

ToxoDB- EnvToxo Hands-on Exercise: Upload and explore data in ClinEpiDB

Background

- **ClinEpiDB** (<https://clinepidb.org/ce/app/>) has a portal called “**My Studies**” (<https://clinepidb.org/ce/app/workspace/datasets/new>) that enables users to upload their own dataset privately and explore it using graphical and mapping tools.
- This tool will be rolled out to other VEuPathDB platforms (including ToxoDB) in the next few months.
- Note: You will need a ClinEpiDB/VEuPathDB account in order to upload data. Register here (it is free and only takes a few seconds): <https://clinepidb.org/ce/app/user/registration>

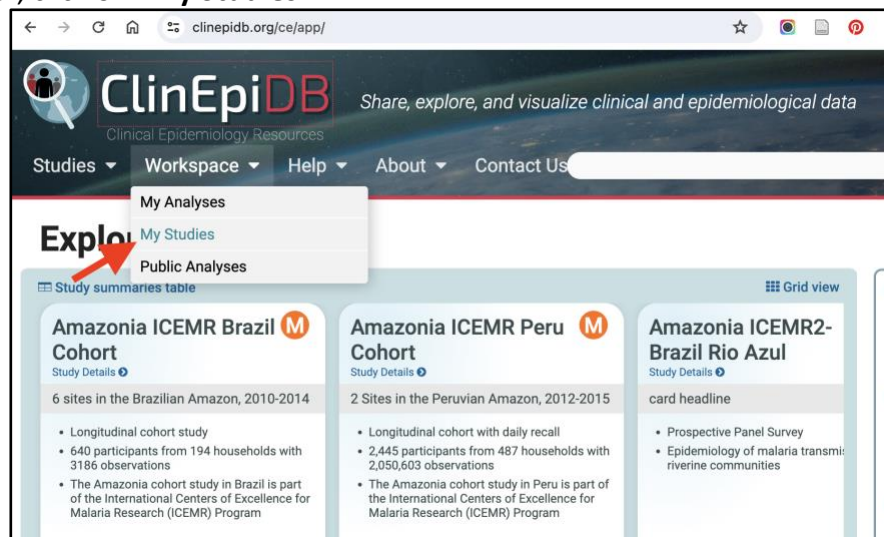
Objectives

- Part 1 (page 1): Upload a practice dataset to the “My Studies” platform
- Part 2 (page 4): Navigate the exploratory data analysis platform to
 - Browse variables and their distribution
 - Use visualization tools to create plots
- Part 3 (page 10): Use interactive mapping tool to explore geospatial data

Part 1. Upload a practice dataset

Note: the dataset provided is for demonstration purposes only and should not be used outside this workshop. Start by saving the practice dataset to your computer.

1. Navigate to the ClinEpiDB home page. Under the “Workspace” drop-down menu in the header, click on “**My Studies**”




2. Upload the practice dataset (see image below)
 - a. Click on the “New Upload” tab
 - b. Add a dataset name
 - c. Add a dataset summary
 - d. Choose the practice dataset file “**toxog**_genotyped_animal.csv” that you previously downloaded
 - e. Click on “Upload Data Set”

My Studies

All **a** New upload Help

Upload My Study

 Before uploading your dataset, please ensure your data is formatted according to the instructions.

Name* **b**

Summary* **c**

Description
optional longer description of the data set including background, study objectives, methodology, etc.

Upload File* **d**

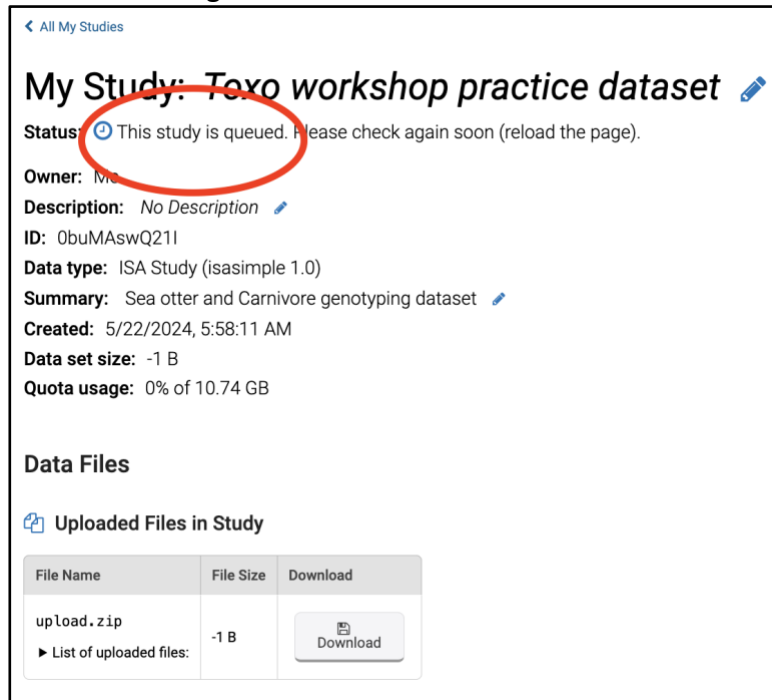
toxo_genoty...d_animal.csv

File must be a .csv, .tsv, or tab-delimited .txt file

e

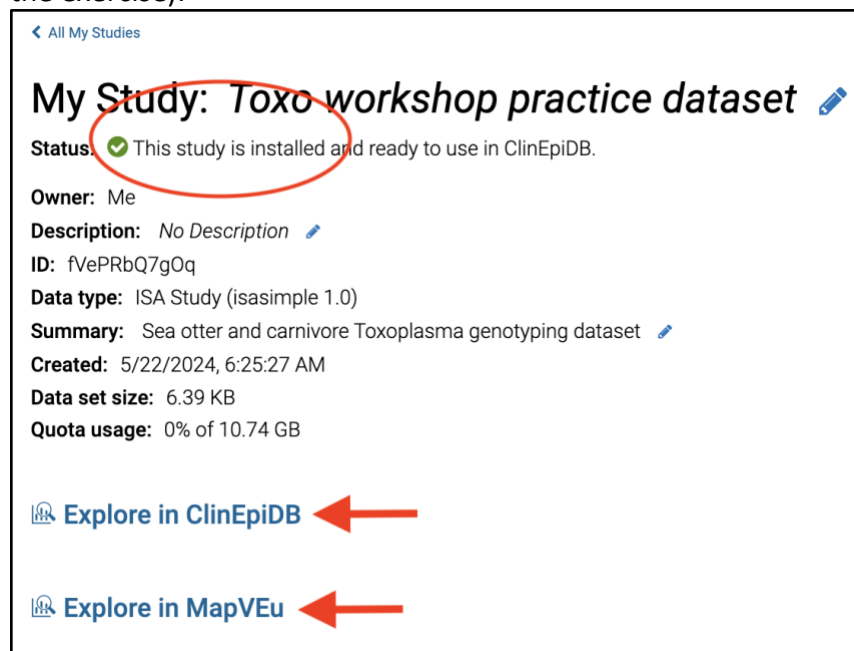
3. The dataset is installed on ClinEpiDB

- a. After you click “Upload” it will open a window that looks like this, indicating that the dataset is being installed.



The screenshot shows the 'My Study' page for 'Toxo workshop practice dataset'. The status is 'queued' with a blue circular icon and a message: 'This study is queued. Please check again soon (reload the page)'. The owner is 'Me'. The description is 'No Description'. The ID is '0buMAswQ21I'. The data type is 'ISA Study (isasimple 1.0)'. The summary is 'Sea otter and Carnivore genotyping dataset'. The created date is '5/22/2024, 5:58:11 AM'. The data set size is '-1 B'. The quota usage is '0% of 10.74 GB'. Under 'Data Files', there is a section 'Uploaded Files in Study' with a table showing 'upload.zip' with a file size of '-1 B' and a 'Download' button. A link 'List of uploaded files:' is also present.

- b. Refresh (reload) the page after 2-5 minutes, and you will now see that the dataset is installed (green check mark). Two links will appear, allowing you to explore the dataset in ClinEpiDB (Part 2 of the exercise) and in MapVEu (Part 3 of the exercise).



The screenshot shows the 'My Study' page for 'Toxo workshop practice dataset'. The status is 'installed' with a green checkmark icon and a message: 'This study is installed and ready to use in ClinEpiDB'. The owner is 'Me'. The description is 'No Description'. The ID is 'fVePRbQ7gOq'. The data type is 'ISA Study (isasimple 1.0)'. The summary is 'Sea otter and carnivore Toxoplasma genotyping dataset'. The created date is '5/22/2024, 6:25:27 AM'. The data set size is '6.39 KB'. The quota usage is '0% of 10.74 GB'. At the bottom, there are two links: 'Explore in ClinEpiDB' and 'Explore in MapVEu', both with red arrows pointing to them.

Part 2. Explore the practice dataset in ClinEpiDB

1. Click on “Explore in ClinEpiDB”. This opens an analysis page like the image below. Examine the layout of the page
 - a. The “entity box” (labeled “a” below) shows that the dataset has 218 observations/rows, in this case, 218 toxo genotyped animals.
 - b. You can give the analysis a name. It will be saved for future use.
 - c. There are 5 tabs that you can toggle between, currently, the analysis is on the “Browse and Subset” tab that allows you to explore each of the variables in the dataset.
 - d. The list of variables is in alphabetical order in the left sidebar- scroll down the variable tree and see what variables the dataset contains.

Toxo workshop practice dataset

Exploring data **b**

a toxo_genotyped_animals
218 of 218

View Study Details **c** Browse and Subset Visualize Download Record Notes

d expand all | collapse all

Find a variable

toxogentyped_animal

- age
- animal
- lat carnivore_FIV
- lat carnivore_FeLV
- lat carnivore_b1_genotype
- lat carnivore_cause_of_death
- lat carnivore_neo_titer
- lat carnivore_predom_habitat
- lat carnivore_sarco_titer
- lat carnivore_toxo160
- lat carnivore_toxo320
- lat carnivore_toxoDB_genotype
- lat carnivore_toxo_titer
- lat carnivore_type
- lat carnivore_year

age

Check items below to apply this filter

218 (100%) of 218 toxogentyped_animals have data for this variable

	Subset of toxogent...	All toxogent...	Distribution	%
<input type="checkbox"/> age	218 (100%)	218 (100%)		
<input type="checkbox"/> Adult	161 (74%)	161 (74%)		(100%)
<input type="checkbox"/> Immature	54 (25%)	54 (25%)		(100%)
<input type="checkbox"/> NA	3 (1%)	3 (1%)		(100%)

2. Click on any variable to examine its distribution.
 - a. Categorical variables are represented as frequency tables, e.g., *animal*. Sea otters represent what proportion of animals in this dataset?

View Study Details Browse and Subset Visualize Download Record Notes

expand all | collapse all

Find a variable

toxogentyped_animal

- age
- animal**
- lat carnivore_FIV
- lat carnivore_FeLV
- lat carnivore_b1_genotype
- lat carnivore_cause_of_death
- lat carnivore_neo_titer
- lat carnivore_predom_habitat
- lat carnivore_sarco_titer
- lat carnivore_toxo160

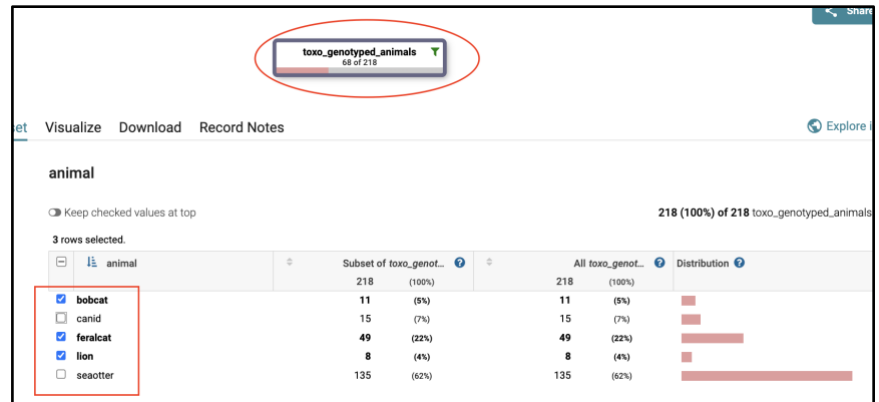
animal

Check items below to apply this filter

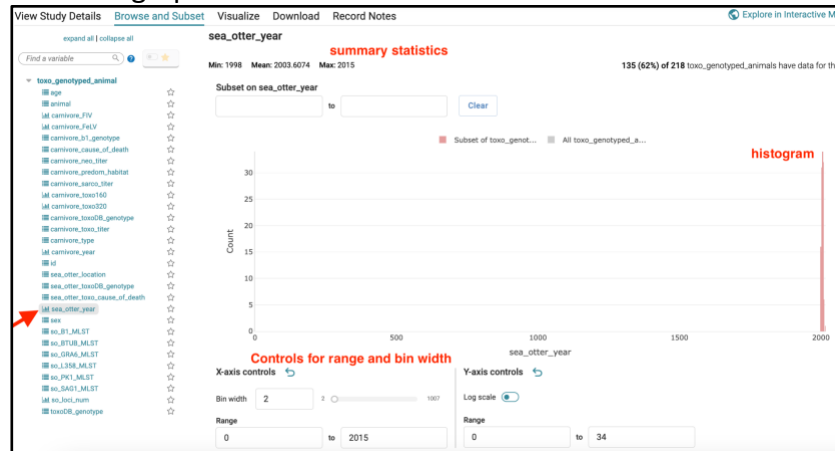
218 (100%) of 218 toxogentyped_animals h

	Subset of toxogent...	All toxogent...	Distribution	%
<input type="checkbox"/> animal	218 (100%)	218 (100%)		
<input type="checkbox"/> bobcat	11 (5%)	11 (5%)		
<input type="checkbox"/> canid	15 (7%)	15 (7%)		
<input type="checkbox"/> feralcat	49 (22%)	49 (22%)		
<input type="checkbox"/> lion	8 (4%)	8 (4%)		
<input type="checkbox"/> seaotter	135 (62%)	135 (62%)		

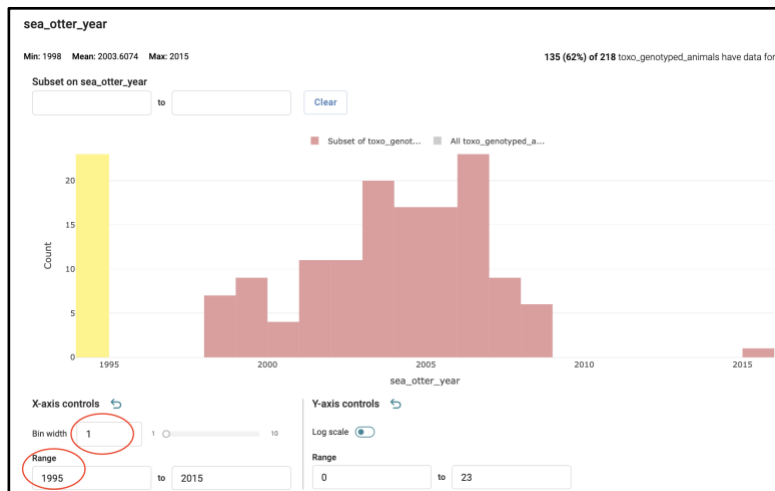
You can subset/filter categorical variables by clicking on the checkboxes, for instance to subset all felids in this dataset (red rectangle below). Observe how filtering updates the entity box (red circle).



- b. Continuous variables are represented as histograms, e.g., *sea_otter_year*. Notice the histogram in pink, the summary statistics at the top, and the controls below the graph.

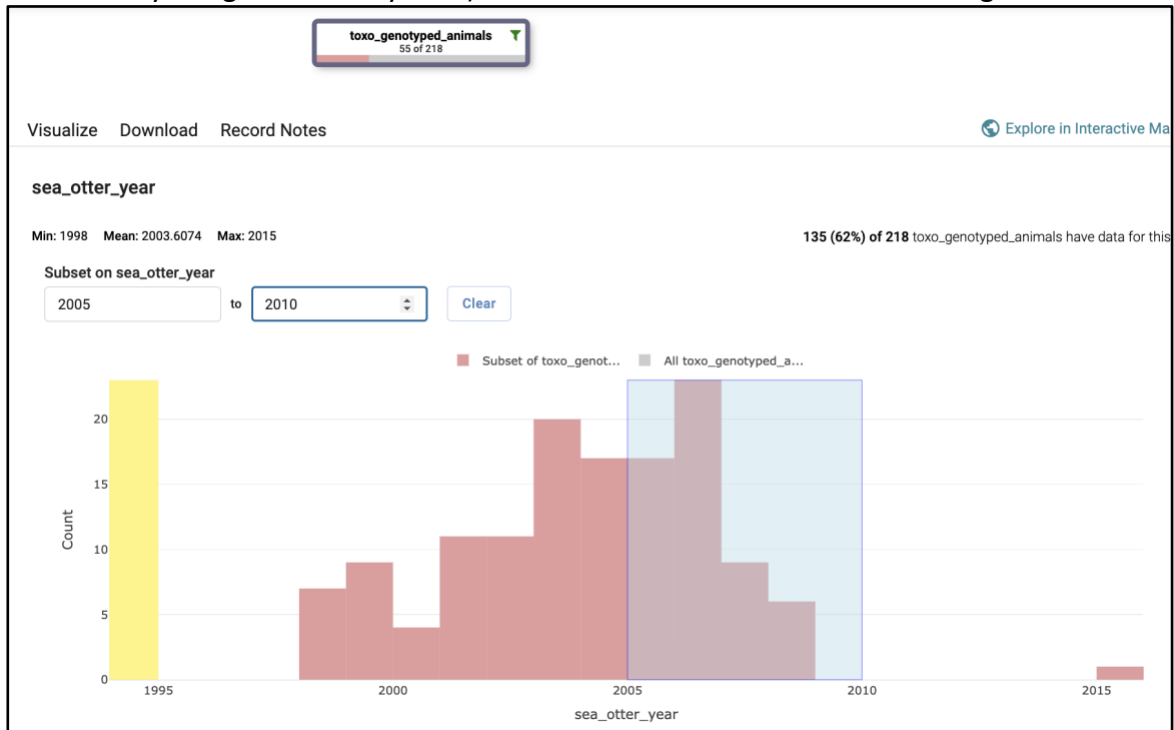


To better visualize this histogram, adjust the **bin width** and **range** under the X-axis controls as shown below.



You can subset/filter continuous variables by dragging to select a portion of the histogram, or by subsetting in the boxes below the summary statistics, for instance to choose observations

between 2005-2010. Observe how filtering updates the entity box above (filtering is represented by the green filter symbol). Click on “Clear” to remove the filtering.

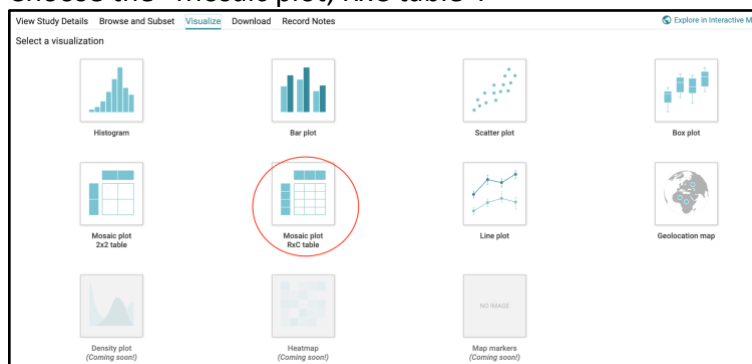


- Click on the “Visualize” tab and then on “New visualization” to access an array of apps that enable you to create graphs and tables to explore associations between two or more variables.

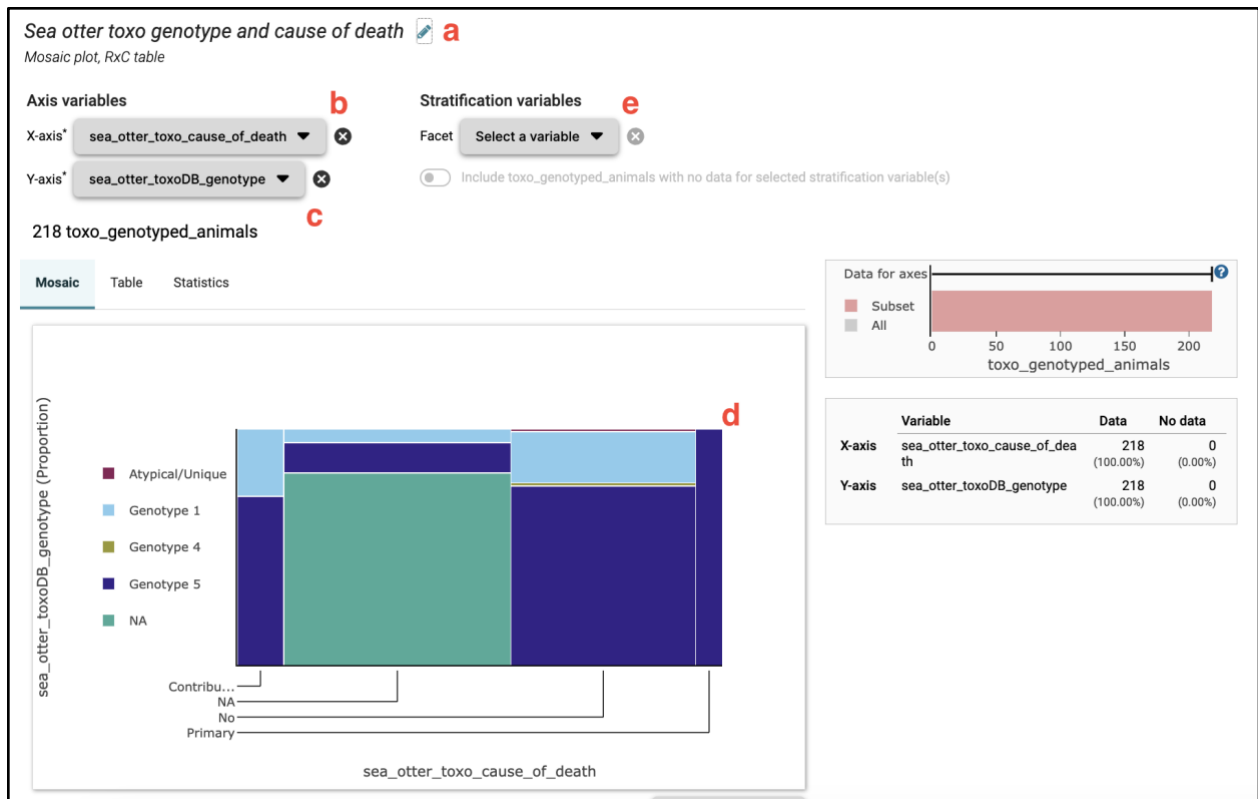
Here are three ideas; please try any or all of them, or make a different graph of your own

Plot 1. What ToxoDB genotypes are implicated in mortality of sea otters?

- Choose the “Mosaic plot, RxC table”.



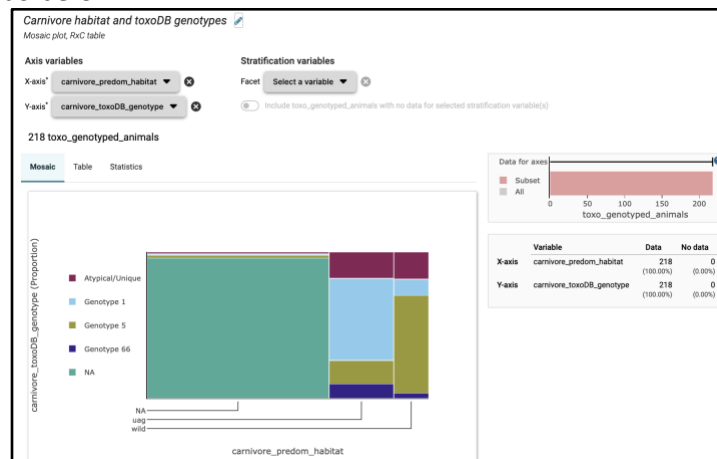
- This opens up an app where you can select parameters



- Give the plot a title
- X-axis= *sea_otter_toxo_cause_of_death* (choose variable from drop-down menu)
- Y-axis= *sea_otter_toxoDB_genotype*
- Observe the resulting plot- when *sea_otter_toxo_cause_of_death*= Primary (X axis), what genotype (Y axis) is it correlated with?
- Optional: Compare the plot to Figure 1 in this publication

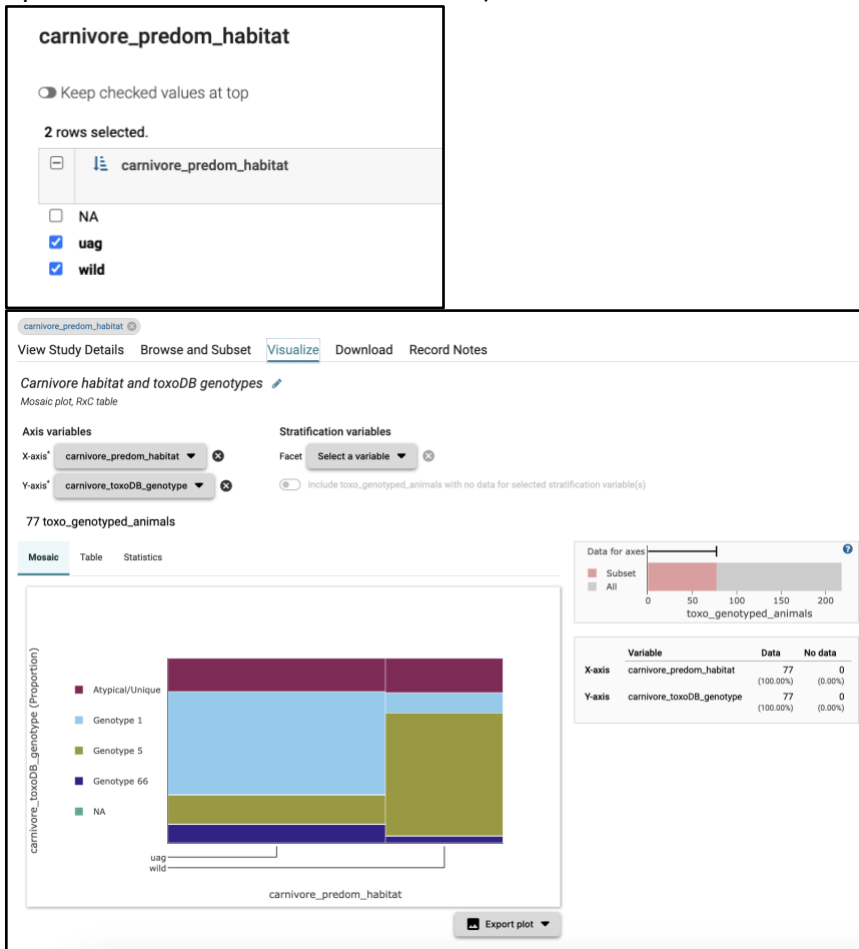
Plot 2. What is the correlation of habitat and ToxoDB genotype in carnivores?

- Click on Visualize > New visualization and select the Mosaic plot, RxC table app again.
- Configure it as below.



- How could you remove the NAs to make the graph easier to visualize? (Hint: go back to the Browse and Subset tab, and for the carnivore_predom_habitat variable, filter out

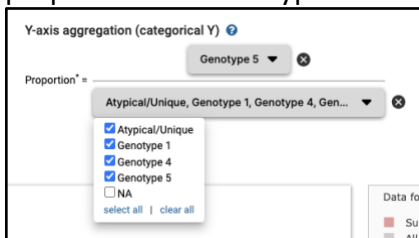
the NAs. Go back to the visualize tab and you will see that the graph automatically updates to reflect the current subset)

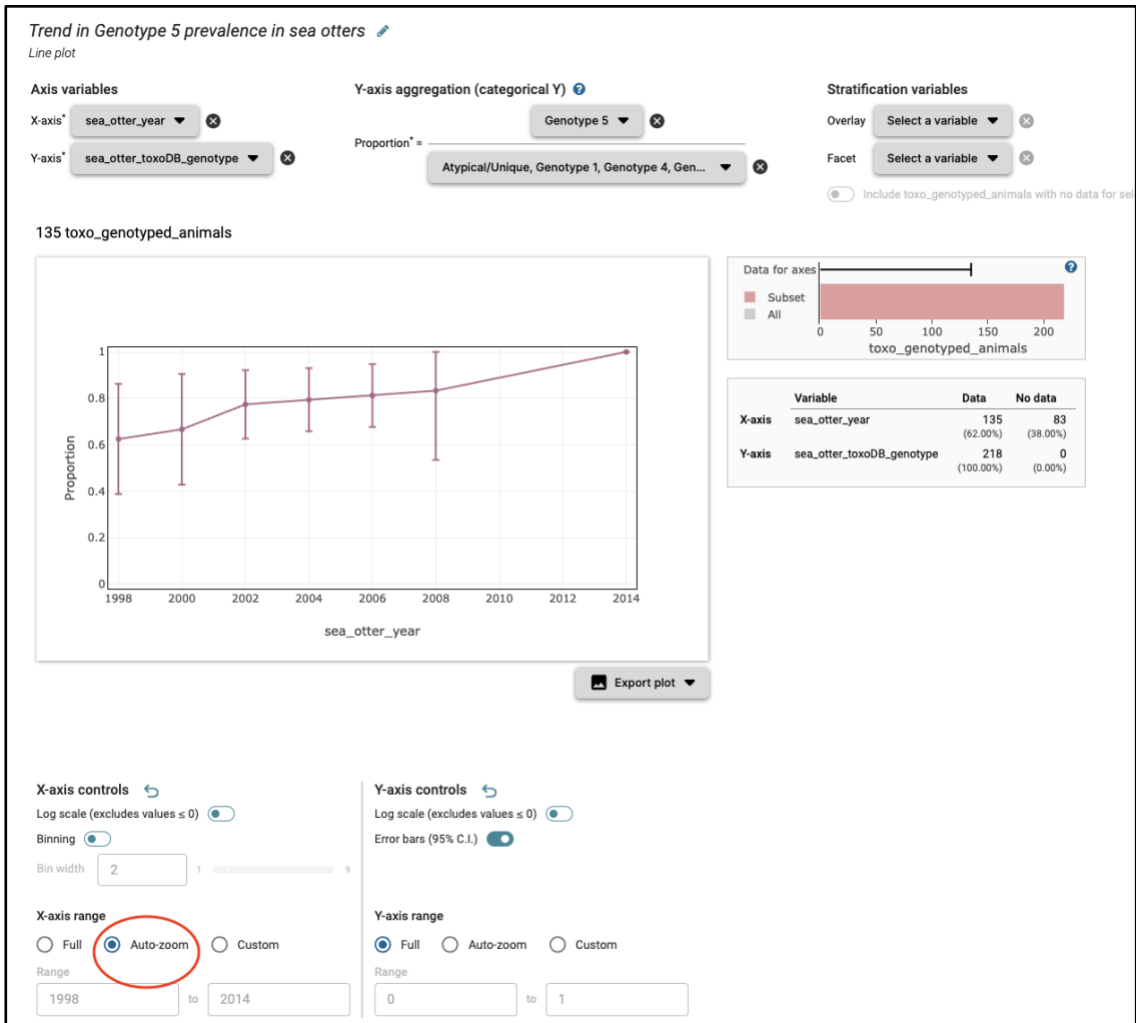


4. In which habitat was Genotype 5 predominantly found?

Plot 3. How did the prevalence of ToxoDB Genotype 5 change in sea otters over the time period of data collection?

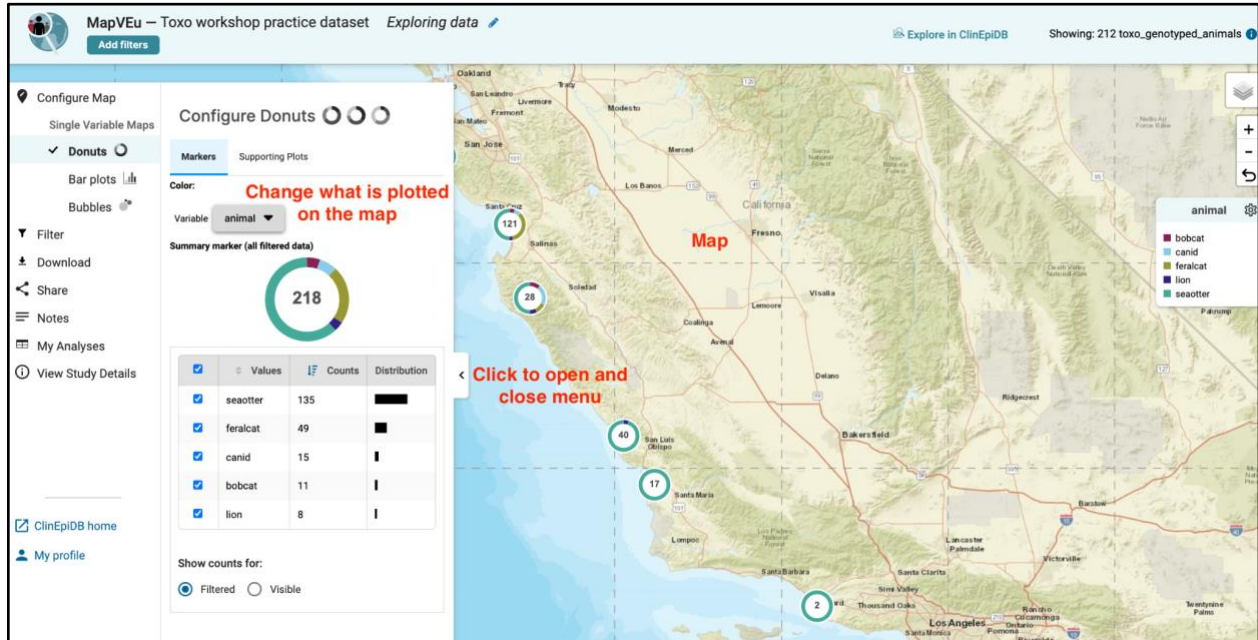
1. Remember to remove the filters from any previous plots before making this new plot.
2. Click on Visualize > New visualization > Line plot
3. Configure the plot as below. Two things to note- auto-zoom on the X-axis, and choose proportions of Genotype 5 as shown.





Part 2. Explore the practice dataset in MapVEu

Towards the top right corner of the analysis page you will notice a link to explore the dataset in the interactive map tool. Click on the link and explore the map that opens.



We can compare ToxoDB genotypes for sea otters and carnivores using the map.

- Copy and paste the URL into a new browser tab to have the map on two separate windows
- In one window, choose sea_otter_toxoDB_genotype for the marker variable
- In the other window, choose carnivore_toxoDB_genotype for the marker variable
- Zoom into each map to explore the data
- More features of the map are explained in this tutorial: https://static-content.veupathdb.org/documents/clinEpi/clinEpi_exercises/Interactive_Map_tutorial_ClinEpiDB_20231127.pdf

–End of exercise–