

AMMnet meeting, Dakar
2025

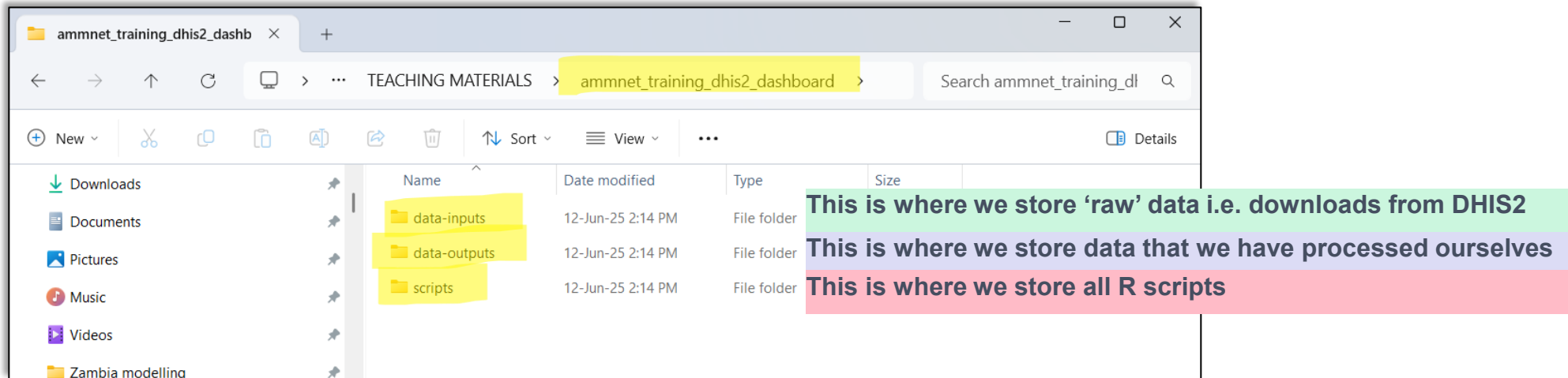
Developing data visualization dashboards in R

Demba Kande
Sarja Jarjusey
Hayley Thompson



Preparation for the session while we wait to start

- 1) ENSURE YOU HAVE **INTERNET ACCESS**
- 2) ENSURE YOU HAVE **R** AND **RSTUDIO** DOWNLOADED
- 3) MAKE A FOLDER IN A SENSIBLE PLACE ON YOUR COMPUTER CALLED SOMETHING LIKE **'ammnet_training_dhis2_dashboard'**
- 4) Within this folder make a three subfolders



- 5) If you know how to create a **project** in R, do so in this same folder (ask a facilitator or neighbor for help if not)

Structure of the workshop

Session 1: How to manually pull data from DHIS2 and clean it ready for use in a dashboard

Session 2: Developing a simple Rshiny dashboard to visualize routine malaria data

Introduction to DHIS2

DHIS2 is an open-source, web-based **health management information system** (HMIS) platform developed by the University of Oslo. It supports data collection, analysis, visualization, and reporting.

It has been adopted by ministries of health in **over 70 countries**, especially across **Africa, Asia, and Latin America**. Commonly used at **national and sub-national levels** for routine health data reporting.

Key features of DHIS2 are:

Captures data from **health facilities and programs** (e.g., malaria, HIV, maternal health)

Enables **real-time dashboards, maps, and charts** for decision-making

Supports **integrated, decentralized data entry** across the health system

Facilitates **monitoring, evaluation, and planning** through tailored analytics

Plan for session 1

- Learn how to access DHIS2
- Explore some existing dashboards to see graphical functionality
- Learn how to make a data query to access the data elements, time period and geographical units of interest
- Download data as a csv
- Read data into R
- Clean data, including name cleaning, visual inspection of outliers/missing data, cross-referencing different variables

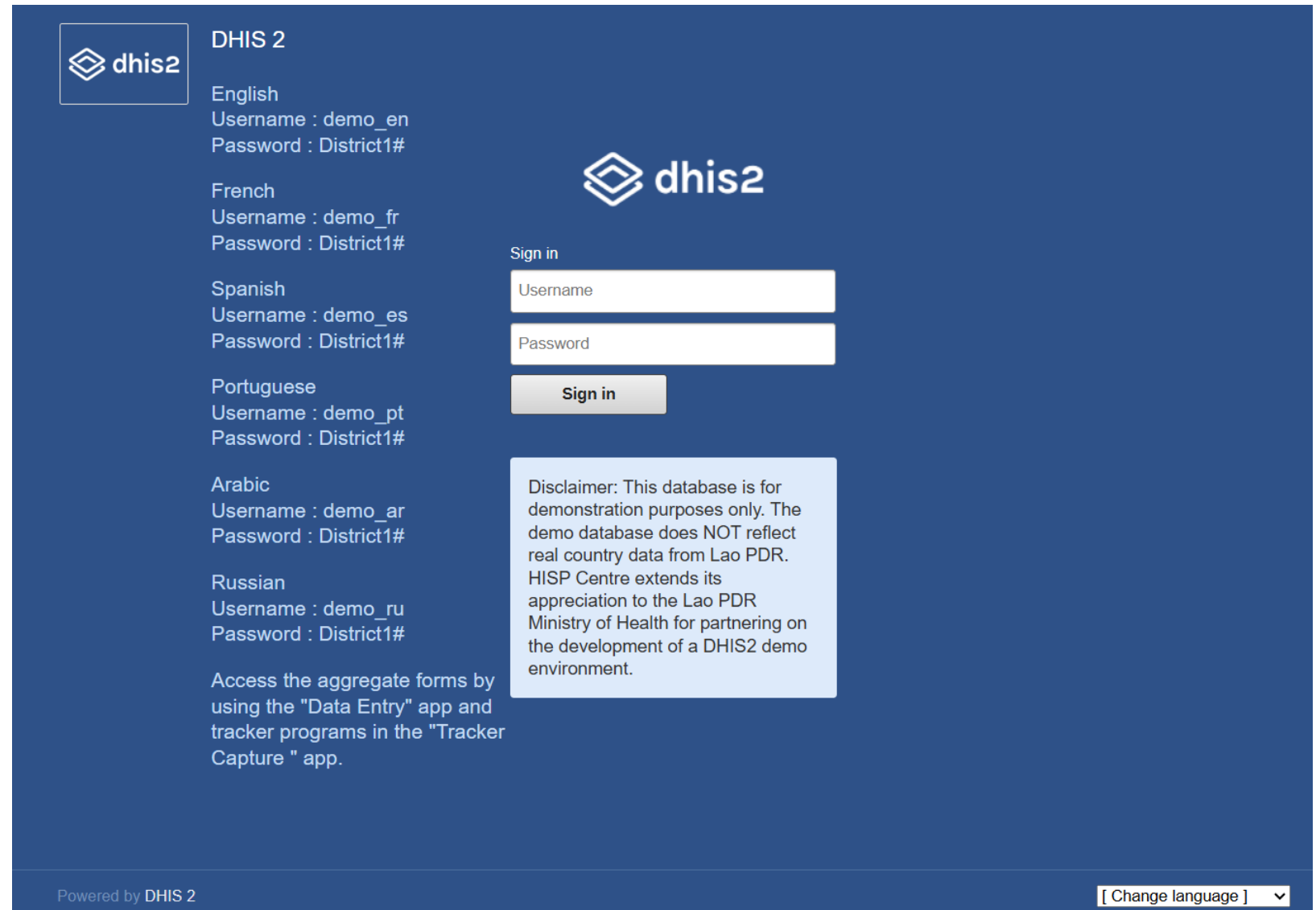
The outcome of this session is a cleaned malaria dataset that we'll use for part 2 (building a shiny dashboard)

Navigating around DHIS2

DHIS2 has a **demonstration database** available based on Lao PDR but containing made-up data

This is what we'll be using in today's workshop

<https://demos.dhis2.org/hmis/>



The screenshot shows the DHIS2 login interface. On the left, there's a sidebar with the DHIS2 logo and a list of languages: English, French, Spanish, Portuguese, Arabic, and Russian. Each language has associated demo credentials (Username: demo_*, Password: District1#). Below the languages, there's a note about accessing aggregate forms. On the right, there's a main login area with the DHIS2 logo, a 'Sign in' label, and input fields for Username and Password. A 'Sign in' button is below the password field. A disclaimer box is also present, stating that the database is for demonstration purposes only and does not reflect real country data from Lao PDR. The footer includes 'Powered by DHIS 2' and a 'Change language' dropdown menu.

DHIS 2

English
Username : demo_en
Password : District1#

French
Username : demo_fr
Password : District1#

Spanish
Username : demo_es
Password : District1#

Portuguese
Username : demo_pt
Password : District1#

Arabic
Username : demo_ar
Password : District1#

Russian
Username : demo_ru
Password : District1#

Access the aggregate forms by using the "Data Entry" app and tracker programs in the "Tracker Capture " app.

Sign in

Username

Password

Sign in

Disclaimer: This database is for demonstration purposes only. The demo database does NOT reflect real country data from Lao PDR. HISP Centre extends its appreciation to the Lao PDR Ministry of Health for partnering on the development of a DHIS2 demo environment.

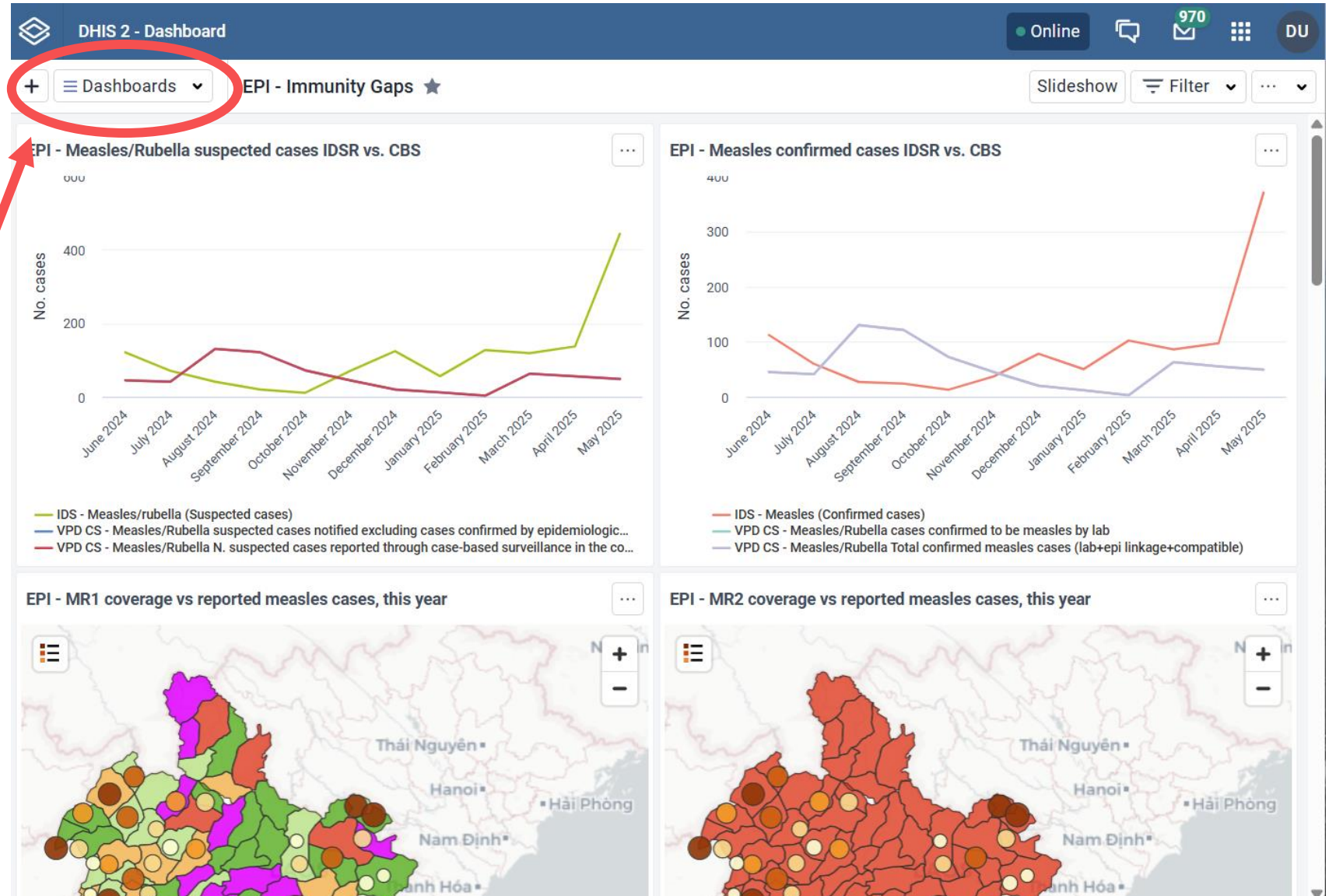
Powered by DHIS 2

[Change language]

Existing dashboards

When you first open this instance of DHIS2 you will see some pre-made dashboards on vaccine-preventable diseases and vaccine coverage rates

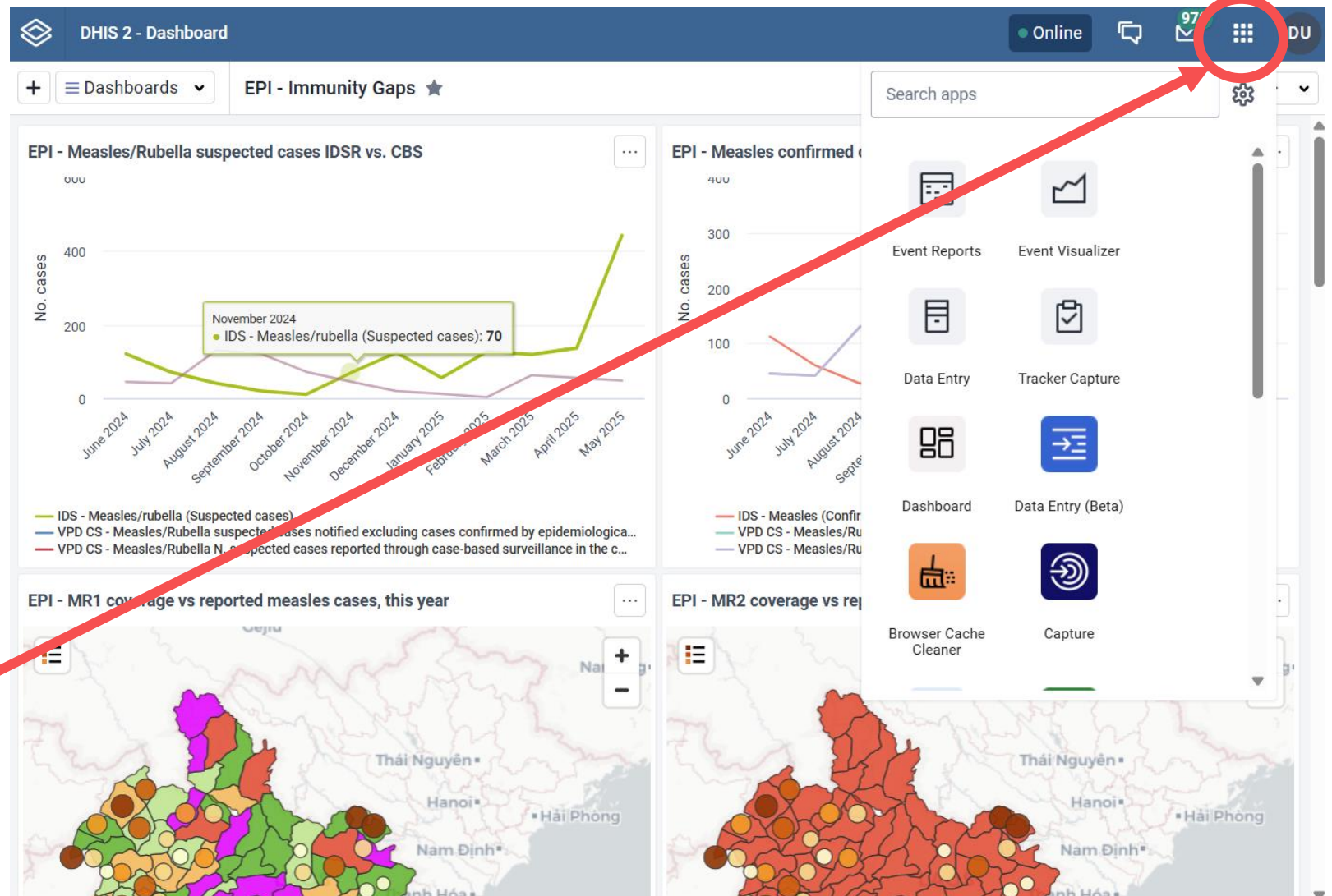
You can explore other pre-made dashboards to see what is possible within DHIS2



Accessing raw data

DHIS2 does have lots of visualization capabilities, however, we often want to extract data from DHIS2 to conduct our own **bespoke analyses** – i.e. a risk stratification or a statistical model looking at intervention impact

To access the raw data, you need to click this 3x3 grid



Finding data visualizer

DHIS2 does have lots of visualization capabilities, however, we often want to extract data from DHIS2 to conduct our own **bespoke analyses** – i.e. a risk stratification or a statistical model looking at intervention impact

Scroll down this menu and select 'data visualizer'

The screenshot displays the DHIS2 Dashboard interface. The top navigation bar includes the DHIS2 logo, the title 'DHIS2 - Dashboard', and user status indicators like 'Online', a notification bell with '970', and a user profile icon labeled 'DU'. Below the navigation bar, the main content area is titled 'EPI - Immunity Gaps' and contains four charts: 'EPI - Measles/Rubella suspected cases IDSR vs. CBS', 'EPI - Measles confirmed cases', 'EPI - MR1 coverage vs reported measles cases, this year', and 'EPI - MR2 coverage vs reported measles cases, this year'. On the right side, a sidebar menu is open, showing various application icons. The 'Data Visualizer' icon, which depicts a green square with a white line graph, is circled in red. A red arrow points from the 'Data Visualizer' icon in the sidebar menu to the text box on the left. Another red arrow points from the top right corner of the dashboard to the user profile icon.

Navigating data visualizer

You should now see this screen – there are several important components of this page

For extracting raw data, we want to use the **pivot table**, but there are options here to present data in different graph formats

This is where you **select the data**, period and organisation unit (i.e. Health facility, ward, district etc).

We'll now look at these one by one on the next slides

The screenshot shows the DHIS 2 - Data Visualizer interface. The top navigation bar includes 'Online', '970', and 'DU'. The main interface is divided into a sidebar on the left and a main content area on the right. The sidebar has three sections: 'MAIN DIMENSIONS' with 'Data', 'Period', and 'Organisation unit' (circled in red); 'OTHER DIMENSIONS' with 'Assigned Categories'; and 'YOUR DIMENSIONS' with a list of various data dimensions. The main content area has a 'Getting started' section with instructions on how to use the visualizer, and a 'Your most viewed charts and tables' section listing several data visualizations.

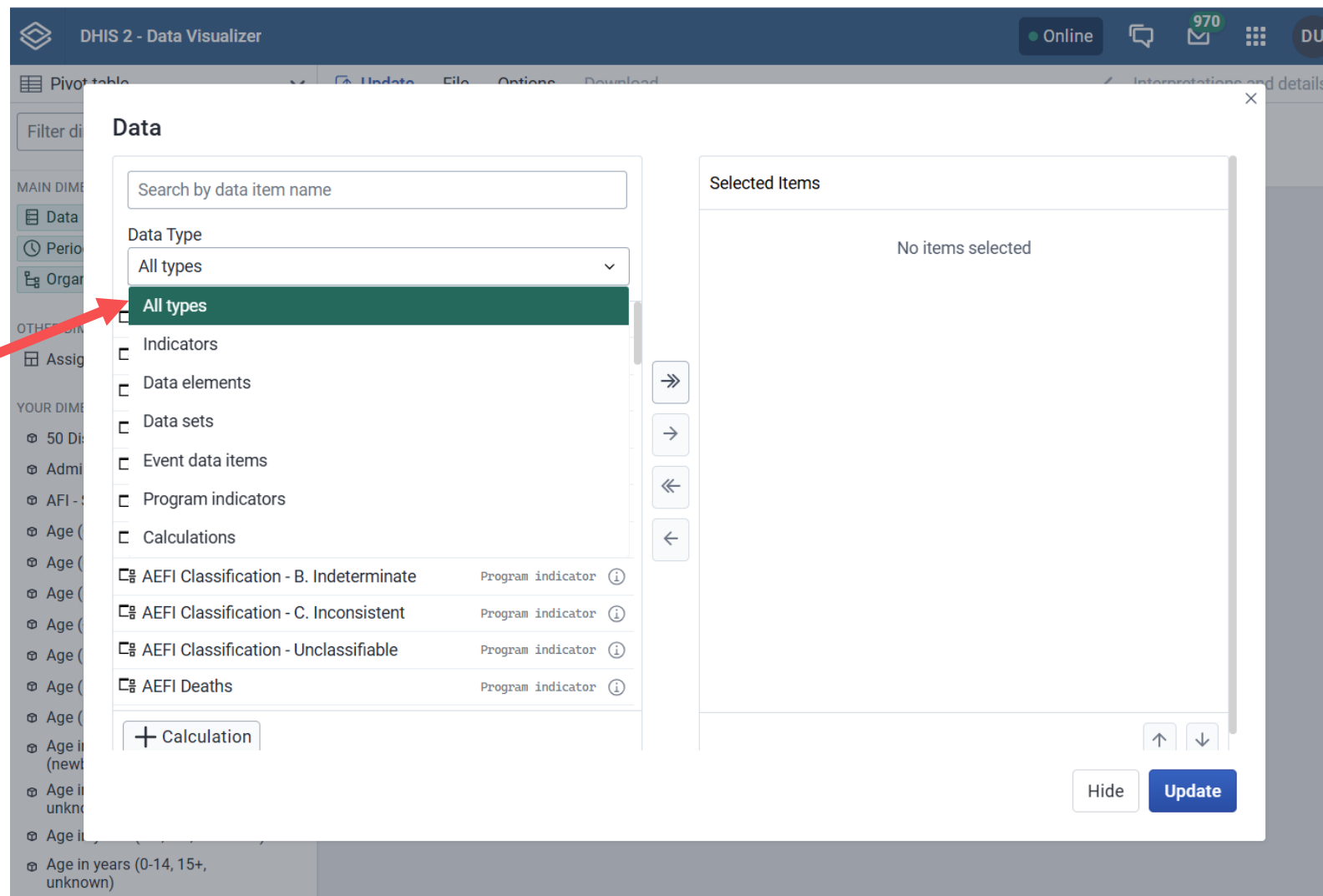
Data dimension

Click on the 'data' option to see this menu

Here you can see there are a range of 'data types'

Today we'll focus on

- **Data elements** – these are the building blocks of data in DHIS2 – its typically a 'count', i.e. malaria cases, or RDTs conducted
- **Indicators** – these are calculated based on a combination of the data elements i.e. Number of tests conducted = # RDT + # microscopy, incidence = cases/population



Finding malaria data (1)

Avoid the malaria indicators with a 'CH' prefix – these relate to cases detected via community case management – today we'll focus on malaria cases reported at health facilities

Firstly, select **data elements**

Next select **malaria burden reduction** in the data element group box

Here we see a wide range of malaria burden variables

For today's example, select, we'll look at **confirmed malaria cases and malaria tests** (RDTs and microscopy) conducted

To select a variable, highlight it as shown here, then press the top 'across' arrow to move it into the **selected items** section

Data

Search by data item name

Data Type
Data elements

Data element group
Malaria burden reduction

Disaggregation
Totals only

- GEN - All-cause inpatients Data element
- GEN - All-cause outpatients Data element
- GEN - Population expected pregnant women Data element
- MAL - ACT issued Data element
- MAL - ACTs courses received by patients Data element
- MAL - All-cause death Data element
- MAL - Confirmed cases (pregnant women) Data element
- MAL - Confirmed cases treated with ACT Data element
- MAL - Confirmed malaria cases Data element**

+ Calculation

Selected Items

No items selected

Hide

Update

Finding malaria data (2)

Data

Data Type

Data elements

Data element group

Malaria burden reduction

Disaggregation

Totals only

- MAL - Plasmodium vivax (RDT)
- MAL - Plasmodium vivax (microscopy)
- MAL - Population at risk for malaria
- MAL - Population with malaria age disaggregation
- MAL - Presumed cases (pregnant women)
- MAL - Presumed malaria cases
- MAL - RDT positive malaria cases (pregnant women)
- MAL - Suspected malaria cases
- MAL - Suspected malaria cases (pregnant women)

Data element

Data element

Data element

Data element

Data element

Data element

Data element

Data element

Data element

+ Calculation



Selected Items

- MAL - Confirmed malaria cases
- MAL - Malaria suspects tested (RDT)
- MAL - Malaria suspects tested with microscopy
- MAL - Microscopy positive malaria cases
- MAL - RDT positive malaria cases

Data element

Data element

Data element

Data element

Data element

↑

↓

Hide

Update

Select these 5 malaria data elements

Finding malaria data (3)

We'll also use some malaria indicators

Another way to find your indicators is simply to type a keyword in the top box

Data

Data Type

Indicators

Indicator group

All groups

✓

CCH - Climatic Suitability for Malaria (All)

Indicator

i

✓

CCH - Climatic Suitability for Malaria (P)

Indicator

i

✓

CCH - Climatic Suitability for Malaria (RH)

Indicator

i

✓

CCH - Climatic Suitability for Malaria (T)

Indicator

i

✓

CH119b - Febrile cases of malaria

Indicator

i

✓

CH120b - Suspected cases of malaria

Indicator

i

✓

CH121 - Malaria RDT positivity rate(%)

Indicator

i

✓

CH121b - Confirmed malaria cases

Indicator

i

✓

CH122 - Malaria cases receiving first line anti-malarial (%)

Indicator

i

+ Calculation

Selected Items

◦

MAL - Confirmed malaria cases

Data element

i

◦

MAL - Malaria suspects tested (RDT)

Data element

i

◦

MAL - Malaria suspects tested with microscopy

Data element

i

◦

MAL - Microscopy positive malaria cases

Data element

i

◦

MAL - RDT positive malaria cases

Data element

i

✓

MAL - Malaria cases tested (total)

Indicator

i

✓

MAL - Confirmed malaria cases (micr + RDT)

Indicator

i

✓

MAL - Confirmed malaria cases per 1000

Indicator

i

Let's also select these 3 indicators

PATH
10A0♦//20

A note on finding the 'right' variables

There are typically a lot of variables that relate to malaria burden, and not all countries regularly report all variables – we normally start by asking a local expert which elements/indicators to use

Failing that, we can manually interrogate the data to see which are reported

Feel free to select a much wider range of malaria variable to explore whether they're all useful, and how they relate to each other

There are also sometimes duplicate variables, for example, this demo DHIS2 has 2 identical indicators for total malaria cases

You can click the circled 'i' Here to get more details on each variables

Data

malaria

Data Type
Indicators

Indicator group
All groups

√ MAL - Percentage of malaria deaths	Indicator	i
√ MAL - Population at risk for malaria	Indicator	i
√ MAL - Presumed malaria cases	Indicator	i

√ MAL - Presumed malaria cases

Name MAL - Presumed malaria cases

Numerator description Presumed cases

Numerator expression MAL - Suspected malaria cases - (MAL - Microscopy positive malaria cases + MAL - RDT positive malaria cases)

Denominator description 1

Denominator expression 1

Annualized No

Indicator type Numerator only (number), 1

Group membership

- Malaria burden reduction
- [OLD] Malaria

Selecting time periods

You can either select a period relative to today (i.e. last 3 months), or you can select fixed periods

In this example, we're going to look at monthly data from 2024 only

First select 'fixed periods' and then select 'monthly' and '2024'

Period

Relative periods Fixed periods

Period type Year

Monthly 2024

Selected Periods

- January 2024
- February 2024
- March 2024
- April 2024
- May 2024
- June 2024
- July 2024
- August 2024
- September 2024
- October 2024
- November 2024
- December 2024

Click the double arrow to bring all months over

Sometimes there is a period auto selected at the top – make sure to remove it by clicking the single right arrow

Selecting organization unit

At the moment we're looking at national totals, but we can also look at the data at smaller spatial units – down to the health facility level

Select organization unit

We can either select all the provinces manually by selecting the boxes – this method is best when we just want to look at a subset of provinces, districts etc.

DHIS 2 - Data Visualizer

Online

Pivot table

Filter dimensions

MAIN DIMENSIONS

- Data
- Period
- Organisation unit

OTHER DIMENSIONS

- Assigned Categories

YOUR DIMENSIONS

- 50 Districts of EPI (P-D)
- Administrative levels c
- Age (0-59 months)
- Age (0-59,60+)
- Age (6-59 months)
- Age (<1- 30+ years)
- Age (GNARF)
- Age (Rehabilitation)
- Age (surveillance)
- Age in days, months a (newborns, children, U
- Age in weeks (28-37, 3 unknown)

Organisation unit

☐ User organisation unit ☐ User sub-units ☐ User sub-x2-units

- ☒ Lao PDR
 - ☒ 01 Vientiane Capital
 - ☒ 02 Phongsali
 - ☒ 03 Louangnamtha
 - ☒ 04 Oudomxai
 - ☐ 05 Bokeo
 - ☐ 06 Louangphabang
 - ☐ 07 Houaphan
 - ☐ 08 Xainyabouli
 - ☐ 09 Xiangkhouang
 - ☐ 10 Vientiane
 - ☐ 11 Bolikhamxai
 - ☐ 12 Khammouan
 - ☐ 13 Savannakhet
 - ☐ 14 Salavan
 - ☐ 15 Xekong

Select a level

Selected: 4 org units Deselect

Organisation unit

☐ User organisation unit ☒ User sub-units ☐ User sub-x2-units

- ☐ Lao PDR
 - ☐ 01 Vientiane Capital
 - ☐ 02 Phongsali
 - ☐ 03 Louangnamtha
 - ☐ 04 Oudomxai
 - ☐ 05 Bokeo
 - ☐ 06 Louangphabang
 - ☐ 07 Houaphan
 - ☐ 08 Xainyabouli
 - ☐ 09 Xiangkhouang
 - ☐ 10 Vientiane
 - ☐ 11 Bolikhamxai
 - ☐ 12 Khammouan
 - ☐ 13 Savannakhet
 - ☐ 14 Salavan
 - ☐ 15 Xekong

Select a level Select a group

Selected: User sub-units Deselect all

Hide Add to Columns

A quicker option is to just select **user sub-units**

Reorienting and downloading data (1)

The data are returned in a wide format, with 18 columns (one per province) for each indicator

For future data manipulation in R, it's preferable to have the data in a **long format**

This is a data format where each row represents a single observation, and each column represents a variable or feature

To make data **long format** we can drag these green boxes all to the **rows' area** (see next slide)

DHIS 2 - Data Visualizer Online 970

Pivot table Update File Options Download Interpretations and

Filter dimensions

Columns Data 8 Organisation unit 1 Filter

Rows Period 12

MAIN DIMENSIONS

- Data
- Period
- Organisation unit

OTHER DIMENSIONS

- Assigned Categories

YOUR DIMENSIONS

- 50 Districts of EPI (P-DL18)
- Administrative levels of care
- AFI - Screening form age group
- Age (0-59 months)
- Age (0-59,60+)
- Age (6-59 months)
- Age (<1- 30+ years)
- Age (GNAP)
- Age (Rehabilitation)
- Age (surveillance)
- Age in days, months and years (newborns, children, Unk)
- Age in weeks (28-37, 38-41, 42+, unknown)
- Age in years (0-4, 5-9, unknown)
- Age in years (0-14, 15+, unknown)
- Age in years (0-4, 5-14, 15+)

Unsaved visualization

MAL - Confirmed malaria cases (micr + RDT)

	01 Vientiane Capital	02 Phongsali	03 Louangnamtha	04 Oudomxai	05 Bokeo	06 Louangphabang	07 Houaphan	08 Xainyabouli	09 Xiangkhouang
January 2024	12	16	1			42	92	12	
February 2024	14	26				21	97	20	
March 2024	24	30		0	50	19	73	17	
April 2024	28	38	2		51	24	60	28	
May 2024	21	43	4	2	52	29	36	28	
June 2024	12	26			64	32	60	42	
July 2024	19	28	3	2	64	30	83	26	
August 2024	45	59	3	7	8	16	87	53	
September 2024	41	40	2	1		14	74	26	
October 2024	16	41		2		9	53	31	
November 2024	54	34	12		1	16	66	57	
December 2024	36	41	4	16		34	62	68	

Useful resources for learning about **long** vs. **wide** data

<https://tavareshugo.github.io/r-intro-tidyverse-gapminder/09-reshaping/index.html>

Reorienting and downloading data (2)

You can change to order of these to see what is most logical for your data

You'll need to **update** the table each time you modify it

DHIS 2 - Data Visualizer

Pivot table ☒ [Update](#) File Options Download

Filter dimensions

Columns

Rows Organisation unit 1 ... Data 8 ... Period 12 ...

MAIN DIMENSIONS

- Data
- Period
- Organisation unit

OTHER DIMENSIONS

- Assigned Categories

YOUR DIMENSIONS

- 50 Districts of EPI (P-DLI8)
- Administrative levels of care
- AFI - Screening form age group
- Age (0-59 months)
- Age (0-59,60+)
- Age (6-59 months)
- Age (<1- 30+ years)
- Age (GNARF)
- Age (Rehabilitation)
- Age (surveillance)
- Age in days, months and years (newborns, children, Unk)
- Age in weeks (28-37, 38-41, 42+, unknown)
- Age in years (0-4, 5-9, unknown)
- Age in years (0-14, 15+, unknown)
- Age in years (0-4, 5-14, 15+)

Unsaved visualization

01 Vientiane Capital	MAL - Confirmed malaria cases (micr + RDT)	January 2024	12
		February 2024	14
		March 2024	24
		April 2024	28
		May 2024	21
		June 2024	12
		July 2024	19
		August 2024	45
		September 2024	41
		October 2024	16
		November 2024	54
		December 2024	36
01 Vientiane Capital	MAL - Confirmed malaria cases per 1000	January 2024	1.26
		February 2024	1.58
		March 2024	2.53
		April 2024	3.05
		May 2024	2.21
		June 2024	1.31
		July 2024	2
		August 2024	4.74
		September 2024	4.46
		October 2024	1.68
		November 2024	5.87
		December 2024	3.79
		January 2024	102

Reorienting and downloading data (3)

We can now **download** our data as a csv

The easiest data format to work with is **plain data source** -> **CSV** -> **Name**

DHIS 2 - Data Visualizer

Pivot table

Filter dimensions

MAIN DIMENSIONS

- Data
- Period
- Organisation unit

OTHER DIMENSIONS

- Assigned Categories

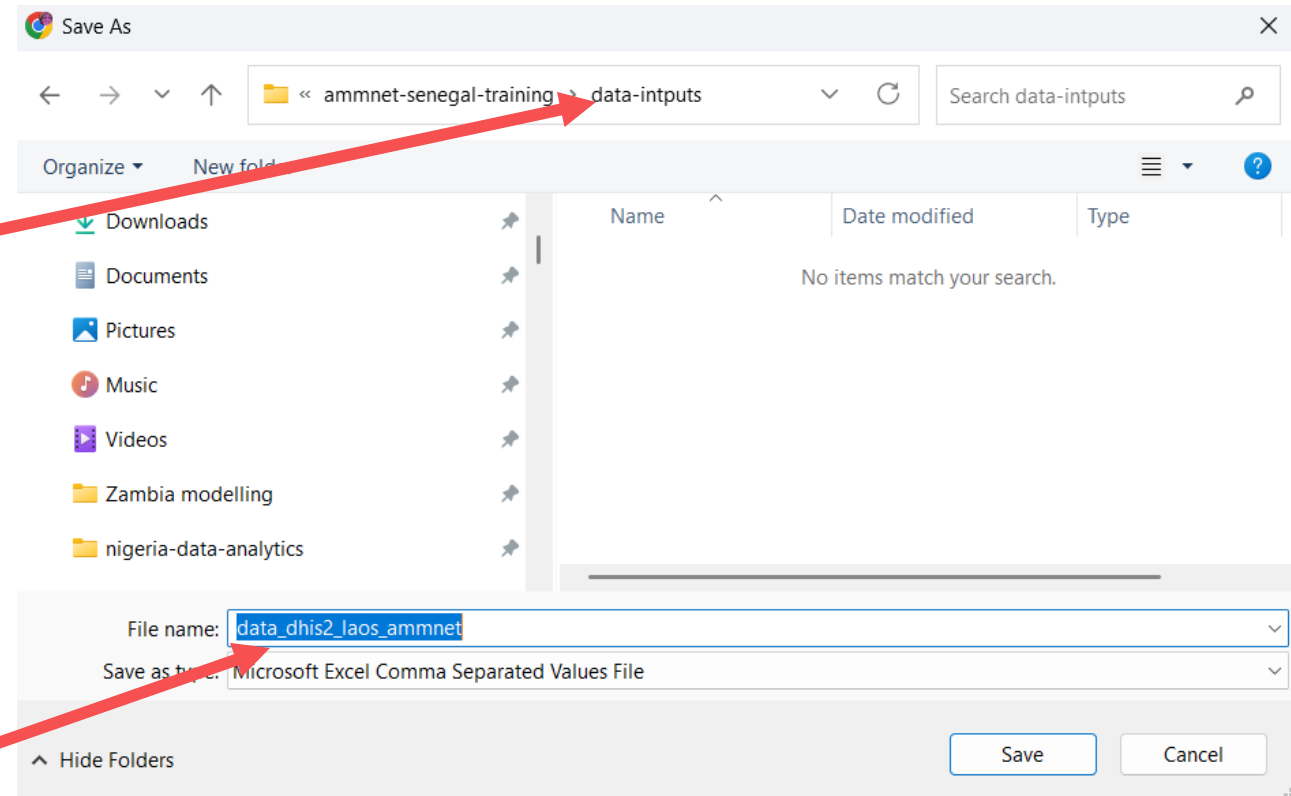
YOUR DIMENSIONS

- 50 Districts of EPI (P-DLI8)
- Administrative levels of care
- AFI - Screening form age group
- Age (0-59 months)
- Age (0-59,60+)
- Age (6-59 months)
- Age (<1- 30+ years)
- Age (GNARF)
- Age (Rehabilitation)
- Age (surveillance)
- Age in days, months and years (newborns, children, Unk)
- Age in weeks (28-37, 38-41, 42+, unknown)
- Age in years (0-4, 5-9, unknown)
- Age in years (0-14, 15+, unknown)
- Age in years (0-4, 5-14, 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85-94, 95-104, 105-114, 115-124, 125-134, 135-144, 145-154, 155-164, 165-174, 175-184, 185-194, 195-204, 205-214, 215-224, 225-234, 235-244, 245-254, 255-264, 265-274, 275-284, 285-294, 295-304, 305-314, 315-324, 325-334, 335-344, 345-354, 355-364, 365-374, 375-384, 385-394, 395-404, 405-414, 415-424, 425-434, 435-444, 445-454, 455-464, 465-474, 475-484, 485-494, 495-504, 505-514, 515-524, 525-534, 535-544, 545-554, 555-564, 565-574, 575-584, 585-594, 595-604, 605-614, 615-624, 625-634, 635-644, 645-654, 655-664, 665-674, 675-684, 685-694, 695-704, 705-714, 715-724, 725-734, 735-744, 745-754, 755-764, 765-774, 775-784, 785-794, 795-804, 805-814, 815-824, 825-834, 835-844, 845-854, 855-864, 865-874, 875-884, 885-894, 895-904, 905-914, 915-924, 925-934, 935-944, 945-954, 955-964, 965-974, 975-984, 985-994, 995-1004, 1005-1014, 1015-1024, 1025-1034, 1035-1044, 1045-1054, 1055-1064, 1065-1074, 1075-1084, 1085-1094, 1095-1104, 1105-1114, 1115-1124, 1125-1134, 1135-1144, 1145-1154, 1155-1164, 1165-1174, 1175-1184, 1185-1194, 1195-1204, 1205-1214, 1215-1224, 1225-1234, 1235-1244, 1245-1254, 1255-1264, 1265-1274, 1275-1284, 1285-1294, 1295-1304, 1305-1314, 1315-1324, 1325-1334, 1335-1344, 1345-1354, 1355-1364, 1365-1374, 1375-1384, 1385-1394, 1395-1404, 1405-1414, 1415-1424, 1425-1434, 1435-1444, 1445-1454, 1455-1464, 1465-1474, 1475-1484, 1485-1494, 1495-1504, 1505-1514, 1515-1524, 1525-1534, 1535-1544, 1545-1554, 1555-1564, 1565-1574, 1575-1584, 1585-1594, 1595-1604, 1605-1614, 1615-1624, 1625-1634, 1635-1644, 1645-1654, 1655-1664, 1665-1674, 1675-1684, 1685-1694, 1695-1704, 1705-1714, 1715-1724, 1725-1734, 1735-1744, 1745-1754, 1755-1764, 1765-1774, 1775-1784, 1785-1794, 1795-1804, 1805-1814, 1815-1824, 1825-1834, 1835-1844, 1845-1854, 1855-1864, 1865-1874, 1875-1884, 1885-1894, 1895-1904, 1905-1914, 1915-1924, 1925-1934, 1935-1944, 1945-1954, 1955-1964, 1965-1974, 1975-1984, 1985-1994, 1995-2004, 2005-2014, 2015-2024, 2025-2034, 2035-2044, 2045-2054, 2055-2064, 2065-2074, 2075-2084, 2085-2094, 2095-2104, 2105-2114, 2115-2124, 2125-2134, 2135-2144, 2145-2154, 2155-2164, 2165-2174, 2175-2184, 2185-2194, 2195-2204, 2205-2214, 2215-2224, 2225-2234, 2235-2244, 2245-2254, 2255-2264, 2265-2274, 2275-2284, 2285-2294, 2295-2304, 2305-2314, 2315-2324, 2325-2334, 2335-2344, 2345-2354, 2355-2364, 2365-2374, 2375-2384, 2385-2394, 2395-2404, 2405-2414, 2415-2424, 2425-2434, 2435-2444, 2445-2454, 2455-2464, 2465-2474, 2475-2484, 2485-2494, 2495-2504, 2505-2514, 2515-2524, 2525-2534, 2535-2544, 2545-2554, 2555-2564, 2565-2574, 2575-2584, 2585-2594, 2595-2604, 2605-2614, 2615-2624, 2625-2634, 2635-2644, 2645-2654, 2655-2664, 2665-2674, 2675-2684, 2685-2694, 2695-2704, 2705-2714, 2715-2724, 2725-2734, 2735-2744, 2745-2754, 2755-2764, 2765-2774, 2775-2784, 2785-2794, 2795-2804, 2805-2814, 2815-2824, 2825-2834, 2835-2844, 2845-2854, 2855-2864, 2865-2874, 2875-2884, 2885-2894, 2895-2904, 2905-2914, 2915-2924, 2925-2934, 2935-2944, 2945-2954, 2955-2964, 2965-2974, 2975-2984, 2985-2994, 2995-3004, 3005-3014, 3015-3024, 3025-3034, 3035-3044, 3045-3054, 3055-3064, 3065-3074, 3075-3084, 3085-3094, 3095-3104, 3105-3114, 3115-3124, 3125-3134, 3135-3144, 3145-3154, 3155-3164, 3165-3174, 3175-3184, 3185-3194, 3195-3204, 3205-3214, 3215-3224, 3225-3234, 3235-3244, 3245-3254, 3255-3264, 3265-3274, 3275-3284, 3285-3294, 3295-3304, 3305-3314, 3315-3324, 3325-3334, 3335-3344, 3345-3354, 3355-3364, 3365-3374, 3375-3384, 3385-3394, 3395-3404, 3405-3414, 3415-3424, 3425-3434, 3435-3444, 3445-3454, 3455-3464, 3465-3474, 3475-3484, 3485-3494, 3495-3504, 3505-3514, 3515-3524, 3525-3534, 3535-3544, 3545-3554, 3555-3564, 3565-3574, 3575-3584, 3585-3594, 3595-3604, 3605-3614, 3615-3624, 3625-3634, 3635-3644, 3645-3654, 3655-3664, 3665-3674, 3675-3684, 3685-3694, 3695-3704, 3705-3714, 3715-3724, 3725-3734, 3735-3744, 3745-3754, 3755-3764, 3765-3774, 3775-3784, 3785-3794, 3795-3804, 3805-3814, 3815-3824, 3825-3834, 3835-3844, 3845-3854, 3855-3864, 3865-3874, 3875-3884, 3885-3894, 3895-3904, 3905-3914, 3915-3924, 3925-3934, 3935-3944, 3945-3954, 3955-3964, 3965-3974, 3975-3984, 3985-3994, 3995-4004, 4005-4014, 4015-4024, 4025-4034, 4035-4044, 4045-4054, 4055-4064, 4065-4074, 4075-4084, 4085-4094, 4095-4104, 4105-4114, 4115-4124, 4125-4134, 4135-4144, 4145-4154, 4155-4164, 4165-4174, 4175-4184, 4185-4194, 4195-4204, 4205-4214, 4215-4224, 4225-4234, 4235-4244, 4245-4254, 4255-4264, 4265-4274, 4275-4284, 4285-4294, 4295-4304, 4305-4314, 4315-4324, 4325-4334, 4335-4344, 4345-4354, 4355-4364, 4365-4374, 4375-4384, 4385-4394, 4395-4404, 4405-4414, 4415-4424, 4425-4434, 4435-4444, 4445-4454, 4455-4464, 4465-4474, 4475-4484, 4485-4494, 4495-4504, 4505-4514, 4515-4524, 4525-4534, 4535-4544, 4545-4554, 4555-4564, 4565-4574, 4575-4584, 4585-4594, 4595-4604, 4605-4614, 4615-4624, 4625-4634, 4635-4644, 4645-4654, 4655-4664, 4665-4674, 4675-4684, 4685-4694, 4695-4704, 4705-4714, 4715-4724, 4725-4734, 4735-4744, 4745-4754, 4755-4764, 4765-4774, 4775-4784, 4785-4794, 4795-4804, 4805-4814, 4815-4824, 4825-4834, 4835-4844, 4845-4854, 4855-4864, 4865-4874, 4875-4884, 4885-4894, 4895-4904, 4905-4914, 4915-4924, 4925-4934, 4935-4944, 4945-4954, 4955-4964, 4965-4974, 4975-4984, 4985-4994, 4995-5004, 5005-5014, 5015-5024, 5025-5034, 5035-5044, 5045-5054, 5055-5064, 5065-5074, 5075-5084, 5085-5094, 5095-5104, 5105-5114, 5115-5124, 5125-5134, 5135-5144, 5145-5154, 5155-5164, 5165-5174, 5175-5184, 5185-5194, 5195-5204, 5205-5214, 5215-5224, 5225-5234, 5235-5244, 5245-5254, 5255-5264, 5265-5274, 5275-5284, 5285-5294, 5295-5304, 5305-5314, 5315-5324, 5325-5334, 5335-5344, 5345-5354, 5355-5364, 5365-5374, 5375-5384, 5385-5394, 5395-5404, 5405-5414, 5415-5424, 5425-5434, 5435-5444, 5445-5454, 5455-5464, 5465-5474, 5475-5484, 5485-5494, 5495-5504, 5505-5514, 5515-5524, 5525-5534, 5535-5544, 5545-5554, 5555-5564, 5565-5574, 5575-5584, 5585-5594, 5595-5604, 5605-5614, 5615-5624, 5625-5634, 5635-5644, 5645-5654, 5655-5664, 5665-5674, 5675-5684, 5685-5694, 5695-5704, 5705-5714, 5715-5724, 5725-5734, 5735-5744, 5745-5754, 5755-5764, 5765-5774, 5775-5784, 5785-5794, 5795-5804, 5805-5814, 5815-5824, 5825-5834, 5835-5844, 5845-5854, 5855-5864, 5865-5874, 5875-5884, 5885-5894, 5895-5904, 5905-5914, 5915-5924, 5925-5934, 5935-5944, 5945-5954, 5955-5964, 5965-5974, 5975-5984, 5985-5994, 5995-6004, 6005-6014, 6015-6024, 6025-6034, 6035-6044, 6045-6054, 6055-6064, 6065-6074, 6075-6084, 6085-6094, 6095-6104, 6105-6114, 6115-6124, 6125-6134, 6135-6144, 6145-6154, 6155-6164, 6165-6174, 6175-6184, 6185-6194, 6195-6204, 6205-6214, 6215-6224, 6225-6234, 6235-6244, 6245-6254, 6255-6264, 6265-6274, 6275-6284, 6285-6294, 6295-6304, 6305-6314, 6315-6324, 6325-6334, 6335-6344, 6345-6354, 6355-6364, 6365-6374, 6375-6384, 6385-6394, 6395-6404, 6405-6414, 6415-6424, 6425-6434, 6435-6444, 6445-6454, 6455-6464, 6465-6474, 6475-6484, 6485-6494, 6495-6504, 6505-6514, 6515-6524, 6525-6534, 6535-6544, 6545-6554, 6555-6564, 6565-6574, 6575-6584, 6585-6594, 6595-6604, 6605-6614, 6615-6624, 6625-6634, 6635-6644, 6645-6654, 6655-6664, 6665-6674, 6675-6684, 6685-6694, 6695-6704, 6705-6714, 6715-6724, 6725-6734, 6735-6744, 6745-6754, 6755-6764, 6765-6774, 6775-6784, 6785-6794, 6795-6804, 6805-6814, 6815-6824, 6825-6834, 6835-6844, 6845-6854, 6855-6864, 6865-6874, 6875-6884, 6885-6894, 6895-6904, 6905-6914, 6915-6924, 6925-6934, 6935-6944, 6945-6954, 6955-6964, 6965-6974, 6975-6984, 6985-6994, 6995-7004, 7005-7014, 7015-7024, 7025-7034, 7035-7044, 7045-7054, 7055-7064, 7065-7074, 7075-7084, 7085-7094, 7095-7104, 7105-7114, 7115-7124, 7125-7134, 7135-7144, 7145-7154, 7155-7164, 7165-7174, 7175-7184, 7185-7194, 7195-7204, 7205-7214, 7215-7224, 7225-7234, 7235-7244, 7245-7254, 7255-7264, 7265-7274, 7275-7284, 7285-7294, 7295-7304, 7305-7314, 7315-7324, 7325-7334, 7335-7344, 7345-7354, 7355-7364, 7365-7374, 7375-7384, 7385-7394, 7395-7404, 7405-7414, 7415-7424, 7425-7434, 7435-7444, 7445-7454, 7455-7464, 7465-7474, 7475-7484, 7485-7494, 7495-7504, 7505-7514, 7515-7524, 7525-7534, 7535-7544, 7545-7554, 7555-7564, 7565-7574, 7575-7584, 7585-7594, 7595-7604, 7605-7614, 7615-7624, 7625-7634, 7635-7644, 7645-7654, 7655-7664, 7665-7674, 7675-7684, 7685-7694, 7695-7704, 7705-7714, 7715-7724, 7725-7734, 7735-7744, 7745-7754, 7755-7764, 7765-7774, 7775-7784, 7785-7794, 7795-7804, 7805-7814, 7815-7824, 7825-7834, 7835-7844, 7845-7854, 7855-7864, 7865-7874, 7875-7884, 7885-7894, 7895-7904, 7905-7914, 7915-7924, 7925-7934, 7935-7944, 7945-7954, 7955-7964, 7965-7974, 7975-7984, 7985-7994, 7995-8004, 8005-8014, 8015-8024, 8025-8034, 8035-8044, 8045-8054, 8055-8064, 8065-8074, 8075-8084, 8085-8094, 8095-8104, 8105-8114, 8115-8124, 8125-8134, 8135-8144, 8145-8154, 8155-8164, 8165-8174, 8175-8184, 8185-8194, 8195-8204, 8205-8214, 8215-8224, 8225-8234, 8235-8244, 8245-8254, 8255-8264, 8265-8274, 8275-8284, 8285-8294, 8295-8304, 8305-8314, 8315-8324, 8325-8334, 8335-8344, 8345-8354, 8355-8364, 8365-8374, 8375-8384, 8385-8394, 8395-8404, 8405-8414, 8415-8424, 8425-8434, 8435-8444, 8445-8454, 8455-8464, 8465-8474, 8475-8484, 8485-8494, 8495-8504, 8505-8514, 8515-8524, 8525-8534, 8535-8544, 8545-8554, 8555-8564, 8565-8574, 8575-8584, 8585-8594, 8595-8604, 8605-8614, 8615-8624, 8625-8634, 8635-8644, 8645-8654, 8655-8664, 8665-8674, 8675-8684, 8685-8694, 8695-8704, 8705-8714, 8715-8724, 8725-8734, 8735-8744, 8745-8754, 8755-8764, 8765-8774, 8775-8784, 8785-8794, 8795-8804, 8805-8814, 8815-8824, 8825-8834, 8835-8844, 8845-8854, 8855-8864, 8865-8874, 8875-8884, 8885-8894, 8895-8904, 8905-8914, 8915-8924, 8925-8934, 8935-8944, 8945-8954, 8955-8964, 8965-8974, 8975-8984, 8985-8994, 8995-9004, 9005-9014, 9015-9024, 9025-9034, 9035-9044, 9045-9054, 9055-9064, 9065-9074, 9075-9084, 9085-9094, 9095-9104, 9105-9114, 9115-9124, 9125-9134, 9135-9144, 9145-9154, 9155-9164, 9165-9174, 9175-9184, 9185-9194, 9195-9204, 9205-9214, 9215-9224, 9225-9234, 9235-9244, 9245-9254, 9255-9264, 9265-9274, 9275-9284, 9285-9294, 9295-9304, 9305-9314, 9315-9324, 9325-9334, 9335-9344, 9345-9354, 9355-9364, 9365-9374, 9375-9384, 9385-9394, 9395-9404, 9405-9414, 9415-9424, 9425-9434, 9435-9444, 9445-9454, 9455-9464, 9465-9474, 9475-9484, 9485-9494, 9495-9504, 9505-9514, 9515-9524, 9525-9534, 9535-9544, 9545-9554, 9555-9564, 9565-9574, 9575-9584, 9585-9594, 9595-9604, 9605-9614, 9615-9624, 9625-9634, 9635-9644, 9645-9654, 9655-9664, 9665-9674, 9675-9684, 9685-9694, 9695-9704, 9705-9714, 9715-9724, 9725-9734, 9735-9744, 9745-9754, 9755-9764, 9765-9774, 9775-9784, 9785-9794, 9795-9804, 9805-9814, 9815-9824, 9825-9834, 9835-9844, 9845-9854, 9855-9864, 9865-9874, 9875-9884, 9885-9894, 9895-9904, 9905-9914, 9915-9924, 9925-9934, 9935-9944, 9945-9954, 9955-9964, 9965-9974, 9975-9984, 9985-9994, 9995-10004, 10005-10014, 10015-10024, 10025-10034, 10035-10044, 10045-10054, 10055-10064, 10065-10074, 10075-10084, 10085-10094, 10095-10104, 10105-10114, 10115-10124, 10125-10134, 10135-10144, 10145-10154, 10155-10164, 10165-10174, 10175-10184, 10185-10194, 10195-10204, 10205-10214, 10215-10224, 10225-10234, 10235-10244, 10245-10254, 10255-10264, 10265-10274, 10275-10284, 10285-10294, 10295-10304, 10305-10314, 10315-10324, 10325-10334, 10335-10344, 10345-10354, 10355-10364, 10365-10374, 10375-10384, 10385-10394, 10395-10404, 10405-10414, 10415-10424, 10425-10434, 10435-10444, 10445-10454, 10455-10464, 10465-10474, 10475-10484, 10485-10494, 10495-10504, 10505-10514, 10515-10524, 10525-10534, 10535-10544, 10545-10554, 10555-10564, 10565-10574, 10575-10584, 10585-10594, 10595-10604, 10605-10614, 10615-10624, 10625-10634, 10635-10644, 10645-10654, 10655-10664, 10665-10674, 10675-10684, 10685-10694, 10695-10704, 10705-10714, 10715-10724, 10725-10734, 10735-10744, 10745-10754, 10755-10764, 10765-10774, 10775-10784, 10785-10794, 10795-10804, 10805-10814, 10815-10824, 10825-10834, 10835-10844, 10845-10854, 10855-1

Reorienting and downloading data (4)

Save your data in your **data-inputs** folder that you created at the start of the session

Rename your file to something informative – I've chosen **data_dhis2_laos_ammnet**



Reading the data in R and cleaning

For this section we just need 2 packages tidyverse and janitor, if you don't already have them, you can install with the commands

```
install.packages("tidyverse")
```

```
install.packages("janitor")
```

Step 1: read in the data, clean the names and look at the data

```
library(tidyverse)
```

```
library(janitor)
```

```
laos_dat <- read_csv("data-inputs/data_dhis2_laos_ammnet.csv") %>%
```

```
  clean_names()
```

```
head(laos_dat)
```


Cleaning (1)

These data names look long and cumbersome – let's shorten them once we decided which ones to use

The download returned the national level data, which we don't want – let's filter this out

These columns don't look useful, so let's remove them from our output

A tibble: 6 x 9

	data	period	organisation	unit	value	numerator	denominator	factor	multiplier	divisor
	<chr>	<chr>	<chr>		<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	MAL - Malaria cas...	Janua...	Lao PDR		11854	11854	1	1	1	1
2	MAL - Malaria cas...	Janua...	01 Vientiane Cap...		102	102	1	1	1	
3	MAL - Malaria cas...	Janua...	02 Phongsali		164	164	1	1	1	
4	MAL - Malaria cas...	Janua...	03 Louangnamtha		122	122	1	1	1	
5	MAL - Malaria cas...	Janua...	04 Oudomxai		169	169	1	1	1	
6	MAL - Malaria cas...	Janua...	05 Bokeo		80	80	1	1	1	

The date is in character format, but we need it in date format for R to know how it should be ordered

```
laos_dat_clean <- laos_dat %>%  
  filter(organisation_unit != "Lao PDR") %>% # filter out national level data  
  dplyr::select(-c(numerator, denominator, factor, multiplier, divisor)) %>% # remove unused columns  
  mutate(data = str_replace_all(data, "MAL - ", ""), # remove starting MAL key  
         data = str_to_lower(data), # lower case  
         data = str_replace_all(data, " ", "_"), # spaces to underscores  
         data = str_remove_all(data, "[()]" ), # remove parenthesis  
         data = str_remove_all(data, "[+]" )) %>% # remove plus sign in one variable name  
  mutate(month = ymd(parse_date_time(period, orders = "B Y"))) # make month a date variable
```

Cleaning (2)

It's a good idea to look at the unique values in each text column to see if they all make sense

```
unique(laos_dat$data)
[1] "MAL - Malaria cases tested (total)" "MAL - Confirmed malaria cases (micr + RDT)"
[3] "MAL - Confirmed malaria cases per 1000" "MAL - Confirmed malaria cases"
[5] "MAL - Microscopy positive malaria cases" "MAL - Malaria suspects tested (RDT)"
[7] "MAL - Malaria suspects tested with microscopy" "MAL - RDT positive malaria cases"
```

We downloaded data that had similar names because we wanted to check for consistency – lets check these now – I'd expect

"MAL - Malaria cases tested (total)" = "MAL - Malaria suspects tested (RDT)" + "MAL - Malaria suspects tested with microscopy"

And

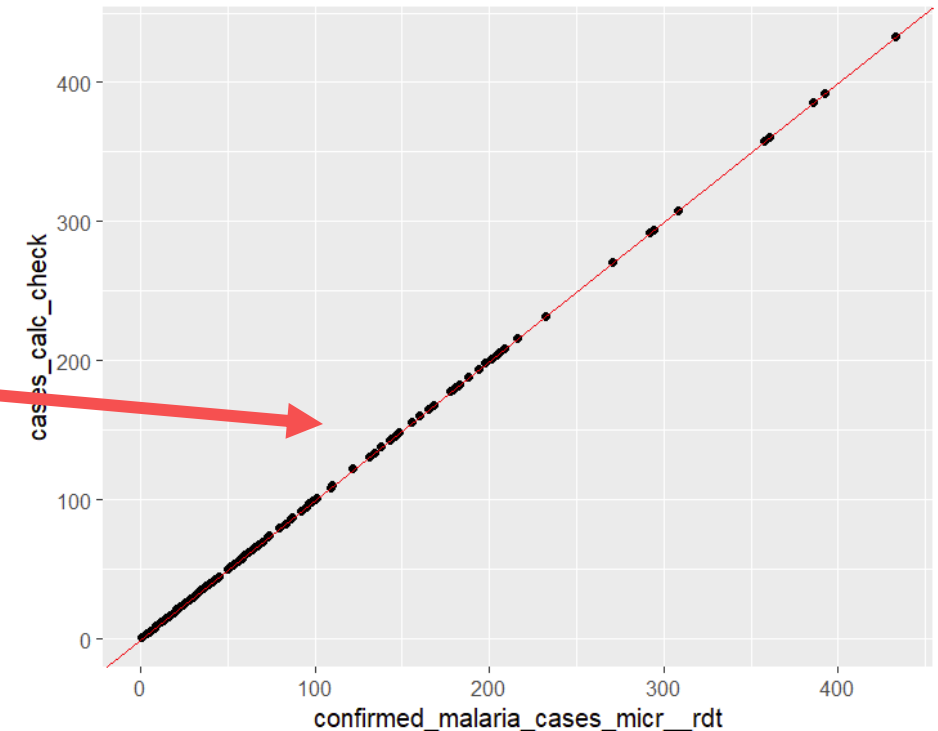
"MAL - Confirmed malaria cases (micr + RDT)" = "MAL - Microscopy positive malaria cases" + "MAL - RDT positive malaria cases" =

Cleaning (3)

```
laos_dat_wide <- laos_dat_clean %>%  
  pivot_wider(names_from = data) %>%  
  mutate(test_calc_check = rowSums(across(c(malaria_suspects_tested_rdt,  
      malaria_suspects_tested_with_microscopy))), na.rm = TRUE),  
  cases_calc_check = rowSums(across(c(rdt_positive_malaria_cases,  
      microscopy_positive_malaria_cases))))
```

```
ggplot(laos_dat_wide) +  
  geom_point(aes(x = malaria_cases_tested_total,  
      y = test_calc_check)) +  
  geom_abline(intercept = 0, slope = 1, color = "red")
```

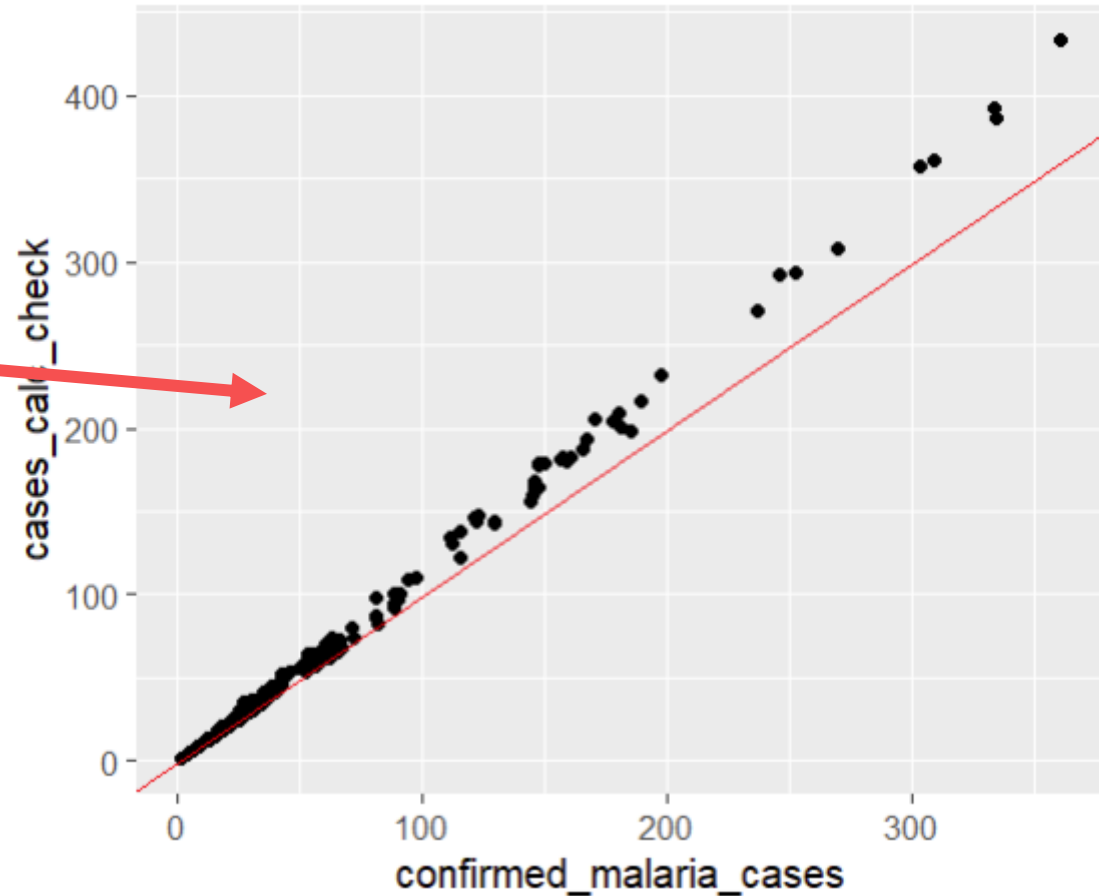
Here we see that these 2 variables are consistent, so we can continue just looking at one



Cleaning (4)

However the same is not true when we compare **cases_calc_check** with **confirmed_malaria** cases (we'd expect the points to fall on the red (x=y) line)

Q: Can you write the code to generate this plot?



This means we, as the data analysts, need to **make a decision** about which variable to use.

Option 1: ask a local **expert**

Option 2: make an **educated decision** and **document** that decision so you can explain your results transparently in the future

For this analysis, we will use **confirmed malaria cases** (as we have spoken to a Laos surveillance officer who reported that in some setting individuals are tested with both RDT and microscopy which can lead to double counting of cases)

Cleaning (5)

Filter only the variable we want for our dashboard

Rename them so they're shorter and easier to work with

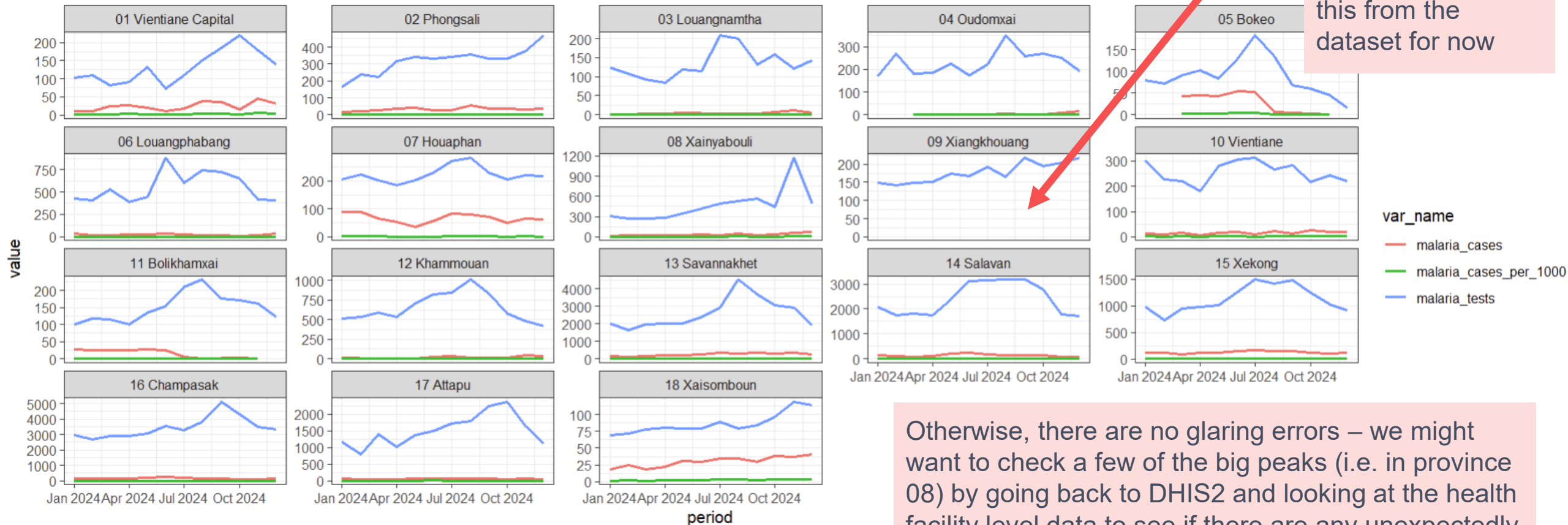
```
laos_dat_clean2 <- laos_dat_clean %>%  
  filter(data %in% c("malaria_cases_tested_total",  
                    "confirmed_malaria_cases",  
                    "confirmed_malaria_cases_per_1000")) %>%  
  
  mutate(var_name = case_when(data == "malaria_cases_tested_total" ~ "malaria_tests",  
                              data == "confirmed_malaria_cases" ~ "malaria_cases",  
                              data == "confirmed_malaria_cases_per_1000" ~ "malaria_cases_per_1000")) %>%  
  
  dplyr::select(organisation_unit, period, var_name, value) %>%  
  arrange(organisation_unit, period)
```

Select only the column we want to use and arrange in a more logical order

ALWAYS plot your data when cleaning – it's the easiest way to help spot anomalies/errors


```
ggplot(laos_dat_clean2) +  
  geom_line(aes(x = period, y = value, color = var_name), linewidth = 0.9)+  
  facet_wrap(vars(organisation_unit), scale = "free_y") +  
  theme_bw()
```

Here we see that province 09 does not have any case or incidence data – lets just remove this from the dataset for now




Otherwise, there are no glaring errors – we might want to check a few of the big peaks (i.e. in province 08) by going back to DHIS2 and looking at the health facility level data to see if there are any unexpectedly large values

Filter out the
province with
missing variables



```
laos_dat_final <- laos_dat_clean2 %>% filter(organisation_unit != "09 Xiangkhouang")
```

```
write_csv(laos_dat_final, "data-outputs/data_dhis2_laos_ammnet_cleaned.csv")
```



Save your output in
the **data-outputs**
folder, making note
that this is now a
cleaned dataset

If you haven't made it through all the data downloading and cleaning and please download **this dataset** so you will be ready for part 2: building a shiny dashboard