ClinEpiDB exercise: Visualization and Analysis Apps

Objective: The following exercise will demonstrate how to use a Visualization App for epidemiologic data exploration and preliminary analysis.

CAUTION! These apps are intended for data exploration only! You may choose to examine any number or variables or relationships between variables as a hypothesis-generating exercise only. 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.

We will use the **SCORE Burundi Mapping dataset** for this exercise
https://clinepidb.org/ce/app/record/dataset/DS_55fd455f5e

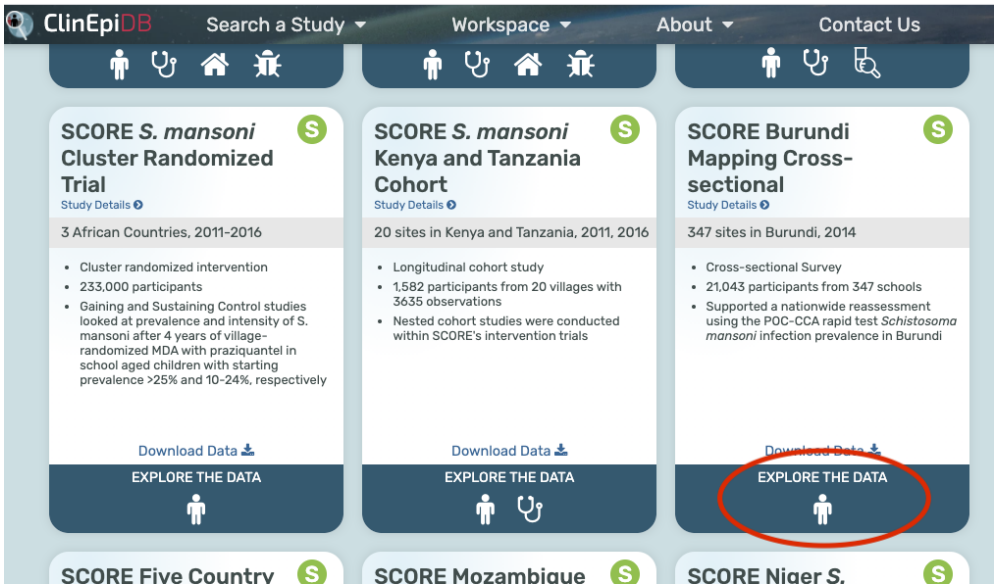
Background: In this dataset, schoolchildren from all over Burundi were sampled for country-wide mapping of the prevalence of schistosomiasis, a parasitic neglected tropical disease. Freshwater snails are intermediate hosts for schistosomiasis, so the study team also collected data on whether there were freshwater bodies near the study sites.

Question:

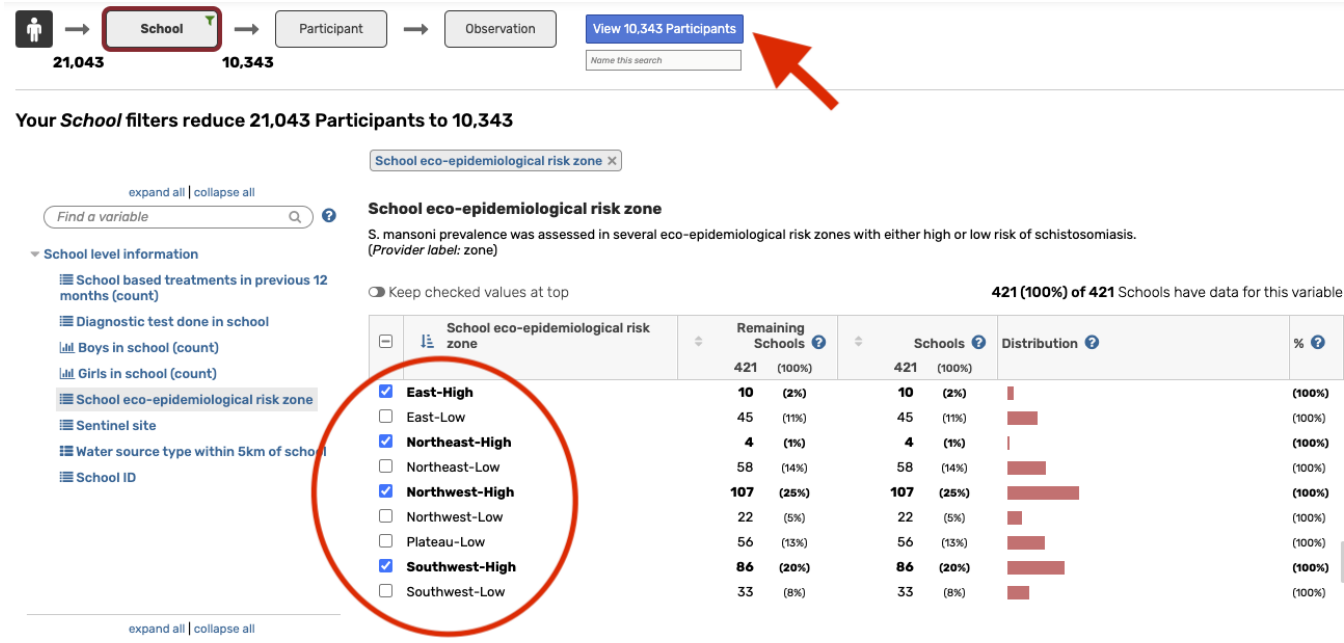
1. Do schoolchildren living near a river have a higher burden of schistosomiasis than those who live further away from a river?
2. Does this relationship vary by whether the sampled school was in a high or low risk zone?
3. Is this relationship different for boys and girls?

Strategy: The burden of schistosomiasis can be measured using the schistosoma egg count variable and visualizing the egg count of the participants as a boxplot, which can be done in the Data Summaries app. We can further stratify the participants by risk zone, river presence/absence and sex to visualize how the schisto burden is related to these parameters.

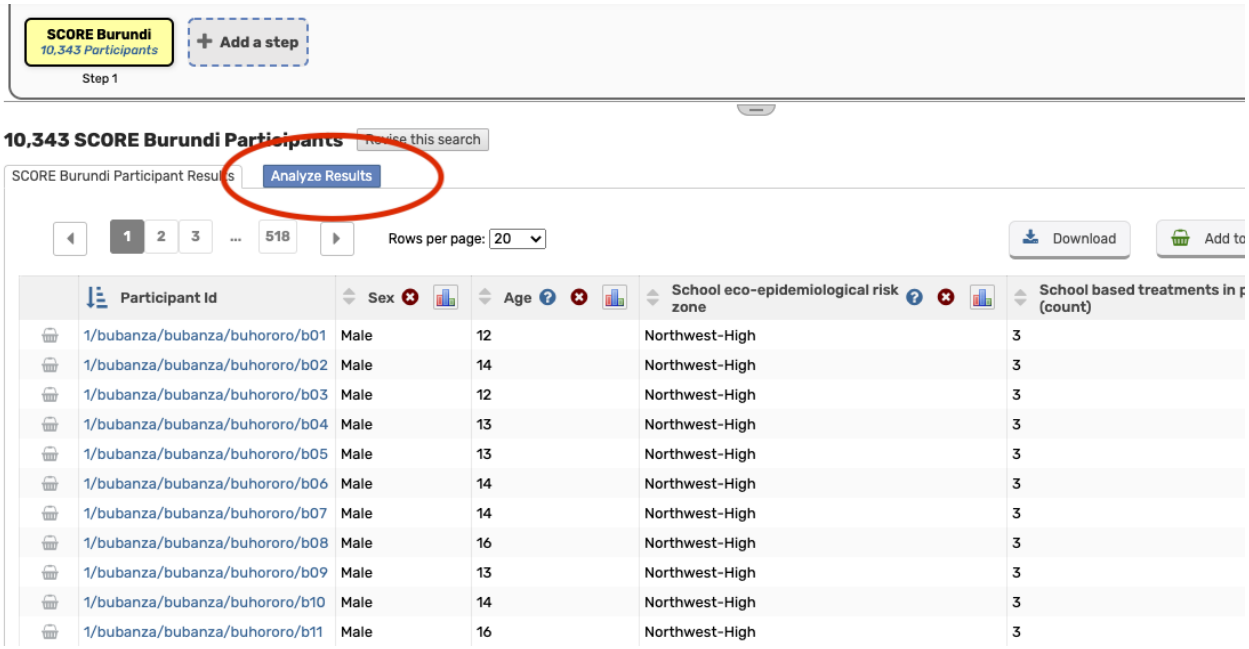
1. Start a Participant-level search of the SCORE Burundi mapping dataset



2. In this study, schistosomiasis prevalence was assessed in several eco-epidemiological risk zones with either high or low risk of schistosomiasis. Let's select all the high risk zones for our analysis (red oval). Then, to get to the Analysis Apps, click the blue "View 10,343 Participants" button (red arrow).



3. To find the Analysis Apps and analyze/visualize the data, click the blue tab next to the Participant Results tab in your data table that says “Analyze Results”



4. There are three Apps available. Click on the box labeled “Data summaries” to initiate that Analysis App.

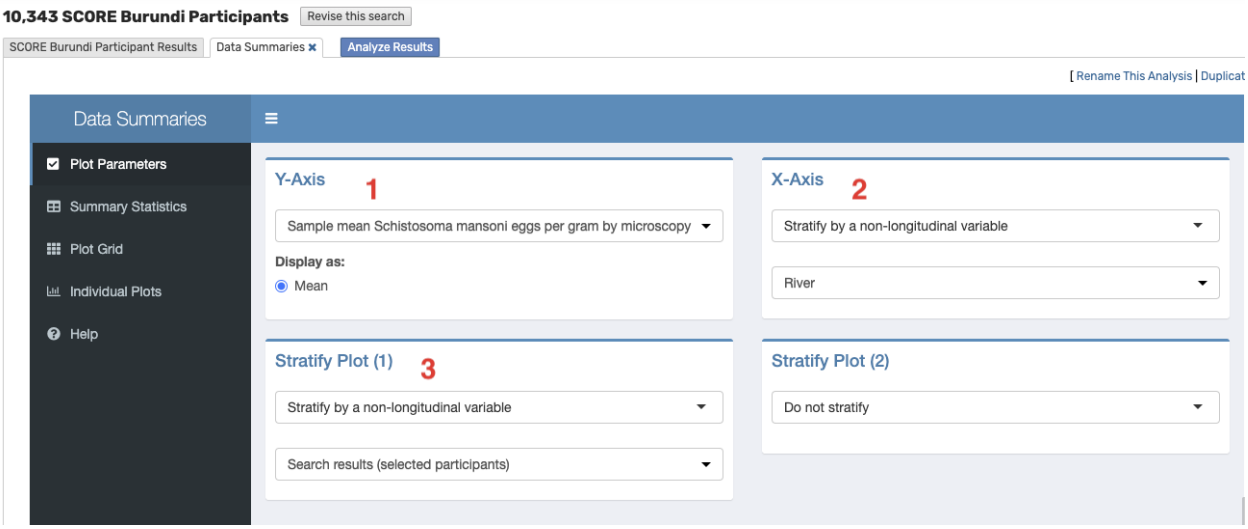
5. When you open an Analysis App, it will have drop down menus that allow you to select which variables you wish to compare.

    1. For the first parameter, the Y-Axis, choose mean Schistosoma mansoni eggs per gram, which represents the schisto burden within the participant.

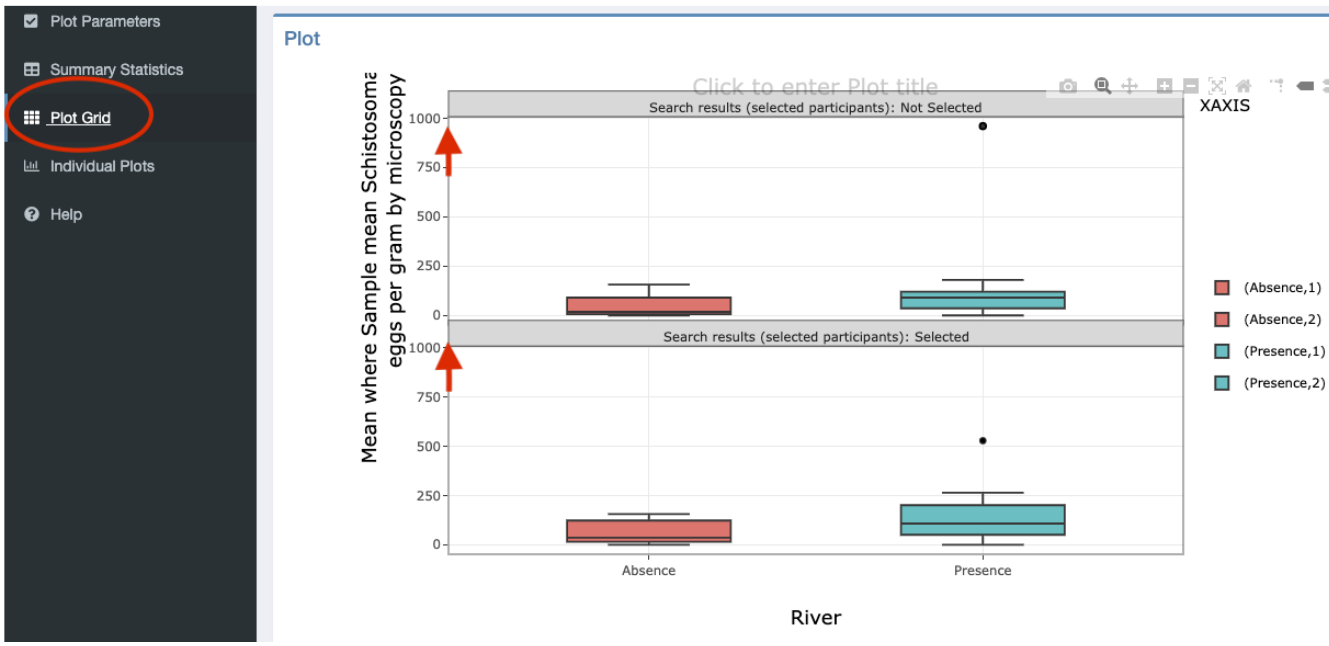
    2. For the second parameter, the X-axis, choose the variable “river” which indicates presence or absence of a river within 5 km of the participant’s school.

    3. By default, the Analysis Apps work on all the data, not just the data you selected in the Search Wizard. However, you can choose to stratify, or subset, your data based on your selected results (in this case the participants from the

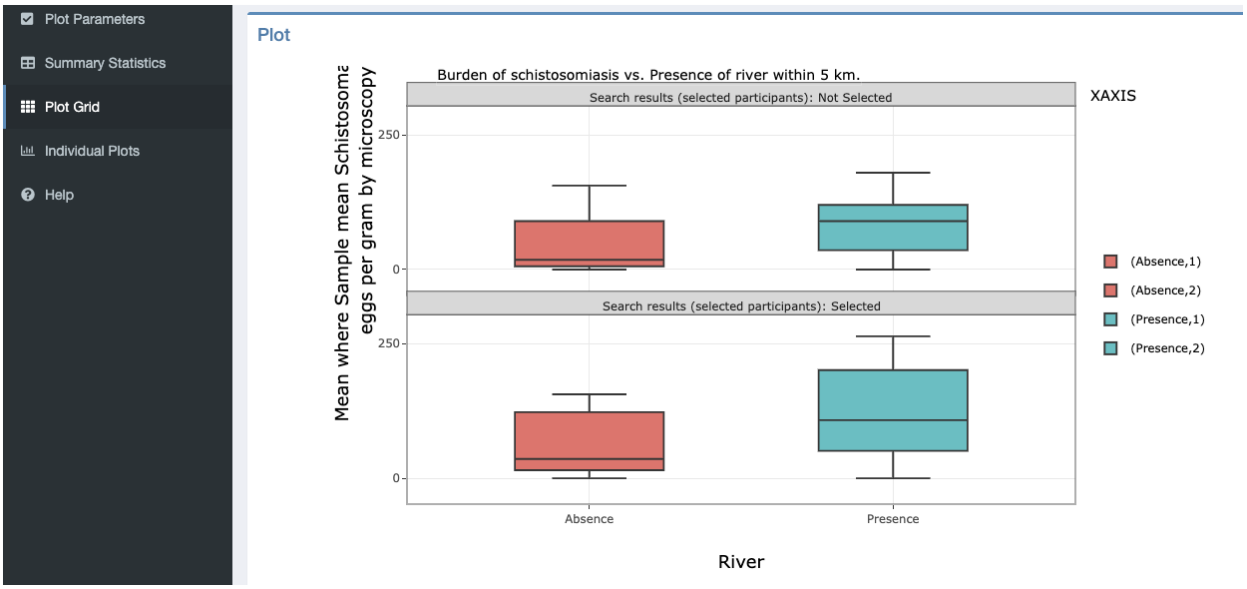
high risk zones). So for the Stratify Plot (1) parameter, let's stratify the participants by the selected results, ie, by the risk zone of the school that we had selected in step 1.



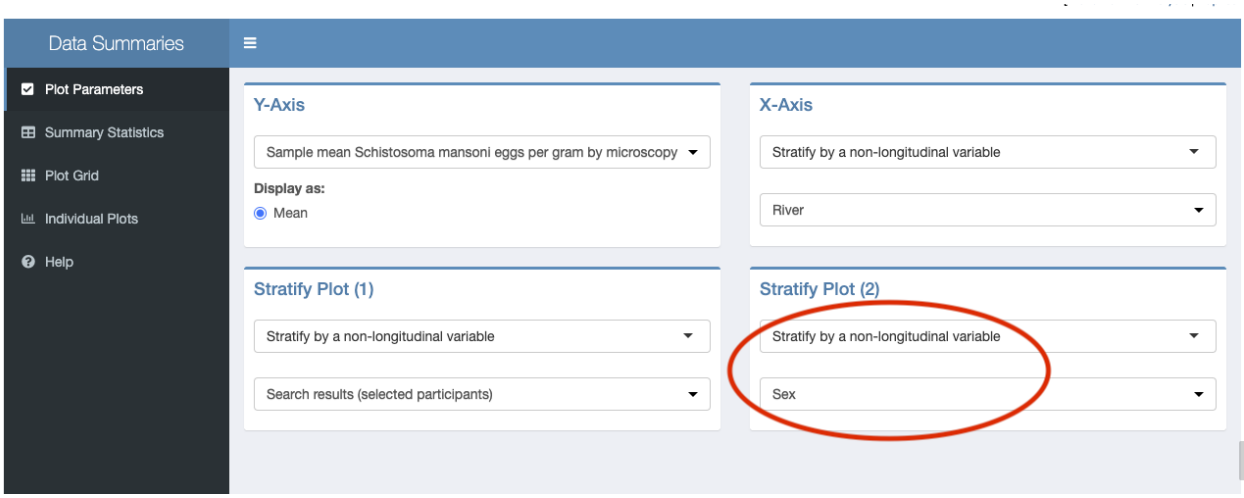
6. Now that we have chosen the parameters, click over on "plot grid" in the menu on the left to look at the plots we made. If you hover near the top of the X axes (shown by red arrows below), a black arrow will appear and will allow you to adjust the axes and get a better look at your data distribution.



7. Examine your plot. Is there a difference in schisto burden between participants who were near a river (presence; blue boxplot) versus those who were not (absence; red boxplot)?[1] Is there a difference in schisto burden among participants in high risk zones ("Selected participants") versus those in low risk zones (not selected)?[2]



8. Now we want to stratify based on the variable Sex, so go back to Plot Parameters, then update Stratify Plot (2).



[1]Yes, schisto burden is consistently higher in schoolchildren within 5 km of a river

[2]Schisto burden is slightly higher in high risk zones but the difference is not dramatic.

9. Once you select this final parameter, go back and look at the plot grid. We are looking at schisto burden (X axis boxplots) by risk zone (low risk zones on the left and high risk on the right) and river presence (red is river absence and blue is river presence) and sex (top row is girls and middle row is boys).
    a. Does presence of river correlate with increased schisto burden?[3]
    b. Is this correlation the same for boys versus girls? Which group shows the most dramatic difference? Can you speculate why this might be? [4]
    c. Is schisto burden always higher in high risk zones compared to low risk zones?[5]

[3]Yes, participants consistently have a higher schisto burden if they are near a river.

[4]Presence of a river seems to have a more dramatic effect on schisto burden in boys compared to girls. This may be related to cultural norms such as boys swimming and playing regularly in the river.

[5]Schisto burden is slightly higher in high risk zones but the difference is not dramatic.