Exercise 1. How does house design relate to malaria incidence in children in Uganda? (40 min)

In this exercise we will use the PRISM cohort study to determine if there is an association between house design and risk of *Plasmodium* infection. The PRISM study collected variables such as wall, floor, and roof type, the type of eaves used, and the presence of airbricks. You will learn how to access the data and explore it using the observation search filters. You will also discover how to use the Shiny Apps to visualize the data while trying to answer the question of how house design relates to malaria in children in Uganda.

1. Navigate to http://clinepidb.org in your browser

Start an Observations search

Take a look at the PRISM Cohort Study card. Notice the 4 icons located under the words 'Search the Data'. Each icon starts a different type of search, Participant, Household, Observation, or Light trap. Within each search, you can filter the data based on any of the variables that were collected in the study, the only difference is in the data tables you get in the end. A Participant search will return one row of data per study participant, an Observation search will return one row of data per observation, etc.

We want to determine how house design relates to malaria incidence in the PRISM cohort study. This study collected data from all participants every three months and from a subset of children every month. Participants were also encouraged to come to the clinic any time they were sick, so each participant has multiple observations. *What type of search do you think we want to do based on this information? Why?[1]*

Start an Observations search by clicking the 'Observation' icon on either the study card or the drop-down menu under "Search a Study". This will take you to a page with the Search Wizard at the top.
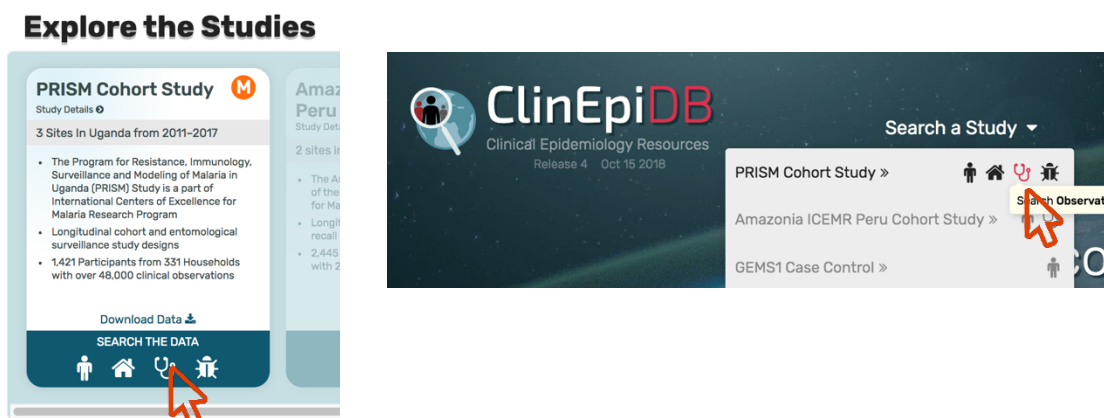
The purpose of the Search Wizard is two-fold. First, it creates a simple way to categorize components of the data allowing for a step-wise approach to building searches. Second, it allows you to explore the data to see what the raw number and distribution of characteristics are in both the full dataset or filtered data.
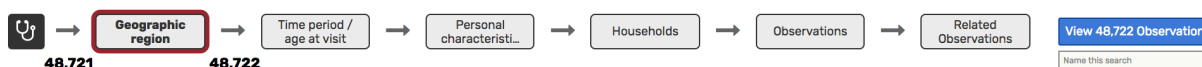
Notice that there is a number below the black square Observation icon. This number represents the total number of Observations that are included in this dataset.

Finding Filters and Examining Data in the Filters – Determining what house design and malaria-related variables were collected

We are interested in looking at how house design relates to malaria in children. *Where might you expect to find age-related variables in the Search Wizard? How about house-related variables? And malaria variables?[2]*

---

[1] An Observation search will return one row of data per observation and allow us to capture data from every individual malaria episode.

[2] Time period/age at visit and Participant have variables related to age, Households has variables related to house design, and Observations has variables related to malaria episodes

1. Click on "Time period/age at visit" in the Search Wizard. We know that data can be either categorical or numerical. For numerical data in our database that has greater than 10 values, the data is displayed as a histogram rather than in table format. Notice that the histogram on the right has an x-axis range of zero to above 75. Can you find the maximum Age at visit (years)? How about the average?
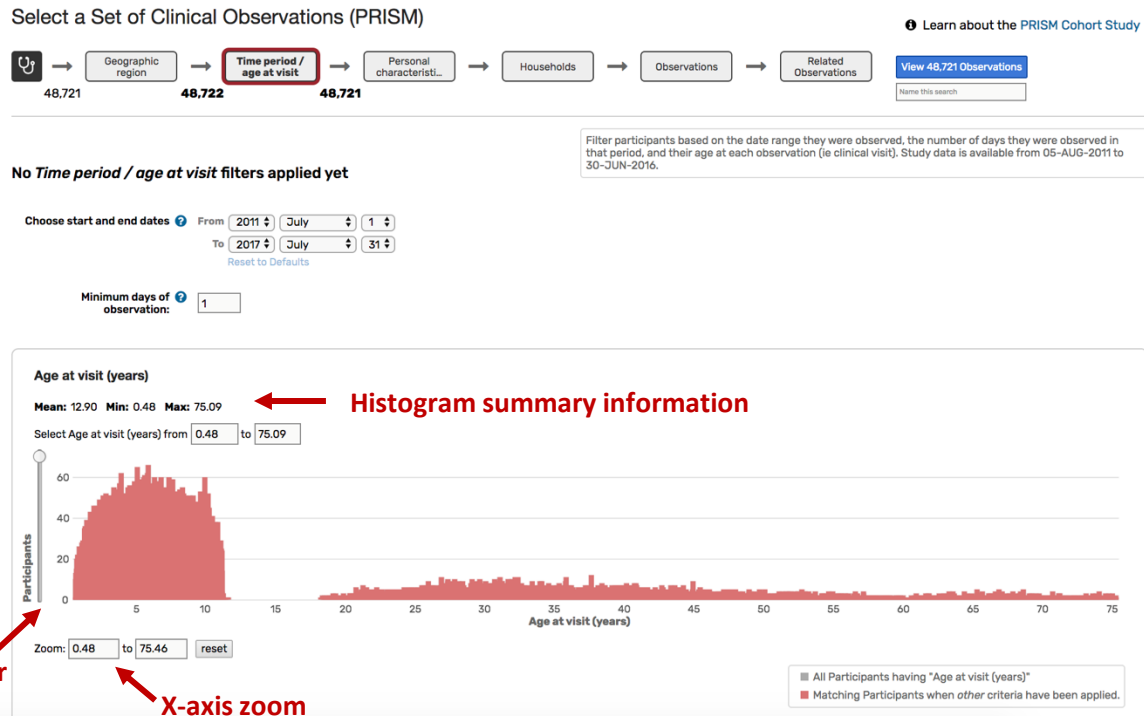
This can easily be determined by the statistics displayed above the histogram. The average age at visit was 12.90 and the maximum age was 75.09. The data for this variable has a very interesting distribution pattern. *Where can you find more information on how a study was conducted?*[3]

2. Select the children in the filter by either typing the age range you want into the boxes or by clicking and dragging on the graph to select the range of interest with your mouse.

---
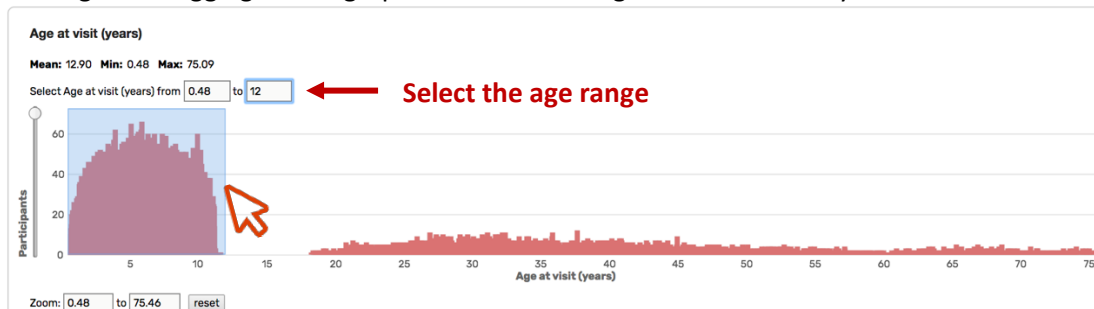
[3] The Study page includes details about the methodology and links to publications. This can be accessed via the link "Learn about the PRISM Cohort Study" in the Search Wizard, by clicking on the body of a study card on the homepage, or by clicking on the study name under the "Search a Study" drop-down menu. In this case, the study enrolled children from age 6 months to 10 years and up to one primary caregiver per household.

3. Notice that the numbers in the Search wizard have updated based on the filter we applied in "Time period/age at visit". *How many observations for children occurred in this study?*[4]

4. Next, take a look at filters (variables) that relate to house design by clicking on "Households" in the search wizard. Examine the available filter categories in blue text on the left-hand side of the page. Clicking a category will reveal the filters within. Clicking the 'expand all' link at the top of the list will reveal all the filters in each of the subcategories. Scroll down and read through these filters. *How many households have open eaves?*[5] *How many households have mud walls?*[6]

5. Now let's look at malaria-related filters. Click on "Observations" in the Search Wizard and look through the available filters.
6. Can you find the information about the presence of asexual parasites? (*Hint: to quickly find subcategories, start typing the subcategory name in the "Find a filter" search box. This will expose the subcategories of interest. Try typing "Asexual" and see what comes up*)

---

[4] 38,145 observations of children
[5] 116 households have open eaves
[6] 202 households have mud walls

Note that the columns "Remaining observations" and "Observations" show different numbers. "Observations" indicates the total data, while "Remaining observations" indicates the filtered data, which in this case is only from children. In the Distribution graphic, grey indicates the data that does not meet the filter criteria, while pink reflects the filtered data. *How many times were children found to be infected with Plasmodium by microscopy*?[7] *How about adults?*[8]

Getting to Know the Results Page – Is a child more likely to be infected or uninfected upon observation?

Now let's look at features of the results page. Notice that the blue button at the far right-hand side of the wizard now says, "View 38,145 Observations". Note: If your button has some other number here, you might have applied additional filters. To see all the filters that you have applied, click on the green filter icon at the top of the Search Wizard.

This will activate a pop-up that lists out all of the active filters you have. You can remove a filter by clicking the 'x', or you can edit it by clicking the blue link.

1. Once you have set your filters to select only children (<12 years old), click the View Observations button in the Search Wizard. This will load the results page. Explore the features of this page by hovering your mouse over some of the links and buttons. By default, the results table displays 20 rows at a time. If you click on 'Advanced Paging' you can display more than 20 rows per page if you wish.

The top of each column in the data table has options for sorting the column in either descending or ascending order and a text description of the column contents. Clicking the 'x' circle will

---

[7] 8,536 child observations were positive for *Plasmodium* by microscopy, or 25% of all child observations.
[8] Since adults have been filtered out by our search you can calculate this by subtracting remaining observations from observations. 9,141-8,536 = 605

remove the column from the table and the small histogram icon will reveal the distribution of data for that column. Use the "Add Columns" button in the upper right to look at other variables.

2. Try clicking on the small histogram icon for "Malaria diagnosis and parasite status". Examine the pop-up box that displays. The y-axis displays the count of Observations and the x-axis displays the selected filter. *Which bin on the x-axis has the largest number of observations?[9] (Hint: Switch from the histogram tab to the Data tab at the top of the graphic to see the full names of each condition along with tabulated values)*
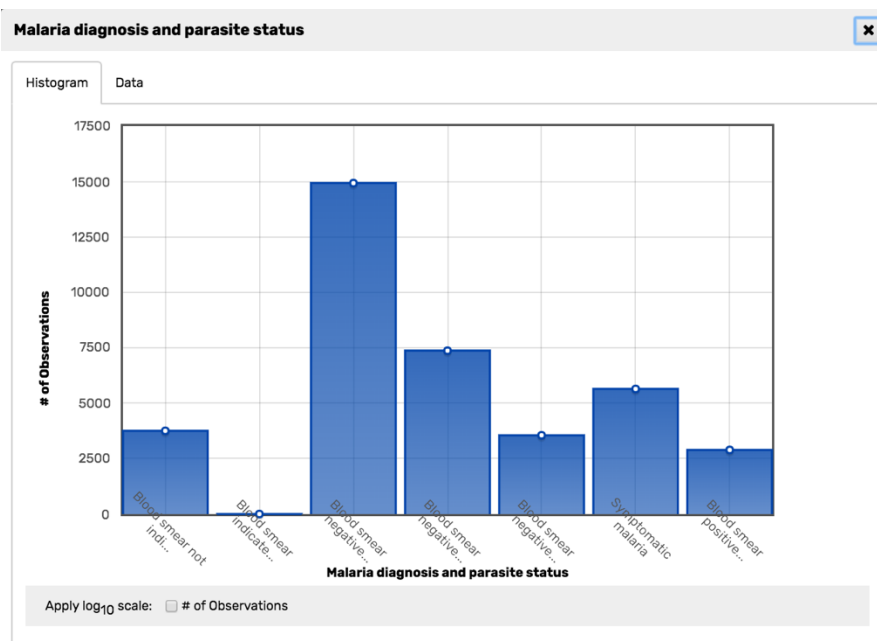
3. Close the histogram and data table and play around with other columns.

Using the Shiny Apps to Visualize Data

We have 3 different Shiny Apps that are available for visualizing data – Data distributions, Contingency tables, and Data summaries. Each app graphs the data in a different way. We want to know how house design is associated with malaria, so let's start by using Data distributions to look at how malaria diagnosis varies based on the house characteristic of "airbricks" and by site.

An important warning about these applications; they are intended for data exploration only! You may choose to examine any number of variables or relationships between variables using these apps but all results should be considered hypothesis generating. More complete, rigorous statistical analyses with proper consideration of study design, possible sources of bias and confounders is strongly recommended before any analyses are considered complete.

---

[9] Blood smear negative/LAMP negative is the most common observation

Data Distributions – How does malaria diagnosis in children vary based on the type of airbricks used?

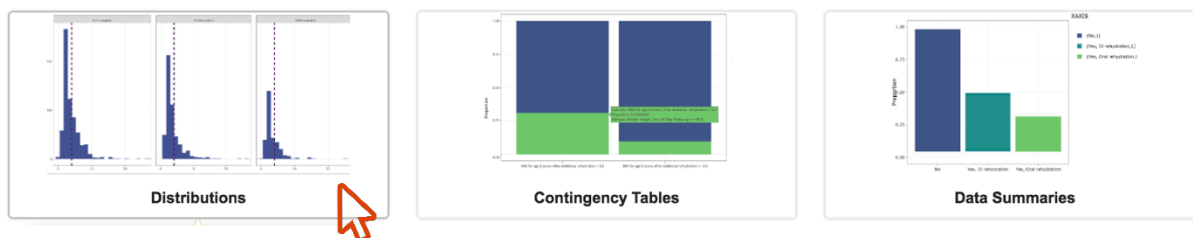1. Click on the "Analyze Results" button from the results page

2. Select the "Distributions" analysis tool

3. When you open a Shiny App, it will have parameters you can edit. By default, the Shiny Apps work on all the data, not just the data you selected in the Search Wizard. However, you can choose to facet, or subset, your data based on your selected results. Here, we are interested in looking at how the Malaria diagnosis and parasite status varies for children based on the presence of air bricks in their home and which site they are from, which will require two facets. Instead of using one facet for choosing only the selected results, we will use the Filter by Age at visit (years) parameter to select for children under 12.
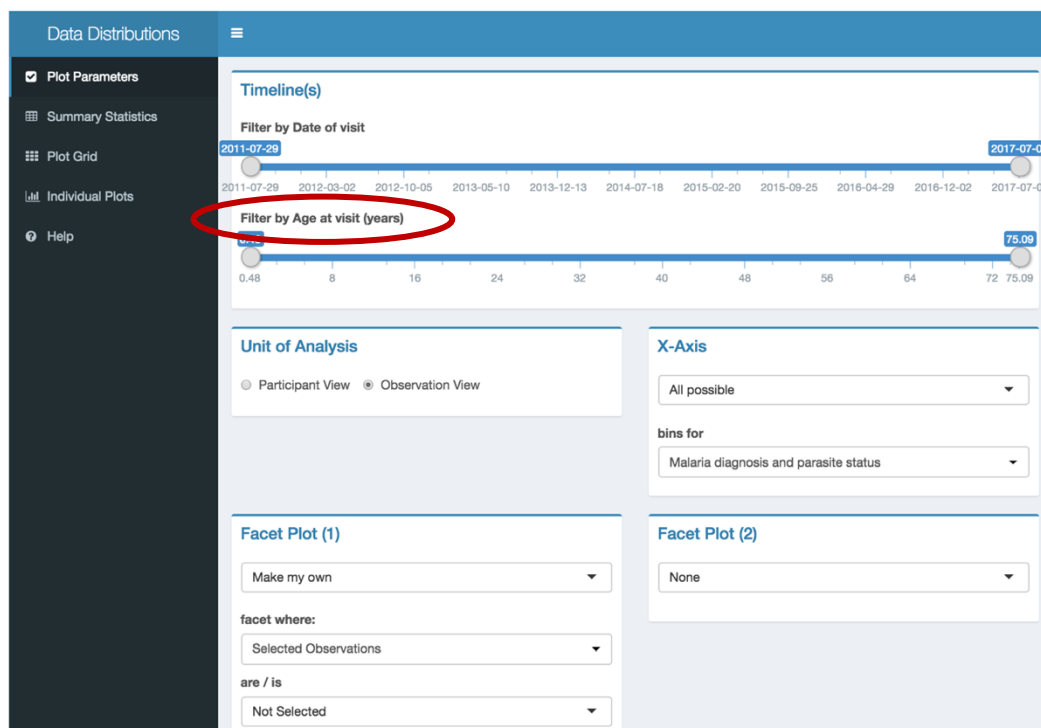
4. Set the "Age at visit (years)" to select for only children
5. In this analysis, we want to see each observation of malaria diagnosis and parasite status tallied, so we will keep the Unit of Analysis on "Observation View" and the X-Axis as "All possible" bins for "Malaria diagnosis and parasite status". *If we want to see how malaria diagnosis and parasite status varies based on air bricks, how will we set our first facet?*[10]
6. Set up the first facet. Notice that as you change "Make my own" to "All possible", the interface updates. When you click the drop-down menu to select the facet you see a filter tree of the possible variables. You can either click through the categories to find your variable or start typing it in the box at the top.
7. The PRISM study chose 3 sites with different transmission intensities. A relationship between house design and geographic region might therefore confound results, so let's set up the second facet for "Sub-county in Uganda".
8. Take a look at the different options on the left-hand menu: Summary Statistics, Plot Grid, and Individual Plots. For Individual Plots, use the drop-down menus at the top to select which distribution plot you want to look at. Hover over the bars to see the specific count. *How many observations of symptomatic malaria are associated with no airbricks in Nagongera? With unscreened airbricks?*[11] *What can you say about the use of airbricks across the different sites?*

---
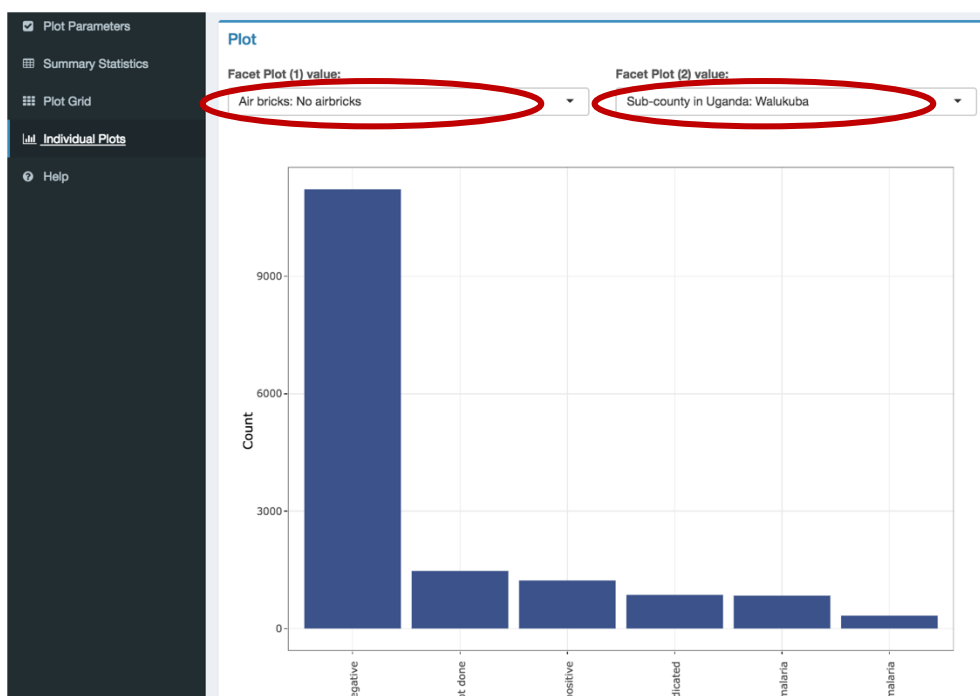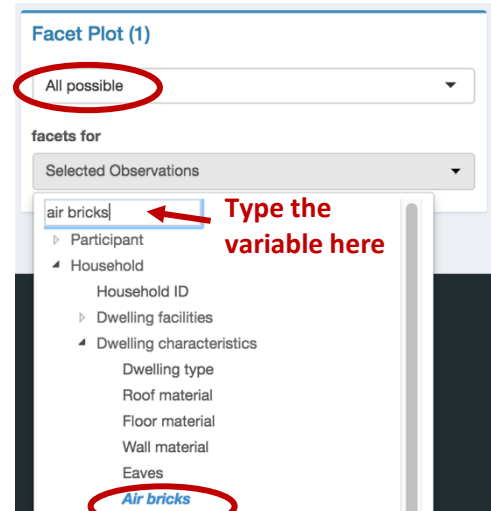
[10] Let's look at "All possible" facets for "Air bricks"
[11] 2,975 observations of symptomatic malaria with no airbricks in Nagongera, 983 observations of symptomatic malaria with airbricks in Nagongera

By looking at the Plot Grid, we can see that the most common type of airbrick varies by site. This Shiny App lets us see how the data breaks down by different variables. Take a moment to try faceting by other variables that you are interested in. Next, we'll use the Contingency Table app to generate 2 x 2 contingency tables and calculate some basic statistics.

Contingency Tables – What's the risk of infection for a child living in a modern vs. traditional dwelling?

The contingency tables app allows you to look at the association between two variables and calculates a p-value, odds ratio (OR), and risk ratio (RR) with 95% confidence intervals.

1. Click on the "Analyze Results" button and select the Contingency Tables app.

Distributions   **Contingency Tables**   **Data Summaries**
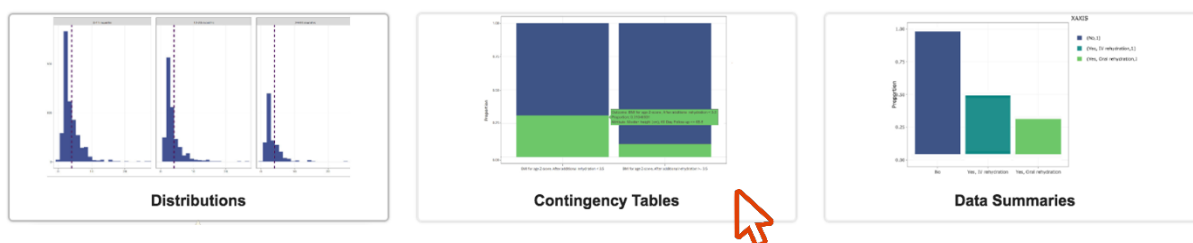
2. We want to look at the relative risk of infection for a child living in modern vs. traditional dwellings and use the variable "Malaria diagnosis and parasite status". However, a 2x2 contingency table only allows two values per variable, and there are more than two potential values for "Malaria diagnosis and parasite status". Can we still use this variable?

Yes, we can still use the variable, but we have to group the values. Here, we are interested in all observations where a child was found to be infected compared to all observations where the child was not found to be infected. We will therefore select "Symptomatic malaria", "Blood smear positive/no malaria", and "Blood smear negative/LAMP positive" from the drop-down menu. The remaining values ("Blood smear indicated but not done", "Blood smear negative/LAMP negative", "Blood smear negative/LAMP not done", and "Blood smear not indication") will automatically become the second group.

3. The PRISM cohort study derived the variable of "Dwelling type" based on the wall and roof material and eaves. Let's select this as our second variable and look at modern homes.
4. Facet Plot (1) is set to facet for "Selected Observations" when they are "Not selected" by default. Let's change this to "Selected". *If we didn't apply this facet, how might we be sure to only use the same data as what we filtered for in the Search Wizard? Why won't this always work?*[12]

---

[12] Since we only filtered for children in the Search Wizard, we could use the "Filter by Age at visit (years)" parameter at the top of the Contingency Tables app. However, if we had applied multiple filters, we would have to facet based on "Selected Observations" in order to restrict our analysis to just the data of interest.

5. Set Facet Plot (2) to facet based on Sub-County in Uganda. Your selections should look like this:

Variable 1

Malaria diagnosis and parasite status ▾

are / is

Blood smear negative / LAMP positive
Blood smear positive / no malaria Symptomatic malaria

Variable 2

Dwelling type ▾

are / is

Modern (wood, cement or brick walls; metal or tiled roof and closed eaves) ▾

Facet Plot (1)

Make my own ▾

facet where:

Selected Observations ▾

are / is

Selected ▾

Facet Plot (2)

All possible ▾

facets for

Sub-county in Uganda ▾

6. Look at the "Summary Statistics". *Is a modern home protective against Plasmodium infection? What's the odds ratio for a child in Walukuba getting infected if they live in a modern home compared to a traditional home?*[13]

Summary Statistics

Facet(s): Selected and Sub-county in Uganda: Kihihi

	p-value	Odds Ratio	Relative Risk
Statistics	<0.0001	0.3873	0.4932
95% Confidence Interval	N/A	0.3605 - 0.4161	0.459 - 0.5299

Facet(s): Selected and Sub-county in Uganda: Nagongera

	p-value	Odds Ratio	Relative Risk
Statistics	<0.0001	0.7894	0.8566
95% Confidence Interval	N/A	0.7519 - 0.8287	0.8159 - 0.8993

Facet(s): Selected and Sub-county in Uganda: Walukuba

	p-value	Odds Ratio	Relative Risk
Statistics	<0.0001	0.4597	0.5064
95% Confidence Interval	N/A	0.4235 - 0.4991	0.4665 - 0.5497

---

[13] The odds ratio is 0.4597, with a 95% confidence interval from 0.4235-0.4991.

7. Check out the plots as well. In this plot of Kihihi. Blue shows observations of infection, while green shows remaining observations. You can see there are significantly fewer infections (blue) for children living in modern homes compared to traditional homes.

8. What other variables are you interested in? Take a look at some of them.
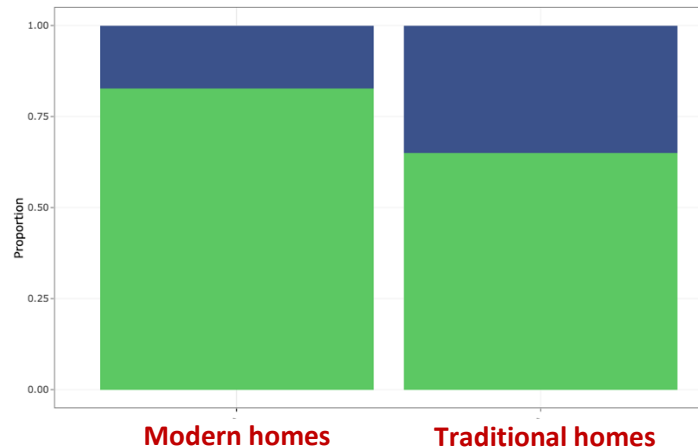
Data Summaries – How does child infection status vary over time based on dwelling type and sub-county?

For longitudinal studies, the Data Summaries app allows you to look at variables over time. Uganda has two annual peaks in malaria transmission which follow the two rainy seasons (March-May and August-October). We might therefore expect to see changes in infection status over time.

1. Click on the "Analyze Results" button and select the Data Summaries app.
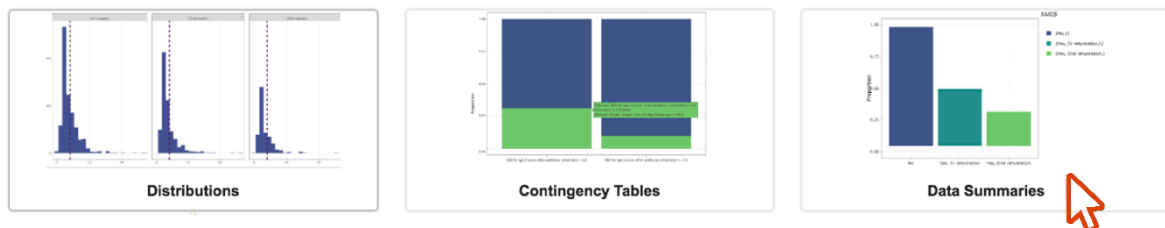
2. This search has app has a few new parameters. The X-Axis parameter allows you to choose between showing observations based on the date they occurred or the age at visit. You can also choose how to bin the data. The Y-axis parameter allows you to set which variable you will be looking at. Facet Line acts similar to Facet Plot, but instead of splitting the data into multiple graphs, it draws a separate line for each possible value.

3. We want to look specifically at children. *What are two ways to limit the data to children?[14]*
4. Set up your parameters so you look at the proportion of observations where the child is infected (hint: use Malaria diagnosis and parasite status) and facet based on Sub-county in Uganda and Dwelling type.

Timeline(s)

Filter by Date of visit

2011-07-29 — 2017-07-07

| 2011-07-29 | 2012-03-02 | 2012-10-05 | 2013-05-10 | 2013-12-13 | 2014-07-18 | 2015-02-20 | 2015-09-25 | 2016-04-29 | 2016-12-02 | 2017-07-07 |

Filter by Age at visit (years)

0.48 — 11 — 75.09

| 0.48 | 8 | 16 | 24 | 32 | 40 | 48 | 56 | 64 | 72 75.09 |

Unit of Analysis

○ Participant View ● Observation View

X-Axis

● My Date Variable ○ My Age Variable

number of bins:

2 — 24 — 40

| 2 | 6 | 10 | 14 | 18 | 22 | 26 | 30 | 34 | 38 40 |

Y-Axis

Malaria diagnosis and parasite status ▼

are / is

Blood smear negative / LAMP positive
Blood smear positive / no malaria Symptomatic malaria

Display as:
○ Count ● Proportion

Facet Line

All possible ▼

facets for

Sub-county in Uganda ▼

Facet Plot (1)

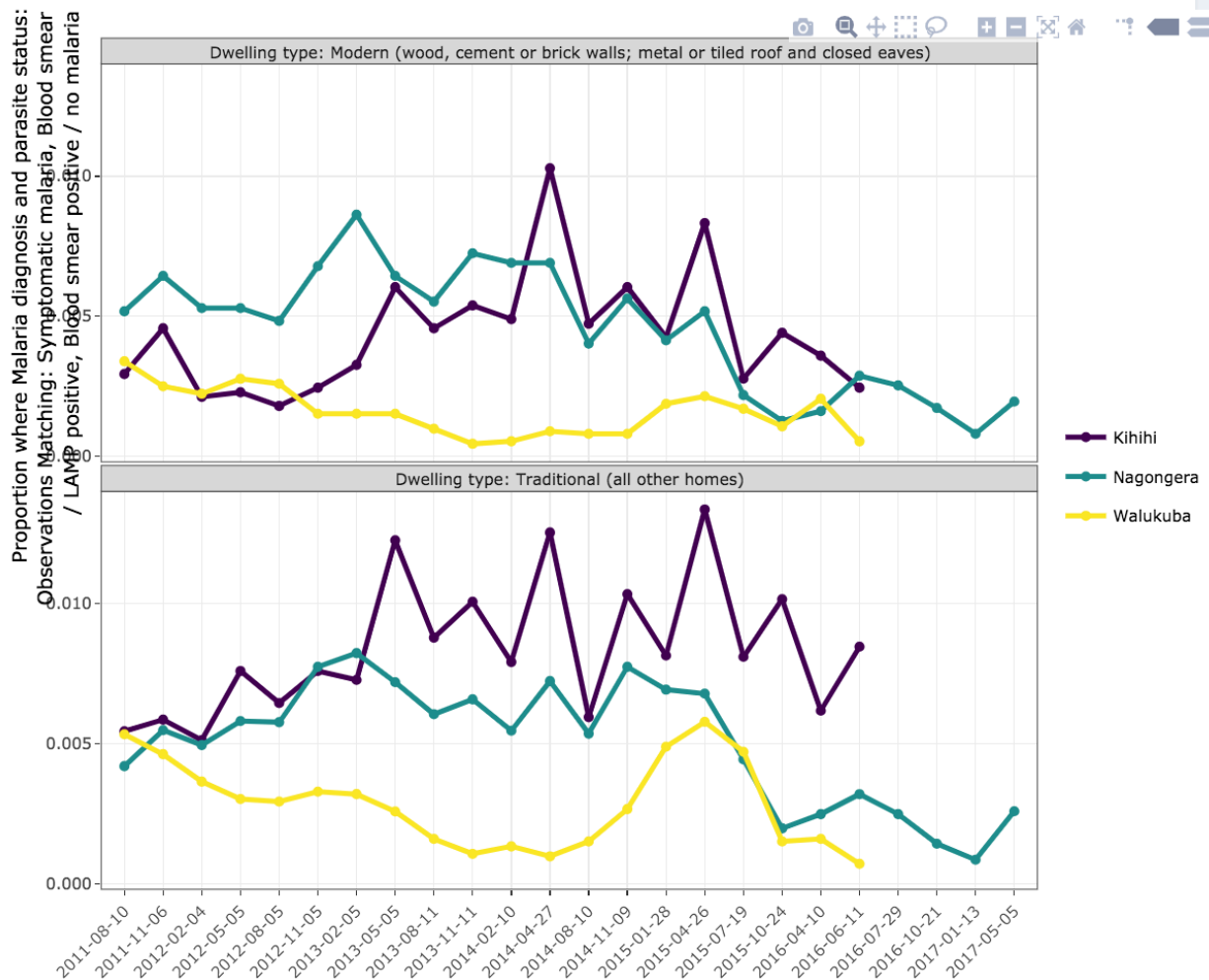All possible ▼

facets for

Dwelling type ▼

Facet Plot (2)

None ▼

---

[14] You can either set the "Filter by Age at visit (years)" or choose to facet on "Selected Observations".

5. Look at the Plot Grid. *How does infection status compare between sites over time? How does it compare between modern and traditional homes? Can you spot signs of an intervention?[15]*

6. Try switching the Facet Line and Facet Plot and playing with other variables.

Do you think house design alters children's risk of *Plasmodium* infection? What other variables would you want to control for? Are there any intervention recommendations you would make based on this data? Why or why not?

You have completed the exercise for this section! Great job! Please let us know if you have any questions or comments via the 'Contact Us' link located in the bottom right corner of any of our web pages. You can also email us at help@clinepidb.org.

---

[15] Walukuba generally has the lowest proportion of infections. Traditional homes are associated with higher proportions of *Plasmodium* infection. Nagongera has a significant decrease in cases in 2015. The Study page indicates IRS was done roughly every 6 months starting in December 2014.