



Co-funded by
the European Union



STRCUU Scientific Data Management Training for Post-Graduate Students and Young Researchers Using R Programming Language

A virtual Training : 24-26 February 2026

Resource Persons

Dr. Thomas Lapaka Odong(Thomas.l.odong@gmail.com)

Dr. Susan Balaba Tumwebaze (susantumwebaze@gmail.com)

Dr. Hellen Namaweje (hnamaweje@gmail.com)

Funded by the European Union under Grant No. 101237180

(ERASMUS -EDU-2025-CBHE-STRAND 1)



Co-funded by
the European Union



Project Title: Strengthening Teaching and Research Capacities in Recently Established Uganda Universities to Deliver Climate Resilience and Green Energy Solutions to Farming Communities

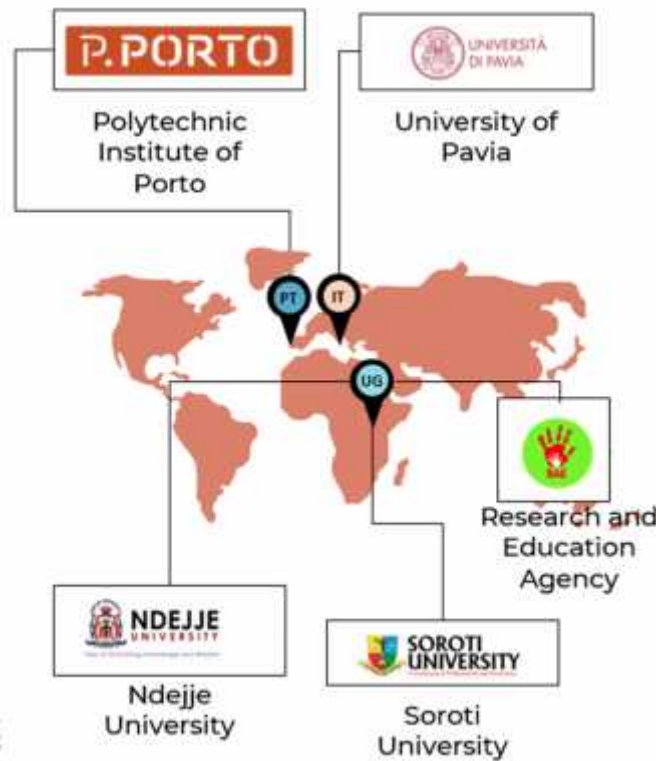
Acronym: STRCUU

Grant Call: ERASMUS -EDU-2025-CBHE-STRAND 1

Grant Number: 101237180

Lead Institution: Ndejje University, Uganda

Partner institutions



The European Commission's endorsement of
cannot be held responsible for any use which may be made of the information contained therein.

not constitute an
the Commission





Co-funded by
the European Union



Executive Summary

STRCUU Scientific Data Management Training for Postgraduate Students and Young Researchers (24–26 February 2026)

The Scientific Data Management Training was organized under the Strengthening Teaching and Research Capacities in Recently Established Uganda Universities (STRCUU) project, funded by the European Union through ERASMUS-EDU-2025-CBHE-STRAND 1 (Grant No. 101237180). The project is led by Ndejje University and aims to strengthen teaching and research capacities in recently established Ugandan universities to deliver climate resilience and green energy solutions to farming communities.

The three-day virtual training (14:00–18:00 EAT) targeted postgraduate students and young researchers across agriculture, environmental sciences, economics, and related disciplines. The course focused on enhancing competencies in scientific data management and statistical analysis using the R programming language. Modules covered: (1) Introduction to R and data management, (2) Exploratory Data Analysis and reporting with R Markdown, (3) Univariate parametric and non-parametric statistical inference, and (4) Simple and multiple linear regression.

The training recorded remarkable regional and international interest, with 841 registered participants from 43 countries and 468 active attendees from 28 countries (56% attendance rate). Participation was highest from Kenya, Ethiopia, Mozambique, and Uganda. Among evaluation respondents (N=106), 73.6% were male and 26.4% female.

Quantitative evaluation findings indicate substantial learning gains. Over 70% of participants reported strong improvement in installing R and importing datasets, while more than half reported strong improvement in

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



regression, parametric, and non-parametric tests. Instructional clarity was rated positively by 90.6% of respondents, and 89.6% expressed overall satisfaction with the training. Additionally, 73.6% reported feeling confident or very confident in using R after the training.

However, challenges were identified. Internet connectivity (31.4%) limited prior knowledge (26.7%), and pace of instruction (19%) were the main barriers. Qualitative feedback revealed strong demand for extended duration (44%), increased practical exercises (21%), advanced modules (14%), and post-training support mechanisms.

Overall, the training successfully strengthened foundational competencies in R programming and scientific data management. The high satisfaction levels and demand for advanced follow-up sessions demonstrate strong engagement and institutional relevance. Strategic recommendations include extending training duration to 5–7 days, increasing hands-on practice, introducing beginner and intermediate tracks, and institutionalizing an R community of practice under the STRCUU framework.

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



Contents

1.0 Background.....	7
1.1 Target audience	9
1.2 Aim	10
1.3 Specific Objectives	10
1.4 Course Outcomes (Expectations)	11
1.5 Delivery Method and Requirements	12
1. 6 Training Pre-requisite	13
1.7 Duration	14
2.0 An overview training content	14
2.1 Module 1- Introduction to R programming language and data management	15
2.2 Module 2- Exploratory analysis and reporting with R Markdown	15
2.3 Module 3- Univariate parametric and non-parametric statistical inference with R.....	16
2.4 Module 4- Linear regressions in R (simple and multiple)	16
3.0 An overview of course outcomes.....	16
3.1 Outcomes for Module 1 on introduction to R programming language and data management	17
3.2 Outcomes for Module 2 on exploratory analysis and reporting with R Markdown	17
3.3 Outcomes of module 3 on univariate parametric and non-parametric statistical inference.....	18
3.4 Outcomes of module 4 on linear regression (simple and multiple)	18

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



4.0 Opening remarks	19
5.0 Evaluation feedback for training 24 th -26 th February 2026	21
5.1 The analytical, interpretive, and strategic sections	29
5.1.2 Part A: Method employed	29
5.1.3 Part B: Some of the representative quotes from participants	30
5.1.4 Part C: Cross-cutting patterns (High-level analysis)	30
5.1.5 Part D: Priority ranking	31
5.1.6 Part E: Actionable recommendations.....	31
5.1.7 Part F: Limitations of the analysis.....	32
5.1.8 Part G: Strategic opportunities from the analysis	32
5.1.9 Proposed Topics for the Follow-Up R Training (April/May/June).....	33
6.0 Closing remarks	39
7.0 Analytical report for scientific data management.....	39
8.0 Links used in the training.....	44
8.1 Google Drive link with training materials	44
8.2 Overall evaluation questionnaire link	44
9.0 Annexes.....	45
9.1 Annex 1: Timelines.....	45
9.2: Annex 2: Training Program.....	45
9.3 Annex 3: Training needs assessment tool.....	46

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



1.0 Background

Scientific data management is a set of practices, procedures, techniques, and tools that enable scientists, in particular PhD, MSc students and young researchers, to manage data rationally in order to produce quality scientific reports. In the same way, its good mastery enhances their ability to meaningfully engage in conducting quality research by developing appropriate research proposals, designing of studies, collecting and analysing data. The purpose is to preserve data in a consistent, accessible, secure and uncluttered form. Without data management, data analysis will produce incorrect results and lead to biased decision making. Indeed, it is currently observed with the technological development and the concern to have more precise and accurate results, experiments or surveys are performed on a large scale sometimes leading to complex designs and to subsequent messy data. Figuring out how to handle data resulting from such experiments/surveys takes time, and getting appropriate assistance is difficult. The students also do not know how to effectively analyze the data using appropriate statistical software, interpret the results, and communicate properly with the target audience. Given these shortcomings, this training is structured to cover the general needs of biometrics and to provide post-graduate students and young researchers with the skills and abilities to conduct their research effectively and

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



efficiently. The content of the modules in this training focuses on the techniques of processing, exploration, inferential analysis and modelling of data (qualitative and quantitative) most commonly encountered in the literature to solve real-life problems. Students and young researchers will also be exposed to the R programming language for data management, analysis and reporting.

R is a free, open source statistical programming language primarily used by statisticians and data miners. According to the PYPL index in 2020, R is popularly ranked 7th among scholarly users worldwide (<https://dairy.solutions/the-most-popular-programming-languages-in-2021/>). It is a very flexible language in performing tasks and anyone

interested in data analysis and any user can quickly learn it, whether they are a data scientist or not. R is accessible to a large number of persons and meets a need of students and researchers in developing countries, most of whom cannot afford to spend large sums of money each year to renew licenses for commercially available statistical software.

In data science with applications in many areas of life such as agriculture, biology, health; environment, economics, insurance, bioinformatics, statistics, etc., R plays a crucial role in pre-processing, exploration, and visualization of data (which can be massive, varied, structured or

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



unstructured), prediction, classification, and clustering thanks to its multiple and diversified functionalities. A user of R can modify the various functions of R and create his own packages or either, he can execute code without a compiler, do many calculations with vectors. It can also be used to develop amazing web applications, has a large community support through bootcamps and R meetings. It is well maintained and R updates are always available on CRAN. Note that R is released under the GNU General Public License and there are no restrictions on its use. As for RStudio, it is an integrated development environment (IDE) for R, which also supports statistical computing, graphics and statistical models. An RStudio user can manipulate data and may be able to store used R commands for future use. The environment also provides an R markdown feature that allows work to be converted into different formats such as Word, PDF, PowerPoint, HTML, etc. This is a boon for academics, as scholarly papers can be written directly in the R markdown environment and then published in a manuscript.

1.1 Target audience

The training is addressed to postgraduate students and young researchers in the following fields: Agricultural Economics, Agroforestry, Plant Protection,

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



Food Science, Aquaculture and Fisheries Science, Natural Resource Management and Environmental Science.

1.2 Aim

The objective of this training is to enhance the efficiency of the flow of agricultural information and research, in particular to achieve the following objectives: to understand the different procedures related to the design of experimental/survey schemes; to implement correctly the different statistical techniques at all stages of research and to present the results in a coherent and efficient way. It will enable postgraduate students and young researchers to acquire not only the technical skills and knowledge necessary to use R programming language for data management, analysis, and presentation of results in a format that ensures their wide dissemination as peer-reviewed publications and policy formulation, but also the practical skills they need to improve the quality of their research publications.

1.3 Specific Objectives

- a) Participants will be able to learn how to download R, RStudio, R packages and install them on their computer.
- b) Participants will be familiar with the techniques of importing data sets in different formats such as Excel, Txt, CSV into R and exporting R data

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



sets to Excel, Txt and CSV.

- c) Participants will learn how to manipulate their field data in R before performing any data analysis using R's interactive command prompts.
- d) Participants will learn to describe data sets with descriptive statistics parameters (central tendency, dispersion and shape parameters) calculated with the R programming language.
- e) Participants will learn how to use R Markdown to store various R commands, write R scripts, add comments which they can then convert to pdf, HTML, Word document and Power Point. Participants will learn various effective data visualization techniques with the R programming language to produce quality graphs for use in manuscripts and technical reports. Participants will be able to use the R programming language to analyze data with inferential statistical tools such as univariate parametric and non-parametric statistical testing, categorical data analysis techniques and linear regression models.

1.4 Course Outcomes (Expectations)

At the end of the training participants would be able to:

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



- (i) Apply key statistical concepts such as exploratory analysis, univariate parametric and non-parametric inferential statistics, categorical data analysis techniques, simple and multiple linear regression models.
- (ii) Use the R programming language to manipulate, explore, analyse and model the state of a biological or agricultural system quantitatively and qualitatively in other fields such as economics and health.
- (iii) Select the appropriate statistical method for the problem and the data available.
- (iv) Interpret the results and graphs obtained and communicate them effectively and coherently

1.5 Delivery Method and Requirements

The delivery mode was mixed, with very little interactive theoretical underpinning and practical work designed to complement the lecture material for real-world problem-solving purposes. The approach was participatory, with students expected to be active learners and to engage in intensive and critical self-directed learning. The assignments were designed to train and test critical thinking skills. Real data sets provided by

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



the facilitators or obtained from the students before the course starts will be used in the examples and exercises. The training was conducted [online](#) and each participant required a laptop, a good connection and a dataset. The daily programme was divided into sections giving an overview of the topics, followed by practical computer exercises and a discussion of the statistical results. First, the basic principles, followed by examples of syntax in the R language was presented. The participants analysed their data using the techniques already introduced in their daily work. Discussions on the interpretation and presentation of results took place each day during the plenary sessions. Participants evaluated the modules on a daily basis and deficiencies were corrected immediately. An overall evaluation of the modules was conducted at the end of the training.

1. 6 Training Pre-requisite

This training builds on the knowledge acquired by participants during their postgraduate and undergraduate studies. It assumes understanding of basic biometrics applied to quantitative and qualitative data, and in addition, numeracy skills acquired overtime. The modules of this training provide a solid understanding of statistical techniques that relate to

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



quantitative/qualitative aspects from application, and analytical perspective, thus balancing between theory and applied concepts.

1.7 Duration

The training program span from **24-26 February 2026; 14:00-18:00 EAT**, allowing for comprehensive coverage of key concepts and ample hands-on practice.

2.0 An overview training content

The following modules cover the fundamental aspects of computer science, data management, analysis and processing, and presentation of results, which provided the framework for the training material. The presentation of these modules were alternated with lectures and illustrations of real-life applications, selected from various fields, agronomic, environmental, socio-economic and health. More emphasis was placed on the principle of statistical methods (univariate and bivariate descriptive statistics, univariate parametric and non-parametric statistical inference, and simple and multiple linear regression), when to use them and the rules for interpreting the results derived from them, rather than on technical or mathematical developments. The functions for implementing each

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



method were presented with an example using the R programming language. Each participant was able to implement the applications on data related to their field or those to be made available to them, followed by presentations.

2.1 Module 1- Introduction to R programming language and data management

Basics of R uses which include downloading R and R packages, arithmetic operations with R, assigning variables, basic data types in R, importing and exporting datasets, defining a directory, saving datasets in R and in excel sheets.

2.2 Module 2- Exploratory analysis and reporting with R Markdown

Types of data, Descriptive statistics and data visualization in R, such as histograms, scatter plots, box plots, and bar charts. With introduction to the production of a statistical report: Write a report including R code, the outputs of this code (results, graphs, tables) and comments with R Markdown whose documents are fully reproducible and obtainable in PDF, HTML, Word and PPT formats among others.

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



2.3 Module 3- Univariate parametric and non-parametric statistical inference with R

Univariate statistical tests (comparison of proportions and means and correlation tests – Parametric and non-parametric tests: all t-tests and ANOVA with their non parametric alternatives, chi-square tests, correlations tests [pearson, kendall, pearson, etc.]). Reporting results, interpreting statistical analyses, discussing limitations, and drawing valid conclusions.

2.4 Module 4- Linear regressions in R (simple and multiple)

Principle and scripts of simple and multiple regression models, estimation and significance test of coefficients in the model, validation of the model, analysis of variance table, coefficient of determination (R^2) and adjusted R^2 , Influential values, graphical and statistical analysis of residuals; correlations between explanatory variables; use of the model in forecasting.

3.0 An overview of course outcomes

Each of the modules to be covered during the training resulted into the following outcomes:

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



3.1 Outcomes for Module 1 on introduction to R programming language and data management

- Download and Installation of R, R studio, and other R packages.
- Define a directory working and create a script in R.
- Import different datasets for example from Excel, Txt or CVS into R and export R datasets into Excel, Txt or CSV.
- Write basic data types and assigning variables in R
- Manipulate a dataset by doing sort, order, divide, stack, subset, replace, insert, extract, etc.
- Perform arithmetic operations with R.
- Save a script or all the work performed in R.

3.2 Outcomes for Module 2 on exploratory analysis and reporting with R Markdown

- Describe the basic statistical terms
- Calculate descriptive statistics parameters (central tendency, dispersion and shape parameters) of a dataset with R
- Give the distribution of data according to the nature of the variables with R

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



- Do data visualization applied in R programming language as well as produce publication-quality graphs.
- Use R Markdown to store the different R commands, write R scripts, write comments and convert R Markdown into pdf, HTML, word document, and power point.

3.3 Outcomes of module 3 on univariate parametric and non-parametric statistical inference.

- Perform comparison of proportions tests with R (Z test, Binomial test, McNemar test, Cochran's Q test)
- Do comparison of means tests with R: Parametric tests- all t-tests and ANOVA with their non parametric alternatives (Wilcoxon signed rank test, Mann Whitney test and Kruskal Wallis test).
- Measure the association between two variables : correlations tests [pearson, kendall, pearson, etc.] and test of independence [chi-square, fisher exact, etc.]

3.4 Outcomes of module 4 on linear regression (simple and multiple)

- Distinguish correlation and regression analysis

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



- Recognize when to apply a regression analysis
- Build models using different scenarios of studies
- Graphical and statistical study of residues Validate final model of the analysis
- Analyse data using correlation and regression analysis and interpret results correctly
- Make predictions from established regression models

4.0 Opening remarks

Opening remarks by Prof. Adipala Ekwamu; Good afternoon distinguished facilitators, dear participants, representatives from RUFORUM, and our valued partners and funders from the European Union.

On behalf of the STRCUU project, I warmly welcome you to this important Scientific Data Management Training. This initiative reflects our shared commitment to strengthening research capacity in our universities and empowering postgraduate students and young researchers with practical analytical skills that directly contribute to climate resilience and sustainable agricultural transformation.

I would like to extend my sincere appreciation to our dedicated resource persons Dr. Thomas Lapaka Odong, Dr. Susan Balaba Tumwebaze, and Dr. Hellen Namaweije for designing and delivering this highly practical and impactful training. Your expertise and commitment are central to the success of this initiative.

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



To our participants joining from across Uganda, Africa, and beyond, your overwhelming response with over 800 registrations from 43 countries demonstrates the growing demand for data-driven research skills. Your commitment to investing time and effort in improving your analytical capacity is commendable. The future of research in our region depends on young scientists who can generate credible evidence and translate it into policy and practice.

I also wish to sincerely thank RUFORUM for providing the platform that enables collaboration and regional outreach, and to the European Union for the financial support under the ERASMUS capacity-building framework. Your investment is not only building skills but strengthening institutions and research ecosystems across our region.

Scientific data management is no longer optional; it is foundational. In an era of complex datasets, climate challenges, and increasing demand for accountability in research, mastery of tools such as R programming ensures transparency, reproducibility, and scientific excellence.

As we proceed, I encourage participants to engage actively, practice continuously, and view this not as a one-time training but as the beginning of a sustained learning journey. STRCUU envisions a continuous training series and a growing community of practice in data science and research methods.

I thank you all and wish you productive deliberations and transformative learning experiences.

Thank you.

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.



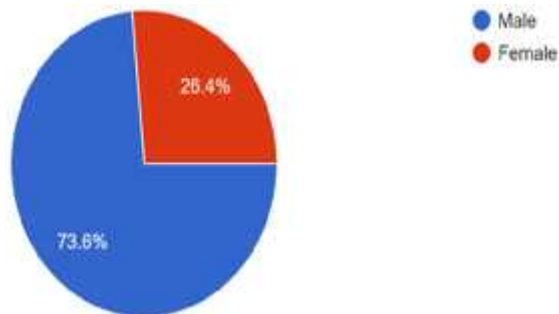


Co-funded by
the European Union



5.0 Evaluation feedback for training 24th-26th February 2026

Qn1: Sex of participants



Of the 106 participants that filled the questionnaire, the males were 78(73.6%) and females were 28(26.4%). This implies that the training had more males than females.

Qn2: Please indicate the level of improvement in the following topics or areas as result of participation in this training

Level of improvement	Not improved	Slightly improved	Improved a lot
Installing R and RStudio	5(4.72%)	22(20.75%)	79(74.53%)
Working with R environment	5(4.72%)	37(34.91%)	64(60.38%)
Basic commands in R	5 (4.72%)	46(43.40%)	55 (51.89%)

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



Importing datasets in R	5(4.72%)	25(23.58%)	76(71.70%)
Simple and multiple regression and correlation	10(9.43%)	39(36.79%)	57(53.77%)
Parametric tests	8(7.54%)	43(40.57%)	55(51.89%)
Non-Parametric tests	9(8.49%)	39(36.79%)	58(54.72%)
Exploratory data analysis	10(9.43%)	46(43.40%)	50(47.17%)

Participants (N = 106) rated their level of improvement in different R-related topics under three categories: that is, Not improved, Slightly improved, and Improved a lot.

The table shows that the majority of participants reported strong improvement ("Improved a lot") across most topics, that is,

The highest improvement areas

These areas had the highest percentage of "Improved a lot": Installing R & RStudio – 74.53%, Importing datasets – 71.70%, Working with R environment – 60.38%

Thus, participants gained strong foundational skills. These are beginner-friendly areas, meaning the training was very effective at introducing participants to R setup and data import.

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



The moderate Improvement Areas were Basic commands – 51.89%, Parametric tests – 51.89%, Non-parametric tests – 54.72% and Regression & correlation – 53.77%

This means that it is slightly lower than installation topics because these require statistical understanding, participants may have struggled with applying concepts, not just running code and some participants reported limited statistics background (see Qn3).

Also, the lowest Strong Improvement Area was Exploratory Data Analysis – 47.17%, this has the lowest “Improved a lot” percentage because EDA requires: interpretation skills, visualization understanding, statistical reasoning and confidence with multiple commands

Therefore, the training was highly effective in Basic setup, Data importing, Environment navigation, but slightly less strong in statistical modeling and analytical interpretation which aligns with Qn3 challenges (pace and prior knowledge).

Qn3: What was the main challenge you faced during the virtual R training?

Main challenge	Percentage
Internet connectivity issues	31.4%
Limited prior knowledge in using R programming language	26.7%

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



Difficulty following the coding demonstrations in real time	14.3%
Pace of training was too fast	19%
Others	8.6%

The topmost issue was internet connectivity (31.4%) because virtual training depends heavily on stable internet, participants likely came from different regions with varying connectivity, and coding demonstrations require live streaming clarity. The second issue was limited prior knowledge (26.7%) this was because many participants were beginners. This explains the requests for slower pace (Qn8), difficulty following live coding and need for extended training duration.

In addition, nearly 1 in 5 participants (19%) felt the training pace was too fast. This supports findings from Qn2 moderate improvement in advanced topics and Qn8 demand for longer duration. The last challenge was difficulty following live coding (14.3%). This may be due to internet delays, screen visibility issues, cognitive overload and lack of hands-on pauses

Therefore, challenges were mainly: technical (internet), skill-level related (beginner gap) and delivery-speed related.

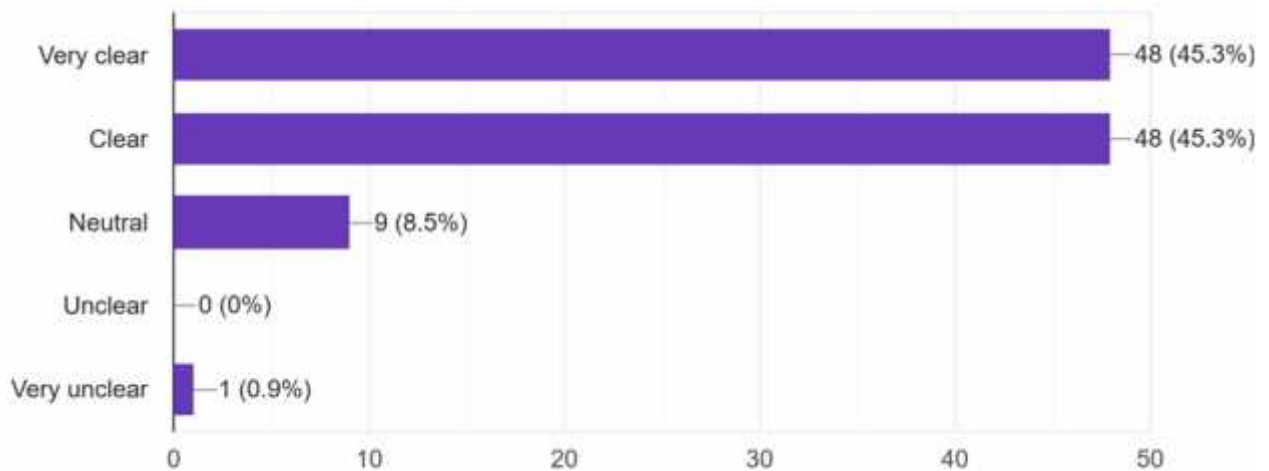
Qn4: How would you rate the clarity of instruction? (N=106 responses)

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



The results indicate that the training instruction was perceived as highly clear by participants. A combined 90.6% (45.3% “very clear” and 45.3% “clear”) rated the instruction positively, demonstrating strong effectiveness in content delivery. Only 8.5% remained neutral, and a very small proportion (0.9%) found the instruction “very unclear.” This suggests that the facilitators communicated concepts effectively and that clarity was not a major limitation of the training. The extremely low dissatisfaction rate reinforces that instructional quality was strong, even though some participants may have struggled due to other factors such as pace or prior knowledge rather than clarity itself.

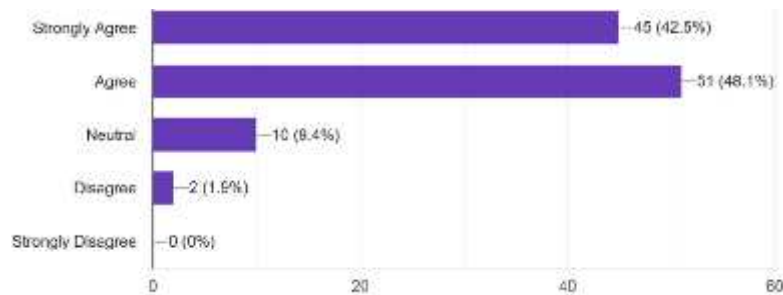
Qn5: How would you rate the hands-on exercises helped me understand the concepts (N=106 responses)

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



The findings show that practical exercises were highly valued by participants. A total of 90.6% either “strongly agreed” (42.5%) or “agreed” (48.1%) that the hands-on exercises helped them understand the concepts. Only 9.4% were neutral, and 1.9% disagreed. This indicates that the practical components significantly enhanced learning and comprehension. Although some participants later requested more practice time (as seen in the qualitative responses), the graph confirms that the exercises themselves were effective in supporting conceptual understanding.

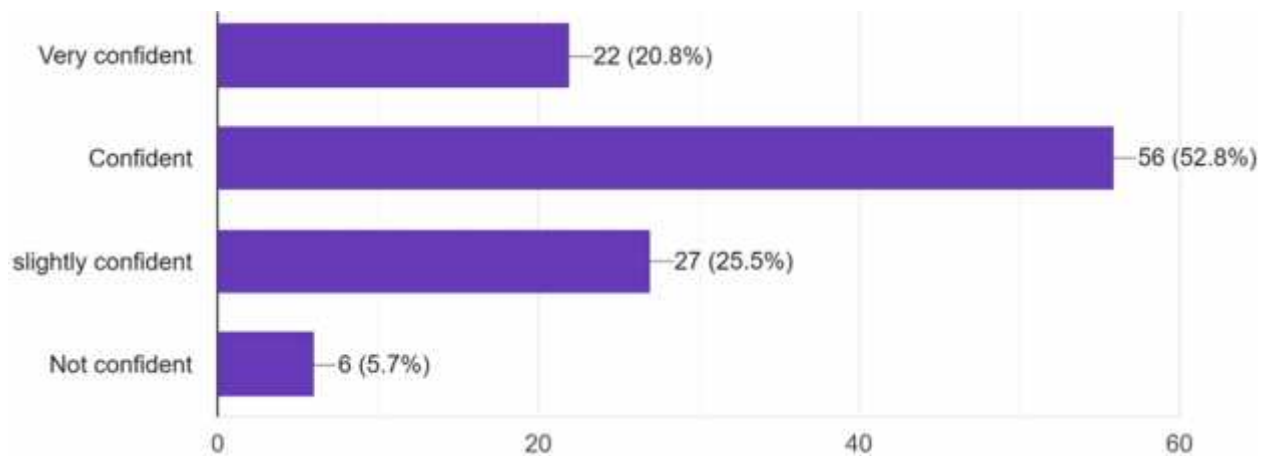
Qn6: After this training, how confident are you in using R? (N=106 responses)

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



The confidence levels reported after the training suggest a generally positive outcome, though with room for improvement. A majority of participants (73.6%) felt either “very confident” (20.8%) or “confident” (52.8%) in using R after the training. However, 25.5% reported being only “slightly confident,” and 5.7% indicated they were “not confident.” This distribution suggests that while the training successfully built confidence for most participants, a substantial minority may require additional practice, extended training time, or follow-up support to strengthen their mastery of R skills.

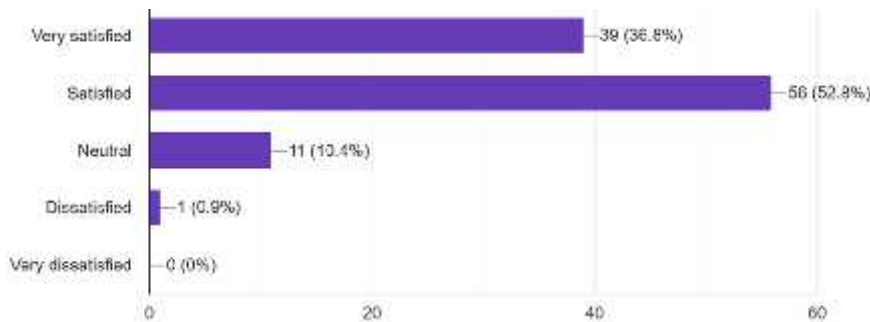
Qn7: Overall, how satisfied are you with this training? (N=106 responses)

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



Overall satisfaction with the training was very high. A combined 89.6% of participants reported being either “very satisfied” (36.8%) or “satisfied” (52.8%). Only 10.4% were neutral, and just 0.9% expressed dissatisfaction. These findings indicate that the training met or exceeded expectations for the vast majority of participants. The high satisfaction rate, together with strong clarity and confidence ratings, suggests that the program was well-received and impactful, even though participants recommended extending its duration and increasing practice opportunities.

Qn8: What is one suggestion to improve future R training?

Theme	Frequency (n=63)	Percentage (%)
Increase training duration/More time	28	44
More practical exercises	13	21
Advanced/Continuous training	9	14

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



Post-training support	7	11
Delivery method improvements	5	8
No suggestion/satisfied	6	10

The evaluation results indicate that the most significant concern (44%) was insufficient training duration, followed by a need for increased practical exercises (21%). These findings suggest that while the training content was valuable, participants require extended time and enhanced hands-on engagement to effectively master R programming skills. Additionally, demand for advanced modules and post-training support demonstrates strong participant interest in continued learning.

5.1 The analytical, interpretive, and strategic sections

The analytical, interpretive, and strategic sections that are useful for decision-making according to question 8 (the qualitative question)

5.1.2 Part A: Method employed

Responses to the open-ended question were analyzed using thematic analysis. Similar suggestions were grouped into broader categories, and frequencies were computed to identify dominant concerns. Each response was coded into its primary theme.

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



5.1.3 Part B: Some of the representative quotes from participants

These include:

Theme: Increase Duration

“R trainings need an extended time of at least a month...”
“It is too difficult to master R in a short period.”

Theme: More Practice

“Participants need to try their hands rather than run codes already written.”

5.1.4 Part C: Cross-cutting patterns (High-level analysis)

The following patterns were developed :

A. Learning Challenge Pattern

Many responses reflected:

-) Cognitive overload
-) Beginner difficulty
-) Pace challenges

Thus, a significant proportion of responses suggest that participants struggled with the pace and volume of content, indicating a need for differentiated instruction based on skill level.

B. Motivation Indicator

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



Requests for:

-) Advanced training
-) Continued sessions
-) Follow-up support

The demand for advanced and follow-up training suggests strong participant engagement and perceived value of the program.

5.1.5 Part D: Priority ranking

This table shows the rank themes by importance:

Priority level	Theme	Justification
High	Increase time	44% mentioned
Medium	More practice	21%
Medium	Advanced training	14%
Low	Delivery improvements	8%

5.1.6 Part E: Actionable recommendations

Short-Term Recommendations

-) Increase training duration to 5–7 days
-) Add structured exercises after each module
-) Share recordings immediately after sessions

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



Medium-Term Recommendations

-) Develop Beginner and Intermediate tracks
-) Create a WhatsApp or online support group
-) Organize monthly R clinics

Long-Term Recommendations

-) Institutionalizing an R training series
-) Introducing in-person regional workshops

5.1.7 Part F: Limitations of the analysis

The analysis reflects self-reported suggestions from participants and may not fully capture silent participants' perspectives. Additionally, responses were categorized under dominant themes, though some responses overlapped in multiple categories.

Despite suggestions for improvement, overall feedback was positive, with several participants expressing satisfaction and appreciation for the training.

5.1.8 Part G: Strategic opportunities from the analysis

-) The high demand for extended training indicates market need.

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



- J There is an Opportunity to create certificate-based programs.
- J Potential to build an R community of practice.
- J Strong regional interest (some mentioned Kenya, Africa).

Conclusion

The quantitative findings indicate that participants experienced substantial improvement across foundational R competencies, particularly in installation and data importation. However, challenges related to internet connectivity, limited prior knowledge, and pace of instruction affected some participants' learning experience. Despite these constraints, overall clarity, confidence, and satisfaction levels suggest that the training was effective, with recommendations focused primarily on extending duration and increasing hands-on engagement rather than altering content quality.

5.1.9 Proposed Topics for the Follow-Up R Training (April/May/June)

This section is derived from qualitative findings (Qn8 – suggestions), improvement gaps in Qn2, confidence levels in Qn6, and challenges reported in Qn3.

The proposed follow-up training is informed by both quantitative and qualitative evaluation findings. While participants reported strong improvement in foundational skills such as installing R and importing datasets, several areas showed comparatively lower improvement levels, including exploratory data analysis and advanced statistical techniques. Additionally, qualitative responses indicated a strong demand for

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



extended training duration, more practical exercises, advanced modules, and post-training support.

Based on these findings, the following thematic areas are recommended for the upcoming follow-up training.

1. Strengthening Foundations Through Extended Practical Sessions

Although installation and basic commands showed high improvement, 25.5% of participants reported being only slightly confident and 5.7% not confident in using R. This suggests the need to reinforce foundational competencies through intensive hands-on practice.

Recommended Topics:

-) Writing and structuring R scripts from scratch
-) Working directory management and project organization
-) Data cleaning and transformation (dplyr, tidyr)
-) Error troubleshooting and debugging common mistakes
-) Building reproducible workflows using RStudio projects

This module will emphasize “learn by doing” with guided exercises after each concept.

2. Exploratory Data Analysis (Advanced Application)

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



Exploratory Data Analysis (EDA) recorded the lowest “improved a lot” rating (47.17%), indicating a need for deeper engagement.

Recommended Topics:

-) Data visualization using ggplot2
-) Customizing plots and interpreting outputs
-) Identifying outliers and data patterns
-) Descriptive statistics automation
-) Reporting EDA results for research papers

Participants will practice using real datasets, including their own data where possible.

3. Advanced Statistical Modeling in R

Qualitative responses explicitly requested deeper and more advanced training (e.g., multivariate and geospatial analysis). Additionally, regression and parametric tests showed moderate improvement levels.

Recommended Topics:

-) Multiple regression diagnostics and assumptions
-) Logistic regression
-) Model comparison and interpretation

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



- J ANOVA (two-way and factorial designs)
- J Introduction to multivariate analysis (PCA, cluster analysis)
- J Introduction to mixed models (if time allows)

This module responds directly to participants' request for more advanced statistical applications.

4. Working with Participants' Own Data

Several participants suggested using their own datasets and receiving consultation support.

Recommended Component:

- J "Bring Your Own Data" practical sessions
- J Small group mentoring sessions
- J Live troubleshooting clinic
- J Data preparation for publication

This increases relevance and boosts post-training confidence.

5. Reproducible Research and Reporting

To move participants from coding to research output, the training will include:

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



- J Introduction to R Markdown
- J Generating automated reports
- J Exporting tables and figures
- J Basic data visualization for publication
- J Organizing scripts for research projects

This aligns with academic and professional research needs.

6. Introduction to Specialized Applications (Optional Advanced Module)

Based on qualitative suggestions:

- J Multivariate analysis
- J Geospatial analysis in R
- J R for social sciences applications
- J Agricultural data analysis (plant traits, soil parameters)

These modules could be offered as optional advanced tracks.

7. Post-Training Support Structure

Evaluation findings strongly recommended structured follow-up.

Proposed mechanisms:

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



- J Creation of a WhatsApp or online support group (already in place)
- J Monthly virtual R clinics
- J Sharing session recordings
- J Providing structured exercise booklets
- J Certification pathway (Beginner → Intermediate → Advanced)

Recommended Training Structure for May/June

Based on feedback requesting extended duration, the follow-up training could be structured as:

- J 5–7-day intensive practical workshop
or
- J 2-week modular training (weekend or evening sessions)

Each day should include:

- J 40% instruction
- J 60% hands-on practice

Strategic Justification

The evaluation findings indicate:

- J High satisfaction (89.6%)

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



-) Strong clarity ratings (90.6%)
-) Moderate-to-high confidence (73.6%)

However:

-) A significant proportion requested more time (44%)
-) More practice (21%)
-) Advanced modules (14%)

This demonstrates strong participant engagement and demand for a structured progression model. The proposed follow-up training therefore builds on strengths while addressing identified gaps in depth, application, and confidence.

6.0 Closing remarks

The closing remarks were given by Prof. Adipala Ekwamu, he thanked facilitators for the job well done, participants for enduring through the training, RUFORUM for giving the platform and informed everyone that there is a series of training going on and encouraged participants to join them.

7.0 Analytical report for scientific data management

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

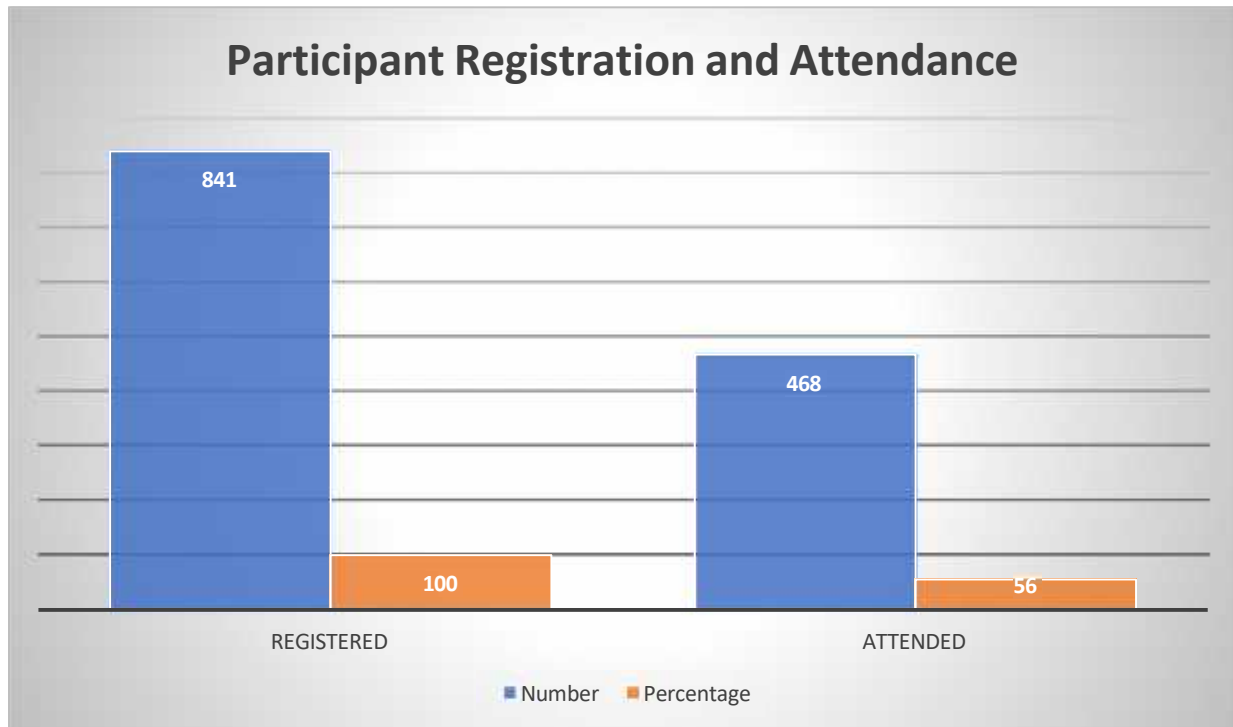




Co-funded by
the European Union



The number of participants: 841 Registered participants from 43 countries, 468 Attendees from 28 countries

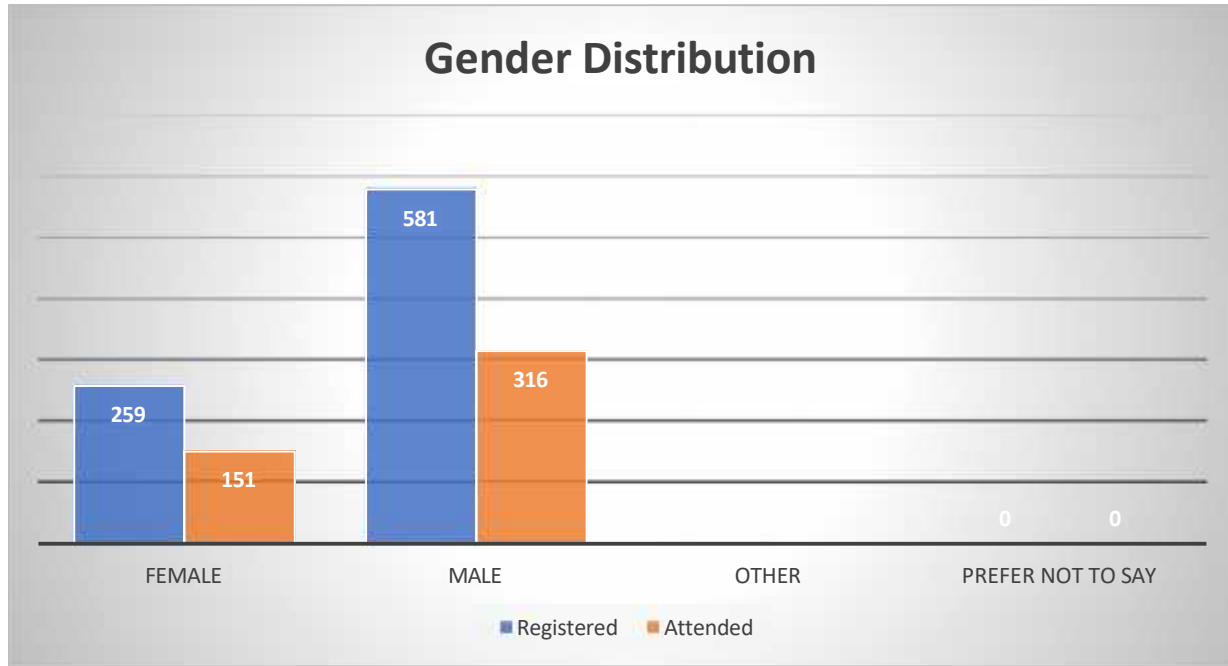


The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union

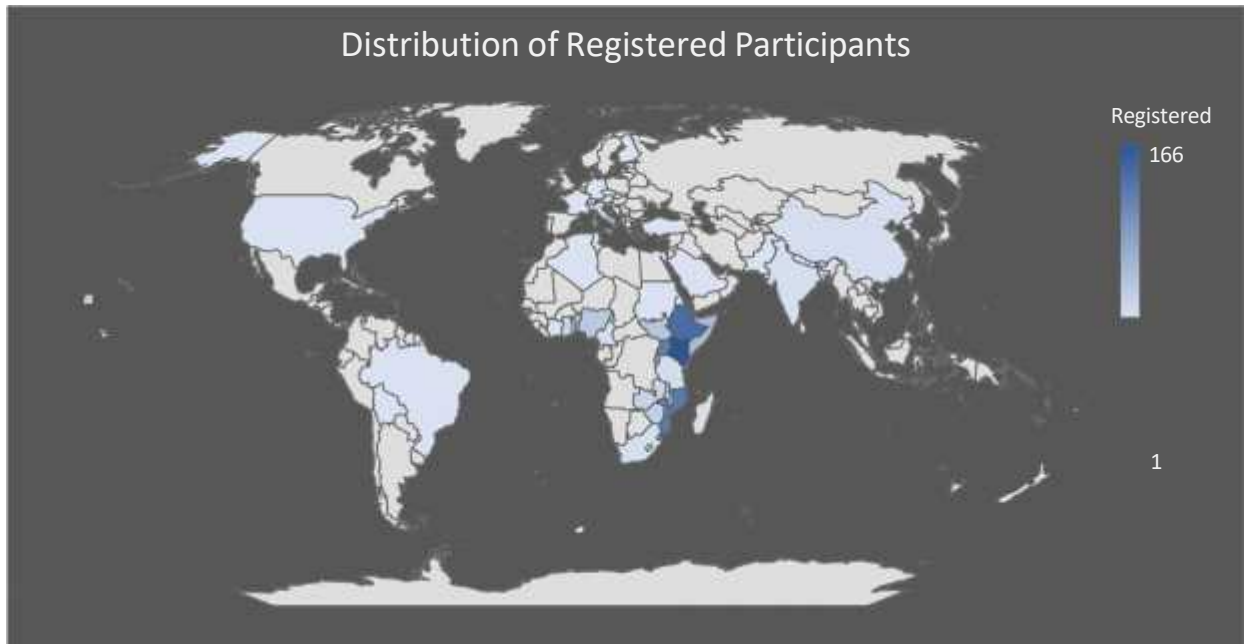


The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.



www.strcuu.org



Co-funded by
the European Union



The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



Participant distribution by country

No.	Country	Registered	Attended
1	Kenya	166	86
2	Ethiopia	137	74
3	Mozambique	120	65
4	Uganda	117	69
5	Somalia	47	22
6	Benin	45	40
7	South Sudan	34	17
8	Nigeria	30	18
9	Italy	23	7
10	Zimbabwe	23	13
11	Ghana	17	9
12	Zambia	17	15
13	Cameroon	7	4
14	Tanzania	7	5
15	South Africa	6	5
16	Congo, Democratic Republic of the	5	4
17	China	4	1
18	Rwanda	4	2
19	Czech Republic	3	2
20	India	3	0
21	Malawi	3	2
22	Germany	2	1
23	Algeria	1	1
24	American Samoa	1	1
25	Bangladesh	1	0

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



26	Bolivia	1	0
27	Brazil	1	0
28	Burundi	1	0
29	Cambodia	1	0
30	Côte d'Ivoire	1	1
31	Eritrea	1	0
32	Finland	1	1
33	France	1	1
34	Iraq	1	0
35	Japan	1	1
36	Martinique	1	0
37	Nepal	1	0
38	Pakistan	1	0
39	Saudi Arabia	1	0

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



40	Sudan	1	0
41	Togo	1	1
42	Turkey	1	0
43	United States	1	0
	Grand Total	841	468

8.0 Links used in the training

Here, the different links used in the training are detailed below.

8.1 Google Drive link with training materials

Training material link

<https://drive.google.com/drive/folders/1xg4euKUBNrGei9QVoKzDFkXy5mlp4uhO?usp=sharing>

The training videos can be accessed on

<https://www.youtube.com/@ResearchEducationAgency>

<https://www.youtube.com/channel/UCIBfm4qp8ZqjEUL7q4v3ePw>

8.2 Overall evaluation questionnaire link

Evaluation link

<https://docs.google.com/forms/d/e/1FAIpQLSfqTnXBnQ7zKYNLzaQ3O3GqIMFvrOD0eMZxgL1ocJCG-nMtdQ/viewform?usp=publish-editor>

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



9.0 Annexes

9.1 Annex 1: Timelines

Day one: Participants will use R programming language to manage data, writing the code for the different statistical methods

After day 1 of training in R programming we expect participants to have understood use of R in data management.

Day two: Participants will be described data sets using numerical summaries and graphical representations and done reporting with R Markdown.

Day three: Participants will be Discovered a complete overview of statistical methods for data analysis. They will know which method to use depending on the data available and the objectives to be achieved. Participants will acquire methodological and practical knowledge of linear regression methods to obtain an explanatory analysis of a phenomenon, to confirm hypotheses, to take decisions or to make forecasts. A global synthesis of the 4 modules will be made followed by a discussion with all participants on these models. This discussion can be related to the tasks they will be given or to their own experiences in data management and analysis.

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



N.B: Evaluation of participant's expectation will be conducted once in three days and an overall assessment of the training.

9.2: Annex 2: Training Program

Time	Tuesday (24/2/26)	Wednesday (25/2/26)	Thursday (26/2/26)
14:00-14:30	Registration Opening Ceremony Objectives & overview of course All	Overview of Exploratory data analysis (What is EDA? Importance of EDA, Types of EDA) Summarizing quantitative variables-univariate using R Summarizing qualitative variables -univariate using R TLO/ST/HN	Relationship between two variables - qualitative and quantitative Relationships between two qualitative variables Relationships TLO /ST/HN
14:30-15:00	Installation of R Introduction to R programming language	Relationship between two variables - qualitative and quantitative	<i>Univariate parametric and non-parametric statistical inference with R (TLO/ST/HN)</i>

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



	HN/ST/TLO	Relationships between two qualitative variables Relationships between two quantitative variables ST/HN/TLO	
16:15-	Health Break	Health Break	Health Break
16:30-	Introduction to R programming language Introduction to R programming language and reporting with R Markdown HN/ST/TLO	Overview of data management (quality measures, integrity, and consistency checks, etc.) Type of variables and measurements TLO/ST/HN	Correlation and Regression analysis Model building steps, validation of assumptions linear models (ANOVA, Regression) ST/HN/TLO
1800-	Closing	Closing	Closing

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



9.3 Annex 3: Training needs assessment tool

Pre-Course Questionnaire for Scientific Data Management Course Schedule for July, 2021

Section A: Personnel Information

Sex: Male Female

2. Name of degree being under taken

3. College/School/Department _____

4. _____ Title _____ of _____ your
research _____

5. Stage of research process

6. Email _____

SECTION B: Competence in Data Management and Analysis

Data Management

11. Knowledge in data management techniques: None Slight
Moderate Good Excellent

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



12. Knowledge in design of spreadsheet for data entry: None Slight
 Moderate Good Excellent
13. Knowledge in data checking: None Slight Moderate
Good Excellent
14. Knowledge in importation to statistical Software: None Slight
Moderate Good Excellent

Basic Statistics and Interpretation of Results

15. Knowledge in some basic descriptive statistics (*measures of central tendency and measures of dispersion*):
 None Slight Moderate Good Excellent
16. Knowledge in some basic inferential statistics (*hypotheses testing, t-test, ANOVA, Linear regression models*) :
 None Slight Moderate Good Excellent
17. Knowledge in data interpretation and reporting: None Slight
Moderate Good Excellent
18. Knowledge in design of data collection tools: None Slight
Moderate Good Excellent
19. Knowledge in presentation of results: None Slight Moderate
 Good Excellent

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



Statistical Software and Computer Skills

20. Which of the following statistical software you have access to and use/used?

Statistical software aware	Aware of	Have access to	Use or used
R			
Python			
SPSS			
SAS			
STATA			
GENSTAT			
MINITAB			
Any other specify:			

21. What is your level of knowledge in R?

None Slight Moderate Good Excellent

22. What is your level of knowledge in other statistics software (e.g. Genstat, Python, Stata; SPSS, SAS, etc.)?

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



None Slight Moderate Good Excellent

23 What is your level of knowledge in Spreadsheet/Microsoft Excel?

None Slight Moderate Good Excellent

SPECIFIC TOPICS

24. Please tick in the appropriate cell in the grid below, your assessment of your needs and capabilities in the topics listed (VD= very deficient, ND= Not deficient, MR= Major enhancement required, LR= Little enhancement required)

Component	Specific Topics	Level of competence on this topic					Level of enhancement required						
		VD				ND	MR				LR		
		1	2	3	4	5	1	2	3	4	5		
Data entry and management	Techniques of data checking												
	Data storage and retrieval												
	Data management strategy												

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



Data analysis and Interpretations	Exploratory data analysis												
	Tests relative to means												
	Correlation tests												
	Proportion tests												
	Independence tests												
	Non-parametric methods for experience/survey data												
	Linear regression models (simple and multiple)												
	Interpretation of statistical results												

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.





Co-funded by
the European Union



This form will be shared with registered participants and will be emailed to
the Resource Persons:

1. Dr Thomas Odong (thomas.l.odong@gmail.com)
2. Prof. Susan Tumwebaze (susantumwebaze7@gmail.com)
3. Dr Hellen Namaweje (namaweje-hnamaweje@gmail.com)

Thank you for taking off time to complete this questionnaire

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

