



University of
St Andrews
FOUNDED
1413

School of Geography and
Sustainable Development



Global Challenges Research
Fund



Holistic Approach To
Unravel Antibacterial
Resistance in East Africa

Handbook for International online workshop: Introduction to statistical analysis using R Studio

Led by Katherine Keenan and Dominique Green

With support from Xuejia Ke

SESSION 1: Monday 14th December 10:00-12:00 (UK time UTC)

SESSION 2: Weds 16th December 10-12 (UK time UTC)

SESSION 3: Thursday 17th December 10-12 (UK time UTC)

List of attendees

Invited participants			
	Country	Name	Email
1	Uganda	Catherine Kansiime	cathie.kansiime@gmail.com
2	Uganda	Benjamin Sunday	warufubenjamin@gmail.com
3	Uganda	Richard Mukasa	mukasa151@gmail.com
4	Uganda	George Sendegye	georgesende@gmail.com
5	Uganda	Ivan muhwezi	muhweziivano@gmail.com
6	Uganda	Eva Laker	elaker@idi.co.ug
7	Uganda	Ruth Mirembe Nabisere	rnabisere@idi.co.ug
8	Kenya	IanaAmke	ianaamke@gmail.com
9	Kenya	Sharon Sang	sangsharon1@gmail.com
10	Kenya	Zamith Oginga	zamithoginga@gmail.com
11	Kenya	John Maina	johnmaina.jnm@gmail.com
12	Kenya	Dr. John Mwaniki	mwanikij@gmail.com
13	Kenya	Fredrick Wanja	waniafredrick356@gmail.com
14	Kenya	Japheth Katana	katanaj1989@gmail.com
15	Kenya	Annette Aduda	aduda.annette@gmail.com
16	Kenya	Susan Wambui	suzikoah@gmail.com
17	Kenya	John Kiiru	kyirow@gmail.com
18	Kenya	Franklin Mwiti	frankusmwitus@gmail.com
19	Kenya	Elsie Isiye	isiyelsie@gmail.com
20	Tanzania	Prof Stephen Mshana	stephen72mshana@gmail.com
21	Tanzania	Dr Joseph Mwanga	jrmwanga@yahoo.co.uk
22	Tanzania	Dr Martha Mushi	mushimarta@gmail.com
23	Tanzania	Miss Perry Msoka	perrymsoaka98@yahoo.com
24	Tanzania	Dr Eveline Konje	ekonje28@yahoo.com
25	Tanzania	Dr Stella Kimwela	stlmugassa@gmail.com
26	Tanzania	Mis Pendo Ndaki	pendondaki@yahoo.com
27	Tanzania	Dr Walter Manyiri	msirywalter91@gmail.com
28	Tanzania	Prof Mariam Mirambo	mmmirambo@gmail.com
29	Tanzania	Dr Jerimiah Seni	senijj80@gmail.com
30	Tanzania	Mr Vitus Silago	vsilago.silago2@gmail.com
31	Tanzania	Mr Betrand Msemwa	b.msemwa@yahoo.com
32	Tanzania	Mr Adam Mwakyoma	amwakyoma@gmail.com
33	Tanzania	Mr Shukran Philip	phillipshukrani@yahoo.com
34	Tanzania	Miss Conjester Mtemisika	conjestermtemisika@yahoo.com
35	Scotland	Katy Keenan	katherine.keenan@st-andrews.ac.uk
36	Scotland	Dominique Green	dlg4@st-andrews.ac.uk
37	Scotland	Mike Kesby	mgk@st-andrews.ac.uk
38	Scotland	Xuejia Ke	xk5@st-andrews.ac.uk
39	Scotland	Alison Sandeman	as7@st-andrews.ac.uk
40	Scotland	Wilber Sabiiti	ws31@st-andrews.ac.uk
41	Scotland	Arun Decano	ad322@st-andrews.ac.uk
42	Scotland	Sarah Huque	sih3@st-andrews.ac.uk

Introduction

This workshop is funded by 'UK Research and Innovation' (sponsored by the UK Department for Business, Energy and Industrial Strategy) under the 'Global Challenges Research Fund'. It is part of a research grant held by Dr Katherine Keenan and Dr Mike Kesby (PIs) called "Maximising analytical and translational capability for social science data on antimicrobial resistance (AMR) in East Africa". The project aims to build scientific, analytical and political capacity to use the available social scientific data on AMR (antimicrobial resistance) to make a positive policy impact. It builds upon and uses data from the HATUA (**H**olistic **A**pproach **T**o **U**n unravel **A**ntibacterial resistance in East Africa) Consortium. Originally this training course was to be an in-person event taking place in Mwanza, Tanzania in April 2020. However, the Covid-19 crisis necessitated that we adapt the training for a series of online events. The course will now take place in three online interactive practical sessions of 2 hours each, held over one week.

This training will introduce the basics of using R Studio (free online statistical software) for producing basic statistics using quantitative data. It is designed to be accessible for beginners to quantitative analysis from many disciplines; everyone regardless of whether you have experience with quantitative analysis, statistics or R Studio can engage with the course.

Aims of the training course:

- Introduce basic statistical concepts including ideas of samples and populations, quantitative data, how it is measured and collected, different types of variables, distribution of variables, univariate and bivariate analysis, hypothesis testing
- Install R Studio and become confident/familiar with R Studio environment
- Learn some basic data management skills (dealing with missing data, recoding variables)
- Learn some basic data visualisation techniques
- Produce univariate and bivariate descriptive statistics using cross-sectional data
- Become more familiar with some variables in the HATUA study and in studying AMR through a social science perspective

Trainers

Dr. Katherine Keenan, Lecturer, University of St Andrews, Co-I of the HATUA Consortium
(Katherine.keenan@st-andrews.ac.uk)

Dr. Dominique Green, Postdoc data analyst on the HATUA Project, University of St Andrews
(dlg4@st-andrews.ac.uk)

Ms Xuejia Ke, PhD student on HATUA project, University of St Andrews

With input from:

Dr. Martha Mushi, , Catholic University of Health and Allied Sciences, Mwanza, Tanzania; **Dr. Catherine Kasiime**, Makerere University, Uganda..**Ms. Annette Aduda**, KEMRI, Kenya.

Where to find training materials

We will post all materials on [google drive](#)- you'll be sent a link to the folder.

Materials will include:

- Slides for each session
- Worksheets and solutions
- List of accompanying resources e.g. useful self-help videos
- HATUA Teaching dataset and codebook
- Videos and recordings of the sessions, how to do key things
- Info on the slack channel to ask questions to us and to each other

Introduction to coding in R Studio

R Studio is open-source software for analysing quantitative data. The Rstudio user community is growing very rapidly in many fields. It is freely available - anybody can install R Studio on their own machine¹. You need an internet connection for R Studio to work smoothly, as it relies on downloading packages from the internet as you go along. Compared with many other proprietary statistical software such as STATA and SPSS, it is **free**. It is also highly flexible – and users are always developing new ‘packages’ – these are add-ins that are available to the entire user community, that can do additional things, solve data problems or help you do certain tasks. It is also well known for being the best software for developing data visualisations.

R Studio is a command-based program, rather than a windows-based program, where you click and point at stuff. You, the user, must write code to tell R Studio what to do. There is a correct (and an incorrect) syntax for this code. It is completely normal to make lots of mistakes at coding, and this is how we correct the mistakes and learn.

On the google drive there is a separate doc with a list of online resources on statistics, and on using R Studio.

¹ Caveat. Older macbooks might be problematic. If you work on a macbook with an older version of the operating system, anything below 10.12 (Sierra), like el Capitan, you will need to upgrade your mac operating system to 10.12 or higher. If this is the case, there are some instructions on how to upgrade your operating system online yourself, or maybe you can seek help from your own IT staff to help you. The process may depend on what operating system you are running.

Session 1: Introduction to Quantitative Analysis; installing and becoming familiar with R Studio

Monday 14th December 10-12 UK time – on zoom (will be recorded)

Join Zoom Meeting

<https://us02web.zoom.us/j/81900931631?pwd=ckh0VUswaW4rK0duRmo4NnN4dldKUT09>

Meeting ID: 819 0093 1631

Passcode: 528526

Led by Katy Keenan – with help from DG, XK, MM and CK.

Proposed schedule:

First hour 10-11 (slides will be provided on google drive; session will be recorded)

- Introductions
- The purpose of quantitative data
- Principles of statistics- sampling, populations and inference
- Types of data, types of variables
- Overview of the HATUA quantitative data
- R Studio advantages / how it differs from STATA, SPSS or SAS

Practical session – 11-12 – instructions will be on google drive

- Everyone installs R Studio on their own machines
- Troubleshooting issues with installation (we may use break out groups for this)

Step-by-step Instructions for installing R Studio:

This is 2-step process, follow the instructions here: <https://rstudio.com/products/rstudio/download/>

1. Click download R studio desktop (the free one)
2. Follow step 1: Download [R 3.0.1+](#).
3. When this is installed, download **RStudio Desktop**- your machine will pick up what type of computer you are using and recommend the correct version.

NB: Important: If you work on a mac with an older version of the operating system, anything below 10.12 (Sierra), like Capitan, you will need to upgrade your mac operating system to 10.12 or higher. If this is the case, and you cannot find how to do this from google, please go to IT helpdesk in the library with your machine and ask for guidance on this.

- (i) Overview of the R Studio environment; dataframes; R scripts; syntax Opening a dataset, saving a dataset

After the session- self help

If you wish, you may continue to explore the training dataset, and watch some videos in preparation for the next session to consolidate your knowledge of statistics and R Studio (see helpsheet in google box)

Session 2: Basic data management, univariate descriptive statistics

Weds 16th December 10-12 UK Time

Led by Katy Keenan

Join Zoom Meeting

<https://us02web.zoom.us/j/86255183372?pwd=c0N4UW4xd3BuZWk2emlQaktBckdsQT09>

Meeting ID: 862 5518 3372

Passcode: 760914

Proposed outline for the session

1. 10-10:45: Intro to univariate descriptive statistics (PPT slides and recordings)

- Questions /issues arising from the last session
- Data management (e.g. recoding a variable)
- Purpose of descriptive statistics and data visualisation
- Ways of summarising different types of variables (univariate statistics)

2. Remaining part of the session- interactive practical session in R Studio (see worksheet and recordings)

- Worksheet tasks for all, which we go through and then show the solutions to the group as we go along
 - Exploring continuous and categorical variables
 - Making plots /graphs using ggplot2
-

Session 3: Exploring relationships between two variables (bivariate statistics)

Thursday 17th December 10-12 UK Time

Led by Katy Keenan and Dominique Green

Time: Dec 17, 2020 10:00 AM London

Join Zoom Meeting

<https://us02web.zoom.us/j/83661300087?pwd=MDIiL3VzWGRnYW1SYStNMIFPdTZGZz09>

Meeting ID: 836 6130 0087

Passcode: 450629

Proposed outline for the session

1. 10-10:45: Intro to bivariate descriptive statistics (PPT slides and recordings)

- Questions /issues arising from the last session
- Purpose of bivariate statistics; hypothesis testing
- Correlation for continuous variables
- Contingency tables for categorical variables
- Chi- square test, fisher's exact for small samples

2. Remaining part of the session- practical session in R Studio (see worksheet and recordings)

- Using HATUA teaching dataset to explore associations between antibiotic misuse and socio-demographic characteristics
 - Contingency tables and testing for differences between two categorical variables
-