The GRAPH Courses

Founding, vision and progress

Data-driven decisions save lives

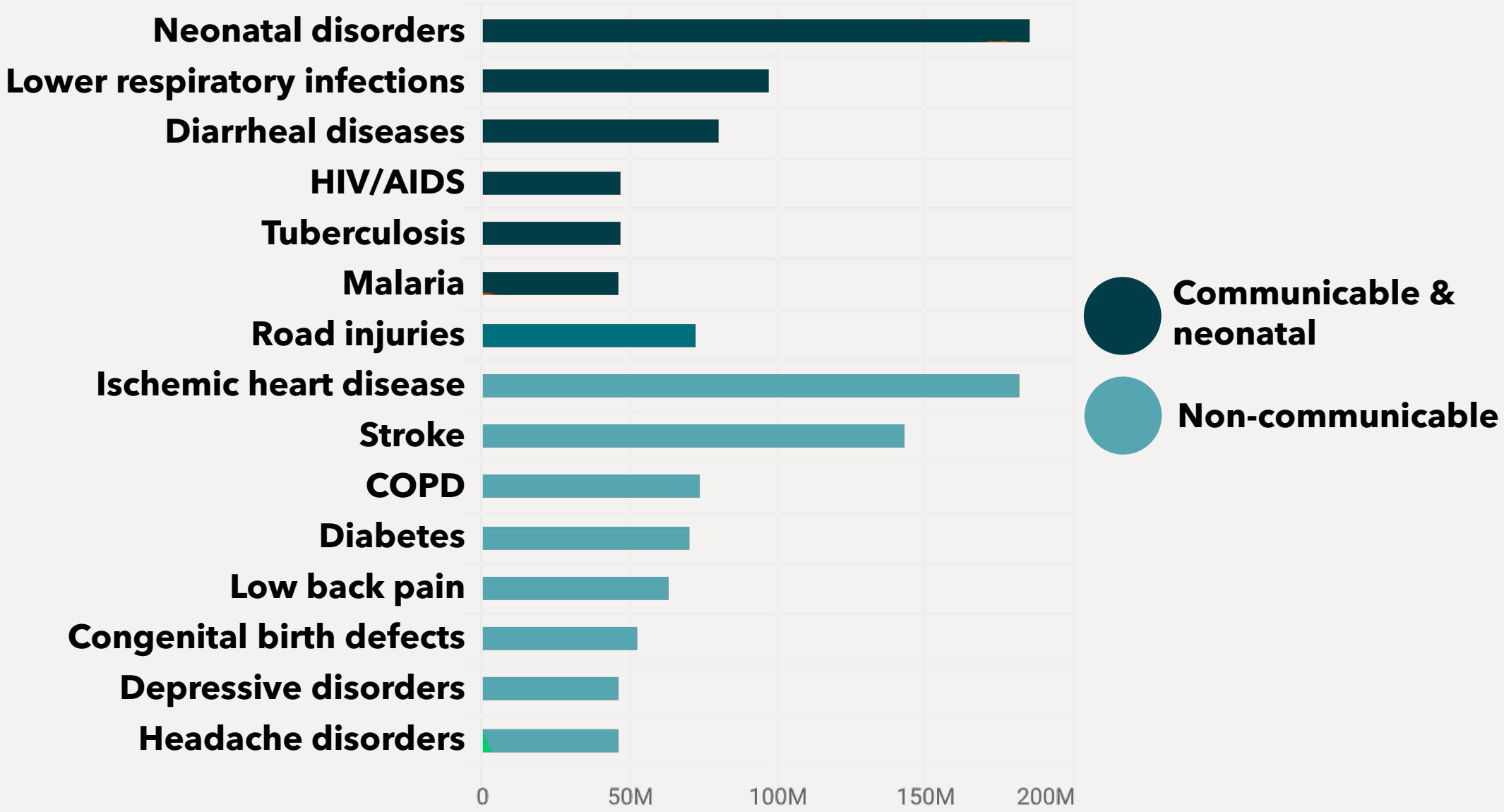
Who is most vulnerable

How to flatten the curve

How to treat

How to prevent

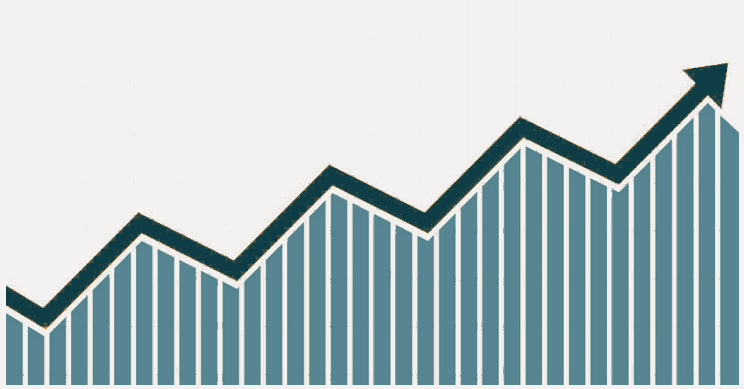
Top 15 causes of death and disease globally (DALYs lost, 2019)



Data source: IHME

Progress

in collecting public health data

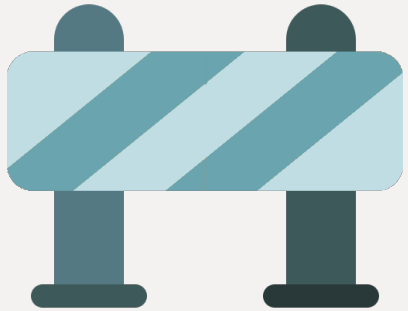


Challenges

in data management

in translating data into insight & policy

particularly severe in low & middle income countries (LMICs)



Solutions

The EpiGraphHub

Connecting to and managing real-world epidemiological datasets

The GRAPH Courses

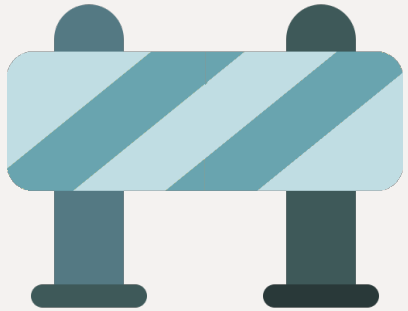
Training a generation of health data analysts

Challenges

in data management

in translating data into insight & policy

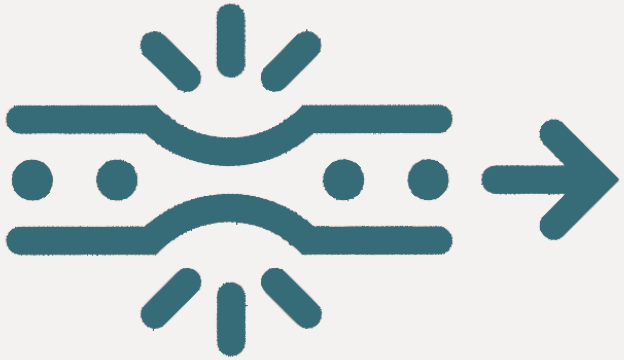
particularly severe in low & middle income countries (LMICs)



Why are we building the GRAPH Courses?

Training bottlenecks in LMICs

Rich pool of potential

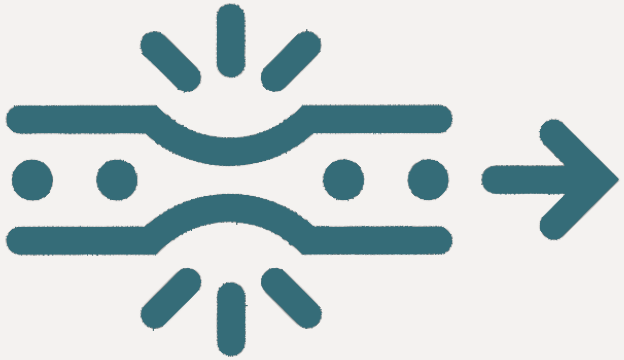


Health data scientists

Large young population
↑ Computer & internet access
↑ Interest in software & data
Ready for quick upskilling with targeted courses

Training bottlenecks in LMICs

Rich pool of potential

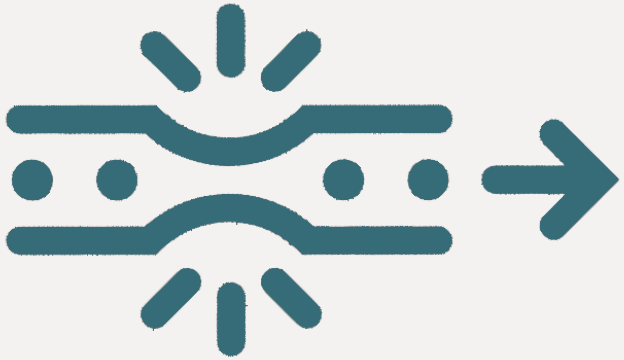


Health data scientists

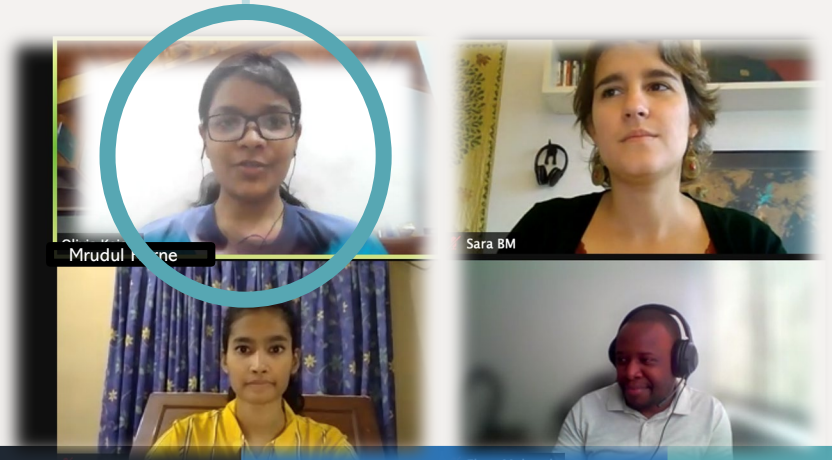
Chigere Amarachi
(HIV key populations program)
Accredited courses from reputable institutions "are usually ridiculously expensive"

Training bottlenecks in LMICs

Rich pool of potential



Health data scientists



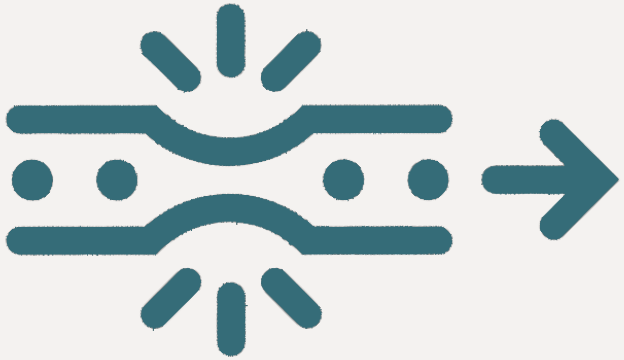
Mrudul Harne

(Hospital administration)

Many courses lack ”practical hands-on experience with projects and datasets”

Training bottlenecks in LMICs

Rich pool of potential



Health data scientists

Elo Nwosu

(Secondary school)

”can’t find classmates or teachers I can ask questions to” in online data analysis courses

Training bottlenecks in LMICs

Rich pool of potential

The GRAPH Courses

Health data scientists

Training bottlenecks in LMICs

Rich pool of potential

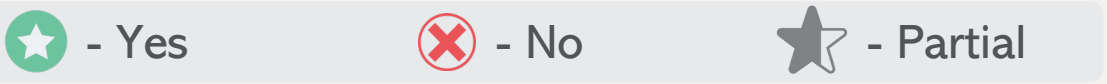
The GRAPH Courses

Health data scientists

Essential features

Epi-projects	• Projects with real-world health data from LMICs
Affordable for LMICs	• Volunteers & philanthropy
Learning community	• knowledge sharing & relationship building
Accreditation	• Tracked assessments & Certificates

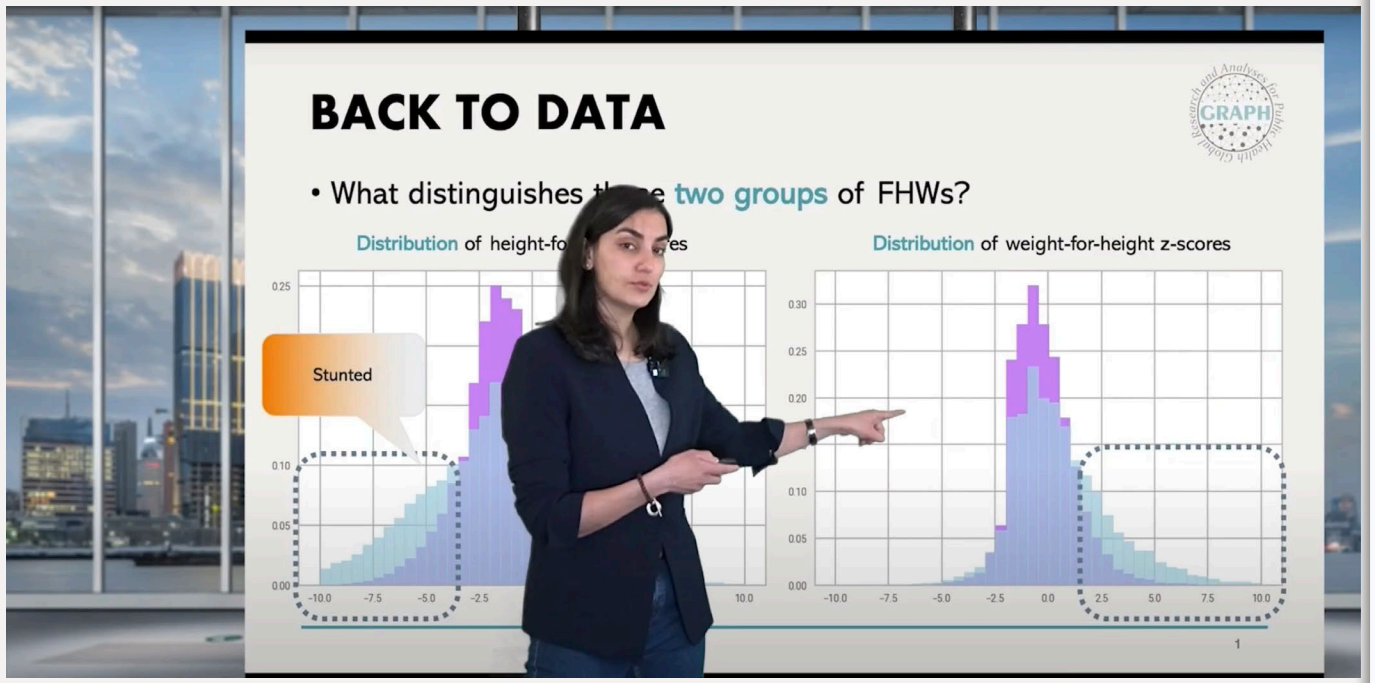
Essential features



	DATA SCIENCE PLATFORMS Datacamp, Dataquest, Codeacademy, etc.	HIGHER-ED MOOC PLATFORMS Coursera, EdX, etc.	R FOR EPI BOOKS R4Epis, Epirhandbook, RMRWR, etc.	OTHER DATA SCIENCE BOOKS R4DS, Datacarpentry Psyteachr., etc.	Our solution The GRAPH Courses
Epi-projects	✗	★	✓	★	✓
Affordable for LMICs	✗	✓	✓	✓	✓
Learning community	★	★	★	★	✓
Accreditation	✓	✓	✗	✗	✓

Platform overview

Lesson structure



2 The Yaounde COVID-19 dataset

In this lesson, we analyse results from a COVID-19 serological survey conducted in Yaounde, Cameroon in late 2020. The survey estimated how many people had been infected with COVID-19 in the region, by testing for IgG and IgM antibodies. The full dataset can be obtained from Zenodo, and the paper can be viewed here.

Spend some time browsing through this dataset. Each line corresponds to one patient surveyed. There are some demographic, socio-economic and COVID-related variables. The results of the IgG and IgM antibody tests are in the columns `igg_result` and `igm_result`.

```
yaounde <- read_csv(here::here('chapter_04_data_wrangling/data/yaounde_data.csv'))
yaounde
```

id	date_survey…	age	age_category	sex	education
BRIQUETERIE…	2020-10-22	45	45 - 64	Female	Secondary
BRIQUETERIE…	2020-10-24	55	45 - 64	Male	University
BRIQUETERIE…	2020-10-24	23	15 - 29	Male	University
BRIQUETERIE…	2020-10-22	20	15 - 29	Female	Secondary
BRIQUETERIE…	2020-10-22	55	45 - 64	Female	Primary

1-5 of 971 rows Previous **1** 2 3 4 5 … 195 Next

Practice in the code console

PRACTICE

- Select the weight and height variables in the `yaounde` data frame. Store the result in `q1`.

```
q1 <- "yaounde %>%"
.check_q1()
.hint_q1()
```

- Subset the `yaounde` data frame to columns 16 and 22. Store the result in `q2`.

```
q2 <- "yaounde %>%"
.check_q2()
.hint_q2()
```

```
yaounde %>% select(sex, age, igg_result)
```

.practice

- Select the weight and height variables (in that order) in the `yaounde` data frame. Store the result in `q1`.

```
{r, include = F}
q1 <- "yaounde %>%"
.check_q1()
.hint_q1()
```

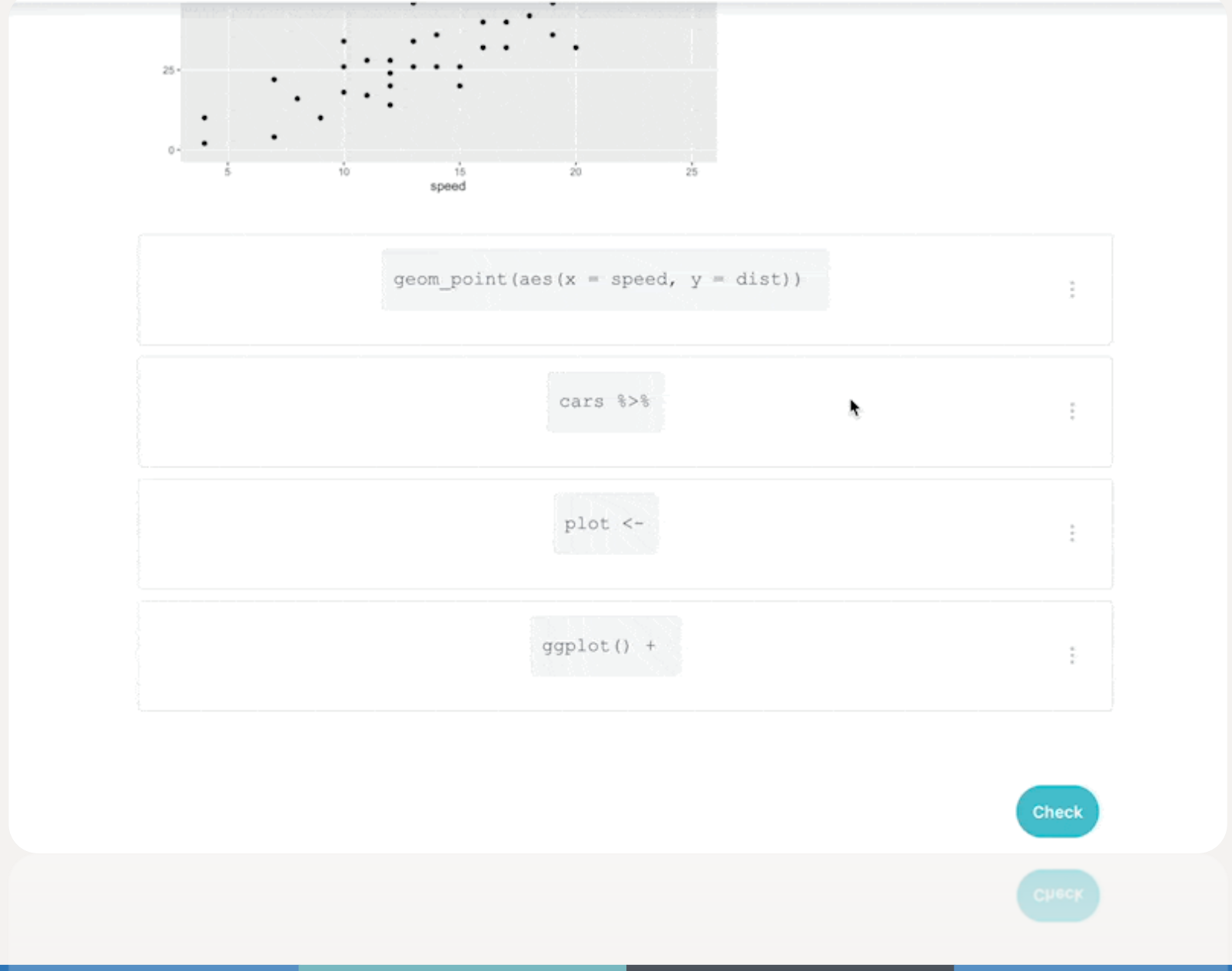
- Subset the `yaounde` data frame to columns 16 and 22 (in that order). Store the result in `q2`.

```
{r include = F}
q2 <- "yaounde %>%"
.check_q2()
.hint_q2()
```

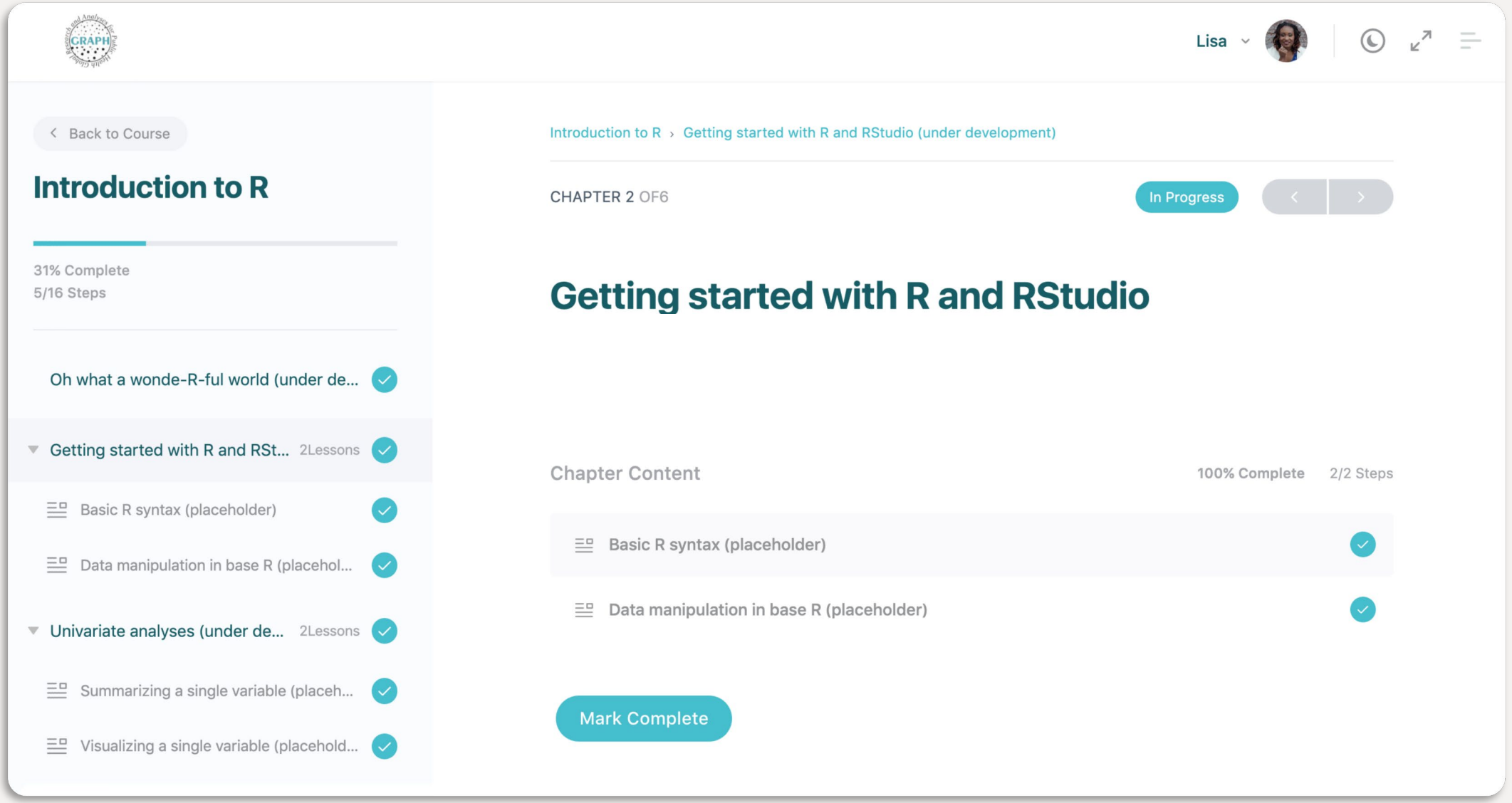
In the examples above we used the pipe symbol `%>%` to pass our data into `select()`. But we can also provide the data as the first argument of `select()`, without the pipe.

```
{r}
```

Quizzes and assessments



Progress tracking

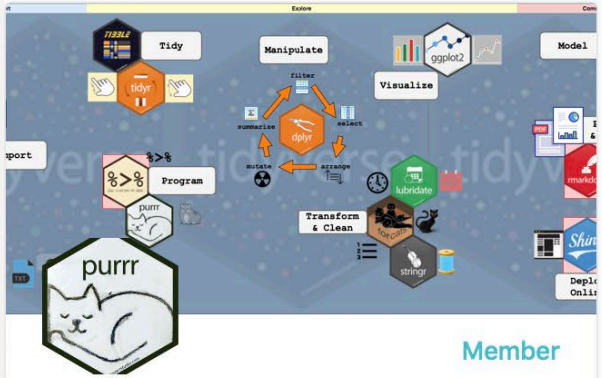


Community features: user-created groups

Home | Courses (under development) | The Network | Laure

All Groups 4 | My Groups 1 | Create a Group

Recently Active

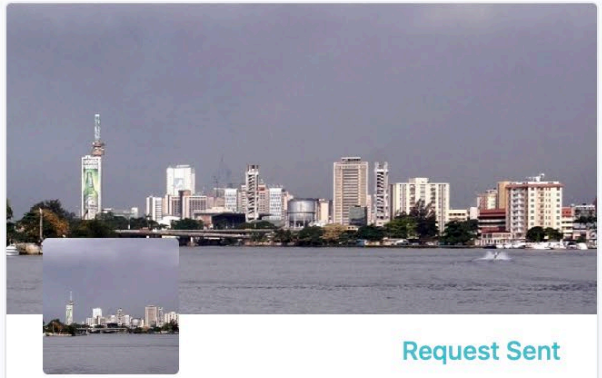


Member

GRAPH Courses Team

Hidden Group

+3 members



Request Sent

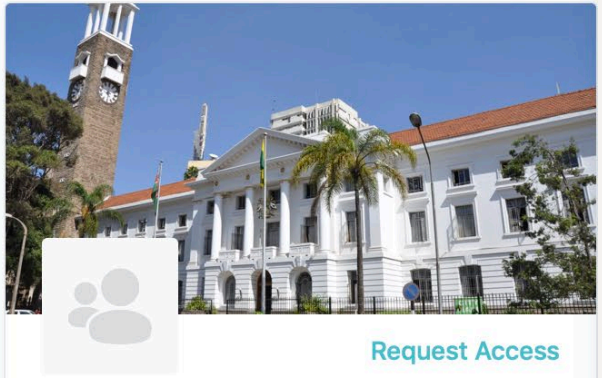
Lagos R meetup

Private Group

Request Access

Tidy Tuesday

Private Group

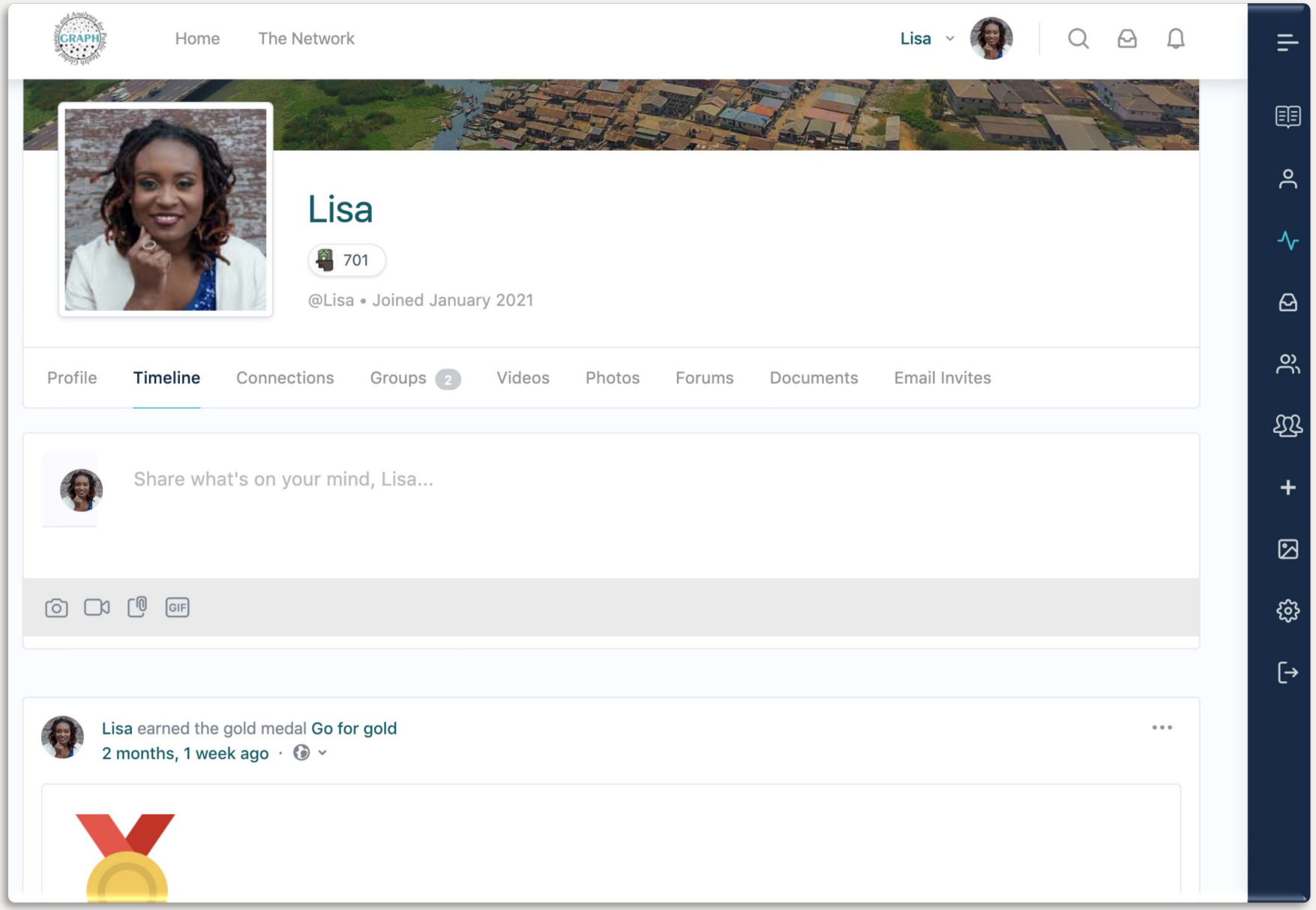


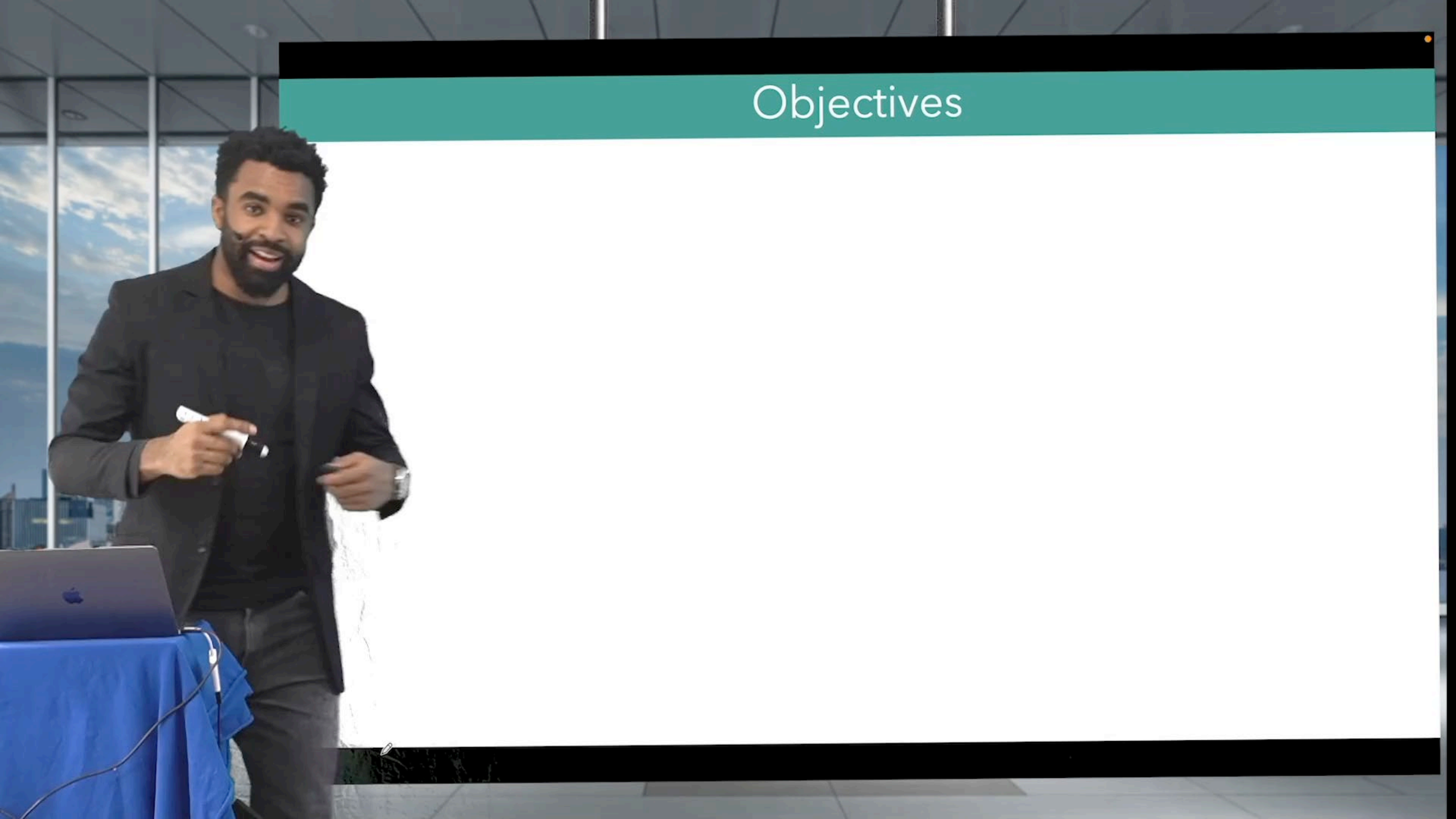
Request Access

Nairobi R meetup

Private Group

Community features: user profiles




Objectives

Sample Lesson

tinyurl.com/graph-sample-lesson

What will we teach?

Courses on offer

by July 2022

- **Introduction to data analysis with R**
- **Further data analysis with R**
- **Introduction to EpiGraphHub** (mini course)

by September 2022

- **Introduction to Epidemiology**
- **Practical Statistics for Epidemiology using R**

Seed funding:

World Health Organization

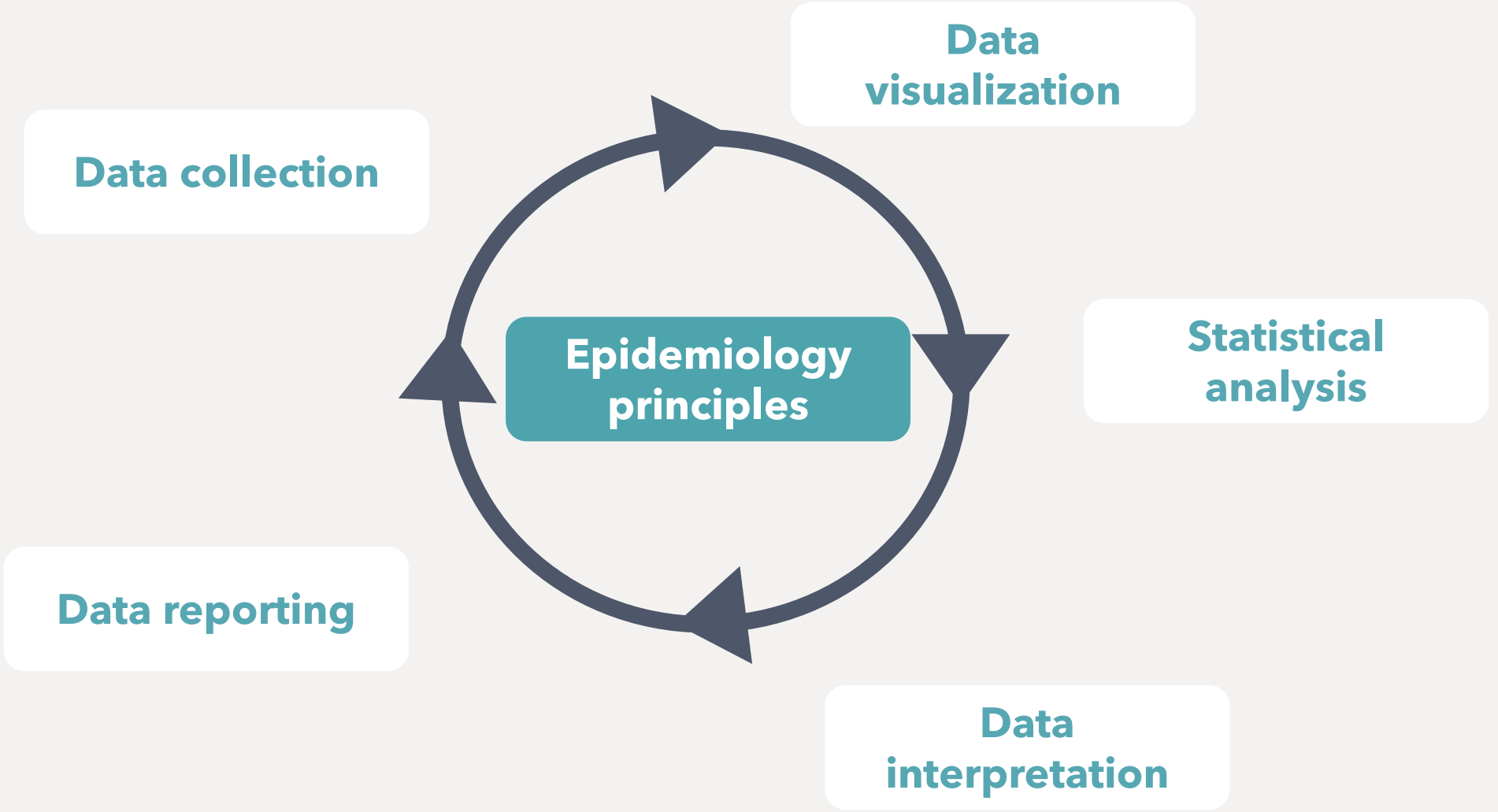
CDC Foundation
Together our impact is greater

Further courses in the pipeline

- **Epidemiological Study Designs**
- **Clinical trials**
- **Importing and analysing DHIS2 data with R**
- **Research Process: From Questions to Applications**
- **Epidemiological data analysis with Python**
- **Survey design and analysis with KoboCollect and koboloadeR**

Our long-term vision

To create a suite of courses that cover the entire "life-cycle" of <u>evidence-based</u> public health practice



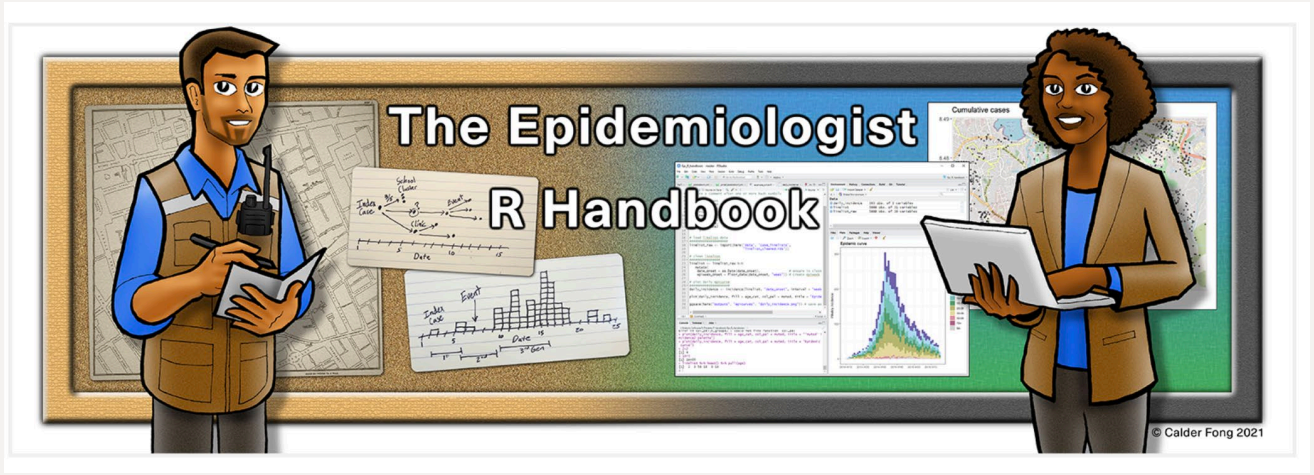
Not reinventing the wheel

- Adapting existing open-source resources
- Here are some groups working on similar problems
- Our content is also open source (CC BY SA)

#PSYTEACHR REPRODUCIBLE RESEARCH

DATA SKILLS | ANALYSIS | STAT MODELS | QUANT FUN | REPRO RES | PSYTEACHR GLOSSARY

R Studio Education



Who are we?

Geneva and Oxford: expertise across epidemiological spectrum

UNIVERSITÉ DE GENÈVE
FACULTY OF MEDICINE
INSTITUTE OF GLOBAL HEALTH

OXFORD POPULATION HEALTH
UNIVERSITY OF OXFORD

Implementation research

IeDEA
International epidemiology Databases to Evaluate AIDS

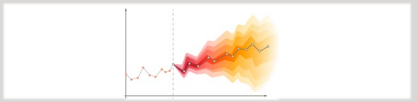
CH-SUR COVID-19 Hospital Based Surveillance

INFO DENGUE

Advanced analytics

Machine learning

Mathematical modelling



Clinical trials

RECOVERY
Randomised Evaluation of COVID-19 Therapy

ASCEND
A Study of Cardiovascular Events iN Diabetes

HPS-4
TIMI 65
ORION-4

Cohort studies and Biobanks

biobank uk
Enabling scientific discoveries that improve human health

CHINA KADOORIE BIOBANK
中国慢性病前瞻性研究

THE MILLION WOMEN STUDY

COVID-19 HIV infectious diseases
cohort studies modelling dengue hepatitis
cost-effectiveness tuberculosis influenza

COVID-19 cancer diabetes
big data ethics cardiovascular
health economics dementia perinatal

Geneva and Oxford: excellent teaching track record

UNIVERSITÉ DE GENÈVE
FACULTY OF MEDICINE
INSTITUTE OF GLOBAL HEALTH

MSc Global Health

UNIVERSITY OF OXFORD
OXFORD POPULATION HEALTH

MSc Global Health Sci & Epi

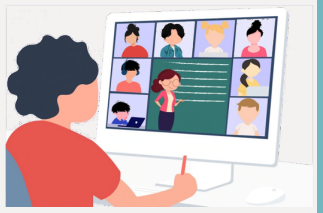
MSc Clinical Trials (online)

Over **500** graduates from diverse setting

How will courses be delivered?

Delivery models

Self-paced courses
Open enrollment
Unlimited capacity

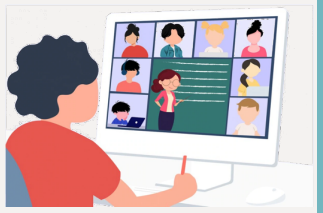


Guided LIVE cohorts
20-30 per cohort

Custom group training
Bespoke, long-term
For govt, companies & NGOs

Funding models

Self-paced courses
Open enrollment
Unlimited capacity

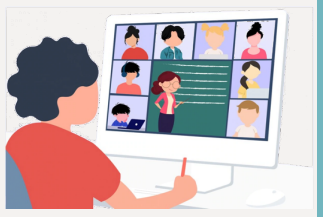


Guided LIVE cohorts
20-30 per cohort

Custom group training
Bespoke, long-term
For govt, companies & NGOs

Funding models

Self-paced courses
Open enrollment
Unlimited capacity



Guided LIVE cohorts
20-30 per cohort

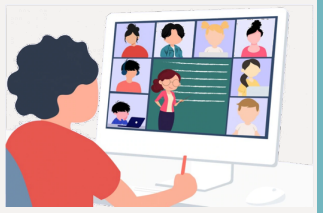
Custom group training
Bespoke, long-term
For govt, companies & NGOs

Seeking course contributors & funders

Funding models

Seeking modest support
Low-cost maintenance

Self-sustainable
"One pays for two"



Initially supported by grants
Fee-charging for financially-abled

Self-paced courses
Open enrollment
Unlimited capacity

Guided LIVE cohorts
20-30 per cohort

Custom group training
Bespoke, long-term
For govt, companies & NGOs

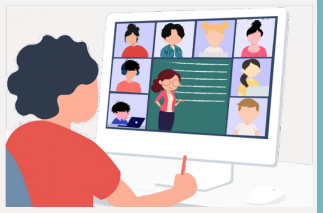
See brochure for more info: tinyurl.com/graph-booklet-2022-05-25

Four-year targets

20,000 students
completing a self-paced course

Self-paced courses
Open enrollment
Unlimited capacity

1000 students
(at least 500 from LMICs)
completing a course



Guided LIVE cohorts
20-30 per cohort

20 customized projects
targeted to LMICs

Custom group training
Bespoke, long-term
For govt, companies & NGOs

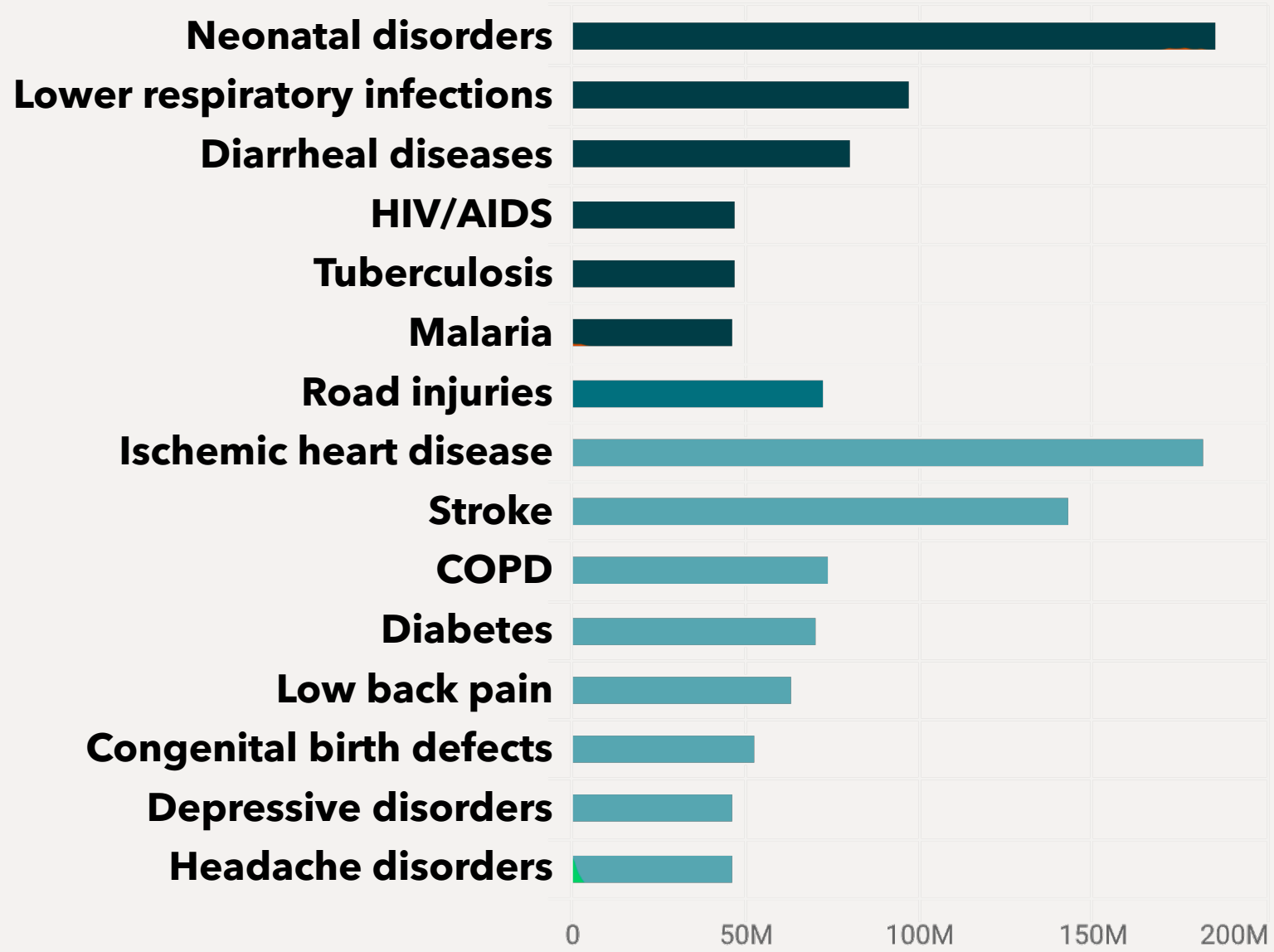
Four-year targets

20,000 students
completing a self-paced course

1000 students
(at least 500 from LMICs)
completing a course

20 customized projects
targeted to LMICs

Top 15 causes of death and disease globally



Four-year targets

20,000 students
completing a self-paced course

1000 students
(at least 500 from LMICs)
completing a course

20 customized projects
targeted to LMICs

Wrapping up

What are we looking for?

- **Users**

Sign up at **thegraphcourses.org**

(Will get update when courses deployed later this year)

- **Course contributors**
- **Partners and funders**

Reach out at **contribute@thegraphnetwork.org**

(you'll all receive an email with these details)

Summary

Public health data science has a huge **return on investment**

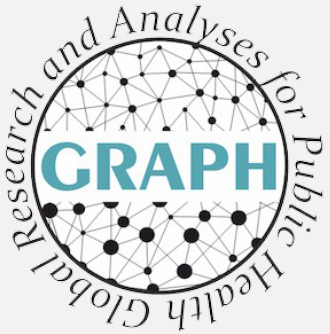
“Data-driven decisions save lives”

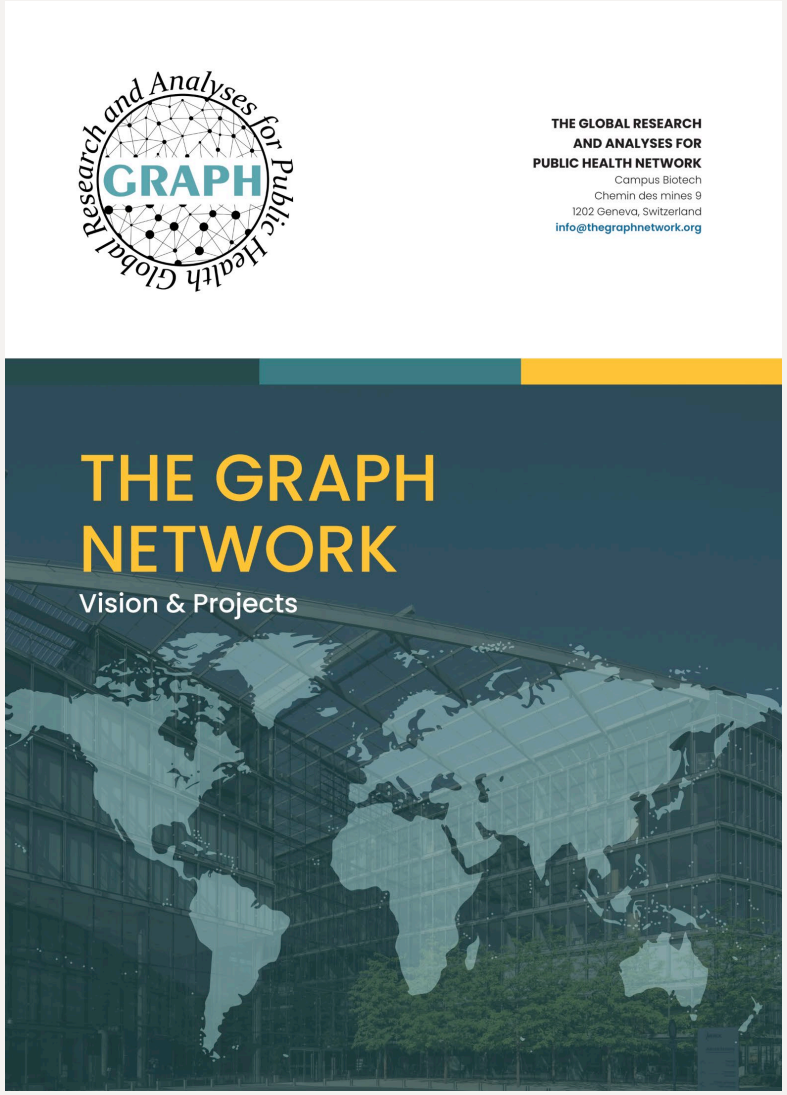
Need for **affordable, high-quality, epi-focused** training for public health analysts in LMICs

We’re working to fill this need, and **we need your support**!

Together, let's train the next generation of leaders in health data science

Thank you!



Brochure link

tinyurl.com/graph-booklet-2022-05-25