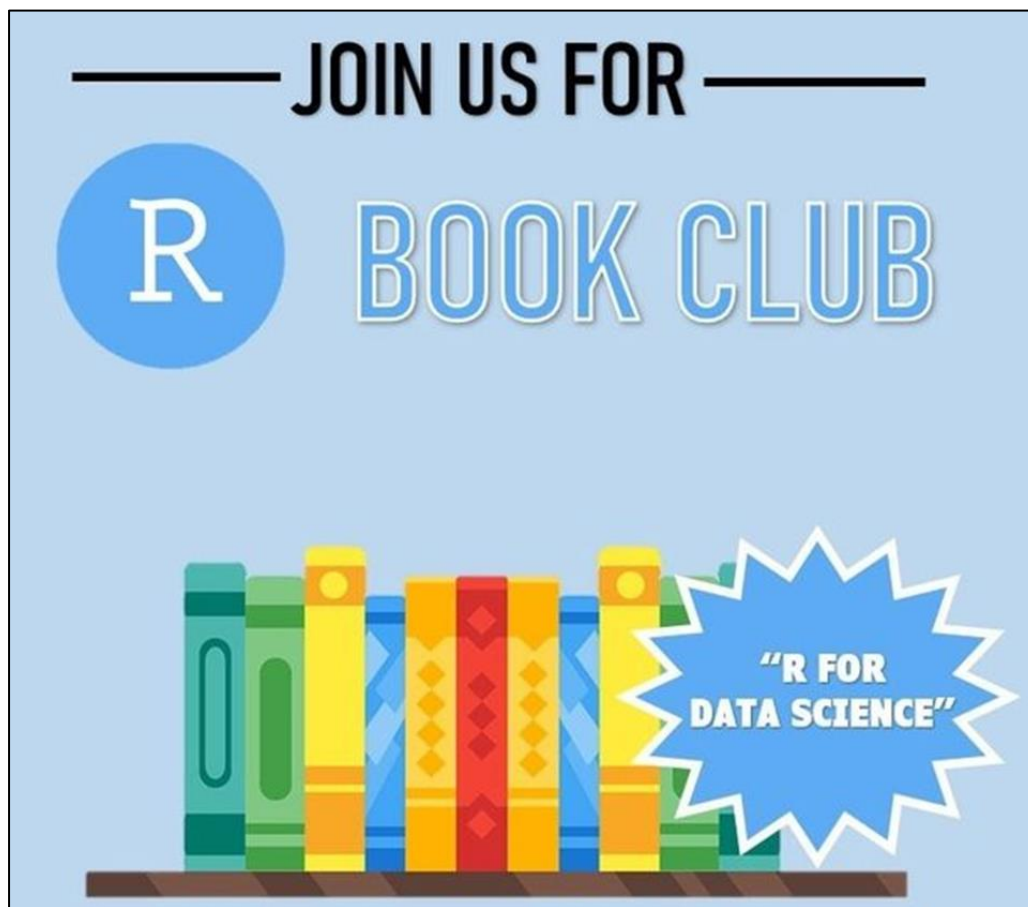
- **Effective introduction to applying R:** R code corners provides a fantastic opportunity for beginners to ease into programming in R, allowing learners to gain confidence before using R in their studies.
- **Guided Learning:** The challenges are broken down into easy-to-follow steps. The learners benefit from this guidance making the tasks easier to complete ultimately boosting confidence when using R in real work situation without this guidance written out.
- **Celebrating Achievements:** Best coders are acknowledged through workplace posts, appreciating achievement and dedication to the challenges. Learning and tips from each challenge are also shared on workplace meaning those who complete the challenge can learn what they could have done better or what they might have missed when completing the challenge. Also gives awareness of what the challenge involved for those who did not complete it.
- **On-demand access:** old challenges are stored online on a shiny app meaning they can be accessed and completed later. Particularly useful if a learner could not complete the challenge when it was released.

Cons:

- **Limited Engagement:** The challenge completion rate remains relatively low, with a consistent group of people completing each challenge.
- **Resource-Intensive:** Planning, creating, and marking challenges demand a substantial investment of time and effort from organizers. Also becomes harder to come up with new ideas for challenges as time goes on.

R Book clubs came about at GSK inspired by Maya Gans' presentation at RStudio::Global 2021 about her own Book Club. The book club initially focused on R for Data Science (Wickham & Grolemund, R for Data Science, 2016) before expanding to other books such as R Packages (Wickham & Bryan, R packages, 2015) and Mastering Shiny (Wickham, Mastering Shiny, 2021). The club involved weekly meetings after reading a section of the book where a nominated member would lead the session each week. The meetings were completely open to interpretation, open to short presentations or code walkthroughs for example. After this, the time is allocated to open discussion on any issues experienced within the chapter, interesting code snippets, and just general conversation about the topic.

2. R Book Clubs:



Pros:

- **Focused Group Learning:** Offers a valuable avenue for a collective dive into learning R skills, resulting in substantial learnings.
- **Organizational Simplicity:** Requires minimal preparation from organizers.

Cons:

- **Substantial Commitment:** Attendees face a significant commitment in reading a chapter weekly and completing associated exercises.
- **Varied Engagement:** Personal benefit varied from week to week and peaked when leading the chapter discussion and creating the presentation, impacting the value gained from different weeks.
- **Discussion Limitations:** Despite its potential, the depth of chapter discussions during meetings was limited.

Finally, we have Hackathons which are a fantastic way to develop digital and analytical skills alongside building agile ways of working and connecting with other people across the business.

3. Hackathons:

GSK

DataLabs

Introduces... Hackathons!

DataLabs will be back on 14th November and this year, we're introducing not one, but **two virtual hackathons!**

Would you like to get hands on experience using Excel, PowerBI or SQL? Or would you prefer wrangling, modelling, and visualising R&D datasets using R or Python?

Take your pick between the **D&A Hackathon** for regular users of D&A and the **Data Science Challenge** for D&A experts.

This is your chance to **develop your D&A skills** whilst working with others from different business units in a **fun environment**... there will be prizes!

Sign-Up here! →

Part of the **GSK Data Academy**

Pros:

- **Focused Skill Development:** An excellent platform for honing R and other data science skills through dedicated work on new projects.
- **Collaborative Learning:** Collaborate with individuals beyond your usual circles, fostering diverse perspectives and learning from others skill sets. Hackathons also create energy and excitement around r through solving interesting and different problems.

Cons:

- **Time Commitment Constraint:** Participating in a hackathon involves dedicating a significant amount of time making it unfeasible for some.
- **No follow-up support:** once the hackathon is over that is it there doesn't tend to be any after activities or continuation of what you've learnt from the challenge.
- **Intimidating and highly demanding:** Hackathons can be intimidating and highly demanding for beginners or those with less experience programming in R as there will potentially be a lot of pressure to complete quickly tasks due to the short nature of the challenge. So does not give much time for learning.

Conclusion:

In conclusion a multi-faceted training approach is important for a successful R training program. Different methods cater to different learners' preferences. Application of learned skills is vital for knowledge retention. You need a training foundation, but then it is about various forms of resource delivered when it matters. Tailored small study team classes just before beginning programming begins have worked the best for us, but the supplementary methods are crucial for teams not yet engaged in R programming. The success of a comprehensive R training program hinges on having a range of methods available to suit learners various learning needs and applications. Training alone does not work.

References:

Wickham, H. (2021). *Mastering Shiny*. Retrieved from <https://mastering-shiny.org/>

Wickham, H., & Bryan, J. (2015). *R packages*. Retrieved from <https://r-pkgs.org/>

Wickham, H., & Grolemund, G. (2016). *R for Data Science*. Retrieved from <https://r4ds.hadley.nz/>

Contact information:

Your comments and questions are valued and encouraged. Contact the author at:

Alanah Jonas

GSK House

980 Great West Rd

London

TW8 9GS

Email: Alanah.x.jonas@gsk.com

Brand and product names are trademarks of their respective companies.