Demonstration on navigating the ClinEpiDB platform

This demo will cover:

1. How to navigate ClinEpiDB.org- analysis workspace and features
    a. Study details
    b. Browse and subset
    c. Visualize
2. How we use ontologies in harmonizing data
3. How to search for variables and datasets

- **Home page**: ClinEpiDB.org
    - Drop down menu of studies
        - Study summaries table
            - See variety of study designs, notice focus on **diarrheal disease**
            - Choose **WASH Benefits Kenya** to explore by clicking on icon
- **Analysis page**- home page of this dataset where we can do several things
    - Dataset diagram
        i. **Shape** of the data- repeated measures for longitudinal data
        ii. **Sample size**
    - View study details tab- wealth of information about the study
        i. **Summary**
        ii. **Useful links**: Publications, related datasets
        iii. Study design, objectives, and methodology
        iv. **Documentation**- codebooks, consent forms
        v. **Study team listing**- they retain ownership of the data
        vi. **72 variables** in this dataset
    - Browse and subset tab
        i. **Variable tree**- organized by categories, shows what kind of data is collected
        ii. Featured vars
        iii. Descriptive stats
            1. Categorical (frequency table and bar chart): ***Household study timepoint***
            2. Continuous- Household > Socioeconomic > ***Persons <=18 years living in house count***- histogram, summary stats, change binning to 2 and see how the plot changes
        iv. How ClinEpiDB uses ontology
            - *Household > Socioeconomic >* ***Persons <=18 years living in house count***. This is the human-readable label that ClinEpiDB has assigned to this variable. The original names of this variable in the dataset are shown here. We also provide the definition here that we pulled from the codebook, making this tab an **interactive codebook**. Epi data and metadata are not standardized, so this is a very helpful and time-saving feature in determining whether this dataset will be useful for one's purpose.
            - What's more, in every dataset where a similar question was asked, we will give it the same harmonized label. This is an **ontology-based**

approach that we take to increase data interoperability. Each term also gets its own identifier (IRI term) you can see in the URL. If anyone is interested in a deeper discussion of ontologies, see one of the facilitators and we would love to tell you more.

v. Subsetting- ***Target child or sibling/neighbor*= Target child**, see changes in dataset diagram and in the ***Diarrhea case*** variable

- Visualize to ask simple questions and look at variable associations
  - Menu of viz tools
  - Make a simple plot, pointing out variable constraints
  - Bar plot (call it diarrhea prevalence)
    - Main: ***Diarrhea case over the last 7 days***
    - Overlay: ***Study timepoint***- Notice that proportion of diarrhea is going down as study progresses from 36% to 28% to 25%
    - Facet by ***Cluster study arm***
      - Passive control shows very little change, nor does water quality
      - Active control (which did not get interventions) shows a decrease too
      - Sanitation arm shows dramatic change, as do combined arms

- In a few minutes, I've demonstrated how to learn more about a dataset, explore variables and make some exploratory graphs. Since I am interested in WASH, I can also search across the site and find other datasets to explore.
- **Site Search**
  - Search for **Water-** I see Studies, Variables and Variable values.
  - Most of what you are looking for will probably be under Variables, so let's look at those first. Click on **Variable** and the **Variable name**
  - Variables: ***Drinking water source*** is a variable in 6 different studies
    - Variable have assorted names in datasets, but we harmonize the label
    - Open it in **LLINEUP** and **PRISM**
  - **Studies**: Matches the **description/summary** in 3 studies
  - **Variable values**- the search term was an option/value/level/category for a variable

With that, we now have an exercise so you can try to navigate ClinEpiDB for yourself, looking at a different diarrhea study called GEMS1.