Exploratory data analysis on the ClinEpiDB Beta Platform

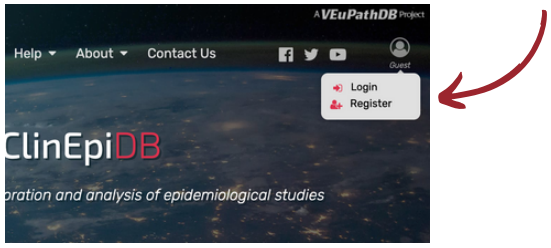
Self-guided

The ClinEpiDB platform is designed for exploratory data analysis of global health studies. In this exercise, we provide some ideas for exploratory questions and list suggested steps for working through an analysis.

Examples of exploratory data analyses on **beta.clinepidb.org**

Dataset	Question	Subset	Visualize
India ICEMR Severe P. vivax and falciparum Cohort	What species of malaria is responsible for most cases of severe malaria in this cohort?	Age at enrollment = 18-40	Use a bar plot to show the distribution of different Plasmodium species
SCORE Five Country CCA Evaluation Cross-sectional	How does the new CCA test compare with the conventional Kato-Katz test for detecting schistosomiasis infection?	Participants with Kato-Katz and CCA results	Use RxC mosaic plot to compare Kato-Katz test with the new CCA test. Which test is more sensitive?
PRISM2 ICEMR Cohort	What age group has the highest prevalence of malaria?	Observation type = enrollment, routine visit, non-routine visit.	Use mosaic plots to look at microscopic and submicroscopic infection, stratified by age group

1. If you would like to save your analysis, make an account by clicking the "Register" link under the "Guest" icon (top right). It only takes a minute to register!



You may also continue as a guest without making an account; the analysis will be accessible as long as the browser window is open.

2. Start at the home page **beta.clinepidb.org**
   - Identify a study of interest by looking through the study cards
   - Click on the title of the study

Exploratory data analysis on the ClinEpiDB Beta Platform

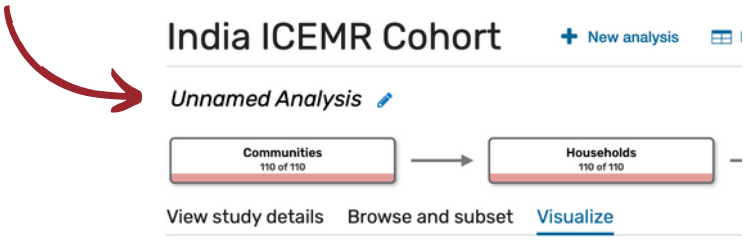
Self-guided (continued...)

3. Click the "**View study details**" tab
   - Read the study summary
   - Scan the objectives and methodology to understand the major objectives, the study design, etc.
   - Use this information to come up with one or more interesting questions to ask of the data, or a hypothesis to explore

4. Click the "**Browse and subset**" tab
   - Identify variables of interest by browsing through the variable tree or searching for variables in the "find a variable" box
   - Star variables that you will use in your analysis
   - Note whether the variables are continuous or categorical; this will be helpful in choosing the visualization tool
   - Use the filtering option to choose a subset of data that is appropriate for your analysis, if applicable

5. Click on the "**Visualize**" tab to plot data
   - Look through the list of visualization tools and identify the appropriate plots to use
   - Make your plots. Do the plots help to answer your exploratory questions?

6. **Name** your analysis



If you check the dropdown menu in the header: Workspace > My analyses, you will see that the analysis automatically appears in this table.

Thank you for completing this exercise on performing an exploratory data analysis on beta.clinepidb.org!