PROJECT DESCRIPTION

I-StAR: Introduction to Statistical Analysis with R

Period: April/May 2021

Pwani University, Kenya.

Address: P.O. Box 195-80108, Kenya
email: r.kigathi@pu.ac.ke

Tel: +2540708860351

Background Information

Graduate level students are required to have data analysis skills to enable them make sense of the data they collect and present it in a meaningful manner as a thesis or publications. Most of the postgraduate students are not equipped with proper data analysis and modelling skills due to software license issues or lack of proper expertise to train on this. The commonly used statistical software's in the region namely STATA®, SPSS® and SAS® can cost over 500 US$ for a single license and over 5,000US$ for a discounted 20 user access. Upgrading the software is an added expense. For example, KEMRI-Wellcome trust program one of our collaborators in Kilifi spend approximately 2,000 US$ upgrading their STATA license annually. At the moment, Public Universities do not have budget (cannot afford) for such statistical software acquisition, hence students and faculties depend mainly on researcher's limited licences. This limits training resulting in a of a large proportion of statistical training in Kenya lacking hands on statistical software use and relevant data, leading to students not being able to connect theory and application.

In addition, lack of code driven data analysis is also a major challenge in our departments with reproducibility of data analysis now being used as a minimum standard for judging scientific claims by journals. The main contributor for this is that most of the students are not trained do reproducible data analysis; thesis/publication linked with executable code and data.

In the last decade or so, the freely available statistical software R© has evolved to be a powerful tool in reproducible statistical analysis and is now used in various universities and institutions as the package of choice. The greatest advantage of R© is that it is available free on-line and has a big user community and numerous freely available tutorials for specialized areas. However, as it is not mostly menu-based, it has a steep learning curve, which makes it hard to use in the beginning. Most users is a need for interactive sessions to first get acquainted with R interface which then allows individuals to explore further once they understand the basics of coding in R. A hands-on 5 half day workshop on this statistical software will arm graduate level students and Faculty with the skills needed to independently perform tasks on R allowing them to then access other freely available materials to enhance their skills.

For the last 5 years, Pwani University in collaboration with KEMRI-Wellcome Trust has offered 2 week face to face R-workshop conducted within Pwani University. However the COVID-19 crisis meant that in 2020 this physical meeting to train in R could not be held. However, Pwani University

co-hosted a virtual DAAD Alumni with the Horticultural Association of Kenya (HAK) on 30th November to 4th December 2020 where we virtually trained 20 students and staff in a pre-workshop training on "*Methods to analyze the effects of climate change on plant growth - The Use of R Statistical Software*". During this training it became evident that the use of R software is not yet well known by both faculty and students (**Figure 1**). We thus focused on training on the use of R software. Due to that 2 day session we have gained experience on doing on-line training of R-statistical software. We thus wish to organize another event, with the expectation of harnessing on the previous experience.

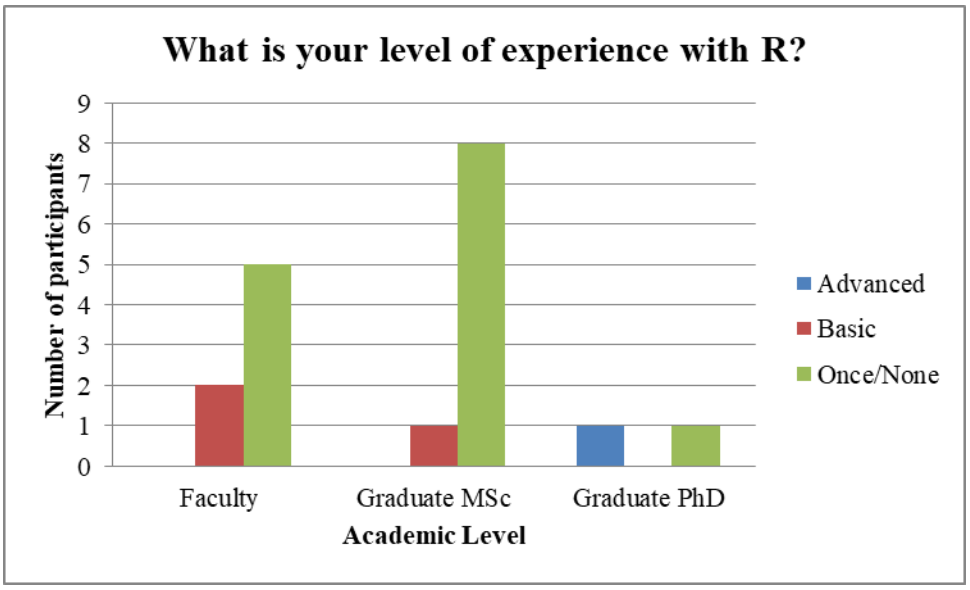

Figure 1: Pre-workshop questionnaire conducted on November 2020 before the HAK- R training session. Graph has excluded undergraduate students.

This will facilitate high quality research and remove the bottle-neck of over reliance on paid software that limit users and increase cost for both students and universities.

Content/Approach

The proposed R statistical software training will focus on improving proficiency in data management and manipulation in R. Training will focus mainly on hands on activities with little lecture sessions to introduce new content. Thus we expect to have atleast 5 facilitators to enhance interaction with participants. Statisticians and trainers from KEMRI-Wellcome Trust program and Komaza (**See Table 1**)will team up with lecturers from Pwani University to offer the five (5) half day workshop on "I-StAR: Introduction to Statistical Analysis with R". We also anticipate that we

will also invite facilitators from Alumni network to support the interactive session and help create an R-user network in Kenyan Universities.

Table 1: List of potential trainers/facilitators

Sr.	NAME	INSTITUTION	ROLE	ALUMNI
1	Ken Mwai	KEMRI-Wellcome Trust	Trainer	NO
2	Moses Kioko	KOMAZA	Trainer	NO
3	Dr. Alice Kamau	KEMRI-Wellcome Trust	Trainer	NO
4	Mark Otiende	KEMRI-Wellcome Trust	Trainer	NO
5	Dr. Kiti Alii	Pwani University	Facilitator	NO
6	Sammy Wambua	Pwani University	Facilitator	NO
7	Rose Kigathi	Pwani University	Facilitator	YES
8	Zaneta Kidiavai	KEMRI-Wellcome Trust/	Facilitator	DAAD Scholar

In this workshop, we will use virtual classrooms provided by Kenya Education Network Trust (KENET) that relies on the BigBlueButton® Conferencing System. This will be used in combination with the RStudio Cloud® (https://rstudio.cloud/) that allows students to log in to the cloud and trainers can access the working spaces of students. These allow direct interaction of students and trainers as they can virtually access the working spaces of students. This allows trainers to directly correct errors and demonstrate exercises where need be. We have used this system before during the HAK-workshop and we thus already have experience with these interfaces.

Goals and Objectives

The objectives of the training are to:

- To equip Faculty, Masters and PhD students with computing skills to import and manage their data with minimum supervision using R.
- Provide a networking opportunity for the trainers and students to engage and foster further collaborations in enhancing the use of quantitative techniques in research work.
- To introduce cost-free statistical environment for statistical computing in an easy and straightforward manner.

Table 1: Matrix of objectives and deliverables

Objective	**Deliverable**
To equip Masters' and PhD students with computing skills to import and manage data with minimum supervision using R.	1) Introduce students to R programming language 2) The participants will be trained on data manipulation and basic r-coding skills to help build confidence in the use of R 3) Participants will be registered to an R online community to ask questions, learn, share, their data analysis, knowledge, and build their careers.
Provide a networking opportunity for the trainers and students to engage and foster further collaborations in enhancing the use of quantitative techniques in research work	1) We will assign each participant to a course facilitator to guide or assist them in case of a major issue during their data analysis 2) The course facilitators will be added to the online forum participate in the online forum.
To introduce cost-free statistical environment for statistical computing in an easy and straightforward manner	1) The course notes will be provided in a reproducible manner – code, lectures and data shared on an online repository open to other institutions of interest. 2) Participants will set up the R environment on their personal devices.

Target group of the event.

The main target is really to empower faculty, who have an immediate multiplier effect. Many staff many have good knowledge of statistics and paid software e.g SAS® or STATA® but no working knowledge in this free software R©. Recently in Pwani University alone we had 10 staff requesting for training in R to improve their training and supervision skills. About 15 Faculty members and 10 MSc./PhD students will be expected to attend (**Table 1**). A brief half day survey to determine interest has drawn 29 DAAD Alumni and 15 non- Alumni (**Table 2 and 3**) with many DAAD Alumni expressing interest of inviting students to join the training. However, only 25 participants will be drawn from the interest list.

Table 1: Expected ratio of participants

	DAAD Alumni	**Non-DAAD Alumni**
Faculty	12	5
Graduate Students	3	5

EVENT PROGRAMME

	DAY 1
	Background, Getting Started, and Nuts & Bolts
8:30 -9:30 am	Introduction to concepts of Data Management (Folders, folder structure)
9:30-10:30	Set up R and RStudio and RStudio Cloud
10:30-10:45	*Health Break*
10:45-11:00	Good data programming skills
11:00-12:45	~~Introduction to R and RStudio- Basic functions~~
	DAY 2
8:30 -9:30	Recap, Basic functions in R
9:30-10:30	Numerical exploration & Data wrangling with R
10:30-10:45	*Health Break*
10:45-11:00	Numerical exploration & Data wrangling with R
11:00-12:45	Numerical exploration of data with R
	DAY 3
8:30 -9:30	RECAP
9:30-10:30	Numerical exploration of data with R
10:30-10:45	*Health Break*
9:30-10:30	Data exploration in R- exercises
11:00-12:45	Introduction to plotting with R - ggplot
	DAY 4
8:30 -9:30	RECAP
9:30-10:30	Data visualization using R - ggplot
10:30-10:45	*Health Break*
9:30-10:30	Exploratory analysis of continuous data
11:00-12:45	Exploratory analysis of continuous data
	DAY 5
8:30 -9:30	RECAP
9:30-10:30	Exploratory analysis of categorical data
10:30-10:45	*Health Break*
9:30-10:30	Exploratory analysis of rates and ratios
11:00-12:45	Reproducible Research in R and wrap up.

LIST OF PROSPECTIVE PARTICIPANTS

Table 2: List of DAAD Alumni who have expressed interest in the R-course. Only 15 Alumni will be selected based on the availability.

Sr.No	Name	Institution	Alumni
1	Dr. Simon Muriu	Pwani University	YES
2	Dr. Fredrick Nyamwala	Moi University	YES
3	Dr. James Gacheru Wanjiku	Taita Taveta University	YES
4	Dr. Regina Bwire	Masinde Muliro University of Science and Technology (MMUST)	YES
5	Victor Odongo	Pwani University	YES
6	Lenard Mounde Gichana	Pwani University	YES
7	Stellamaris Ogutu		YES
8	Dr. John Mumbo	National Environment Management Authority	YES
9	Leahbell Walusuna	Pwani University	YES
10	Daniel Rexford Nyamekye	YARA Farming Solutions	YES
11	Dr. Pascaline Jeruto	University of Eldoret ,Kenya	YES
12	Bidii Ngalah	Heidelberg University	YES
13	Fredrick Okumu	Jaramogi Oginga Odinga University of Science and Technology (JOOST)	YES
14	Victor Odari	Masinde Muliro University of Science and Technology (MMUST)	YES
15	Collins Handa	Technical University of Kenya	YES
16	Vitalis Ogemah	Masinde Muliro University of Science and Technology (MMUST)	YES
17	James Linturi	South Eastern Kenya University	YES
18	Moses Muraya	Chuka University	YES
19	Alexander Muthanga	Pwani University	YES
20	Dr. Nicholas Ajwang'	Pwani University	YES
21	Everline Nyokabi	Maseno University	YES
22	Dr. Joel Okutoyi	Maseno University	YES
23	Duncan Gathungu	Jomo Kenyatta University of Agric &Tech (JKUAT)	YES
24	Jane Wangui Mugo	Nairobi University	YES
25	Dr. Patrick W. Okanya	Technical University of Kenya	YES
26	Dorcas Lusweti	Moi University	YES
27	Alex Okaru	Nairobi University	YES
28	Urbanus Mutwiwa	Jomo Kenyatta University of Agric &Tech (JKUAT)	YES
29	Felista Waihuini Muriu-Ng'ang'a	South Eastern Kenya University	YES

Table 3: List of other individuals who have expressed interest in the R-course. Only 15 Alumni will be selected based on the availability.

	Name	**Institution**	**Alumni**
1	Bonventure Obeka	Pwani University	Student
2	Timothy Musa	Pwani university	Student
3	Stanely Wanjala	Pwani University	Student
4	Dennis Olumeh	Humboldt University in Berlin	Student
5	Maureen Anyango Olwero	Masinde Muliro University of Science and Technology	NO
6	Danstone Lilechi Baraza	Masinde Muliro University of Science and Technology(MMUST)	NA
7	Purity Ncabira	Taita Taveta University	N/A
8	Esther Magembe	Kibabii University	N/A
9	Virginia Kimani		N/A
10	Mary Mwangi	South Eastern Kenya University	N/A
11	Obiero GF		N/A
12	Dr. Virginia Gichuru	Pwani University	NO
13	Edith Muwawa	Pwani University	NO
14	Nyinge N. Mwadzombo	Pwani University	NO
15	Dr. B. Fulanda	Pwani University	NO

Detailed Budget:

Item	**Persons**	**Days**	**Rate (KES)**	**Amount (KES)**	**EUR Equiv.***
Communication Costs					
Airtime support for Alumni	15	5	1,000.00	75,000.00	625.00
Airtime for students/others	10	5	1,000.00	50,000.00	416.67
Airtime support for trainers/facilitators	6	5	1,000.00	30,000.00	250.00
Organizing/pre workshop Airtime	6	1	500.00	3,000.00	25.00
Technical Equipment					
Payment for R Cloud Services	1	1	1,500.00	1,500.00	12.50
Honoraria					
Honoraria for Trainers	4	5	15,000.00	300,000.00	2,500.00
Grand Total				**459,500.00**	**3,829.17**

*Exchange rate used 120Kshs per 1 Euro.

Note: Airtime takes into account use of RStudio Cloud for data download/upload, practice and assignments outside of the training hours.