Visualizing Drinking Water Quality in Mercer County, PA

Author: Sandra Karcher
Date: August 16, 2024

Overall Goal of the Project

The overall goal of the project was to provide a way for residents to review their drinking water quality over the past 10 years.

Specific project requirements:

-Display sampling locations on a map

-Make the application available to the public

-Allow users to select one or more sampling locations and review the water quality data associated with those locations

-Allow users to easily select one or more substances to review

Data Source

The data used is this project were obtained from the Pennsylvania Department of Environmental Protection Drinking Water Reports (Figure 1).

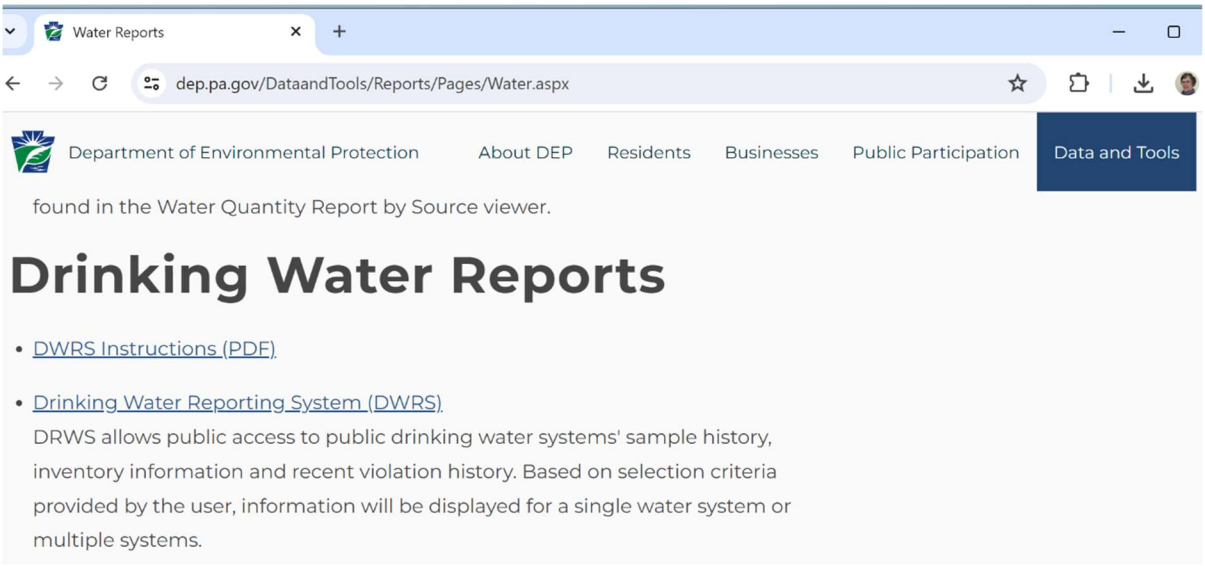

Figure 1: Source of Water Sampling Data

Data from 2014 through 2023 were downloaded and transformed for use (Figure 2).

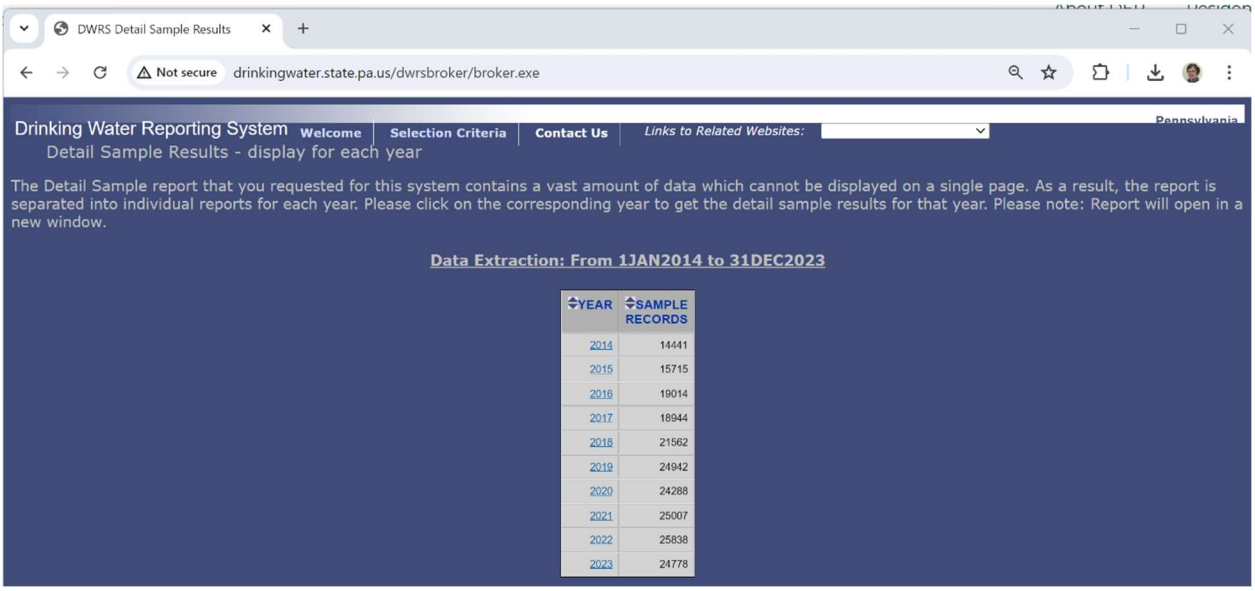

Figure 2: Range of Sampling Times of Data Selected for Visualization

The data were transformed as described in detail in the project GitHub Repository (https://github.com/sandra444/waterdata/blob/main/down/down_water_prep.py).

My Experience Using Tableau Public

The Tableau Public interface is conceptually similar to Excel, but with different categories of tabs: Data Sources, Worksheets, Dashboards, and Stories. For this project, I used one data source, three worksheets, and one dashboard per workbook. Three worksheets were pulled into the dashboard: one that displayed the map, one that displayed the plots of the time series data, and one that provided information about the units. The connection between the map and the plots was established using an “action”.

Data were aggregated in two different ways and saved as two different “Vizzes”. A viz is conceptually like an Excel workbook. In one viz, data from individual locations are plotted on individual graphs (Figure 3), in the other, data from multiple locations are plotted on the same graph (Figure 4).

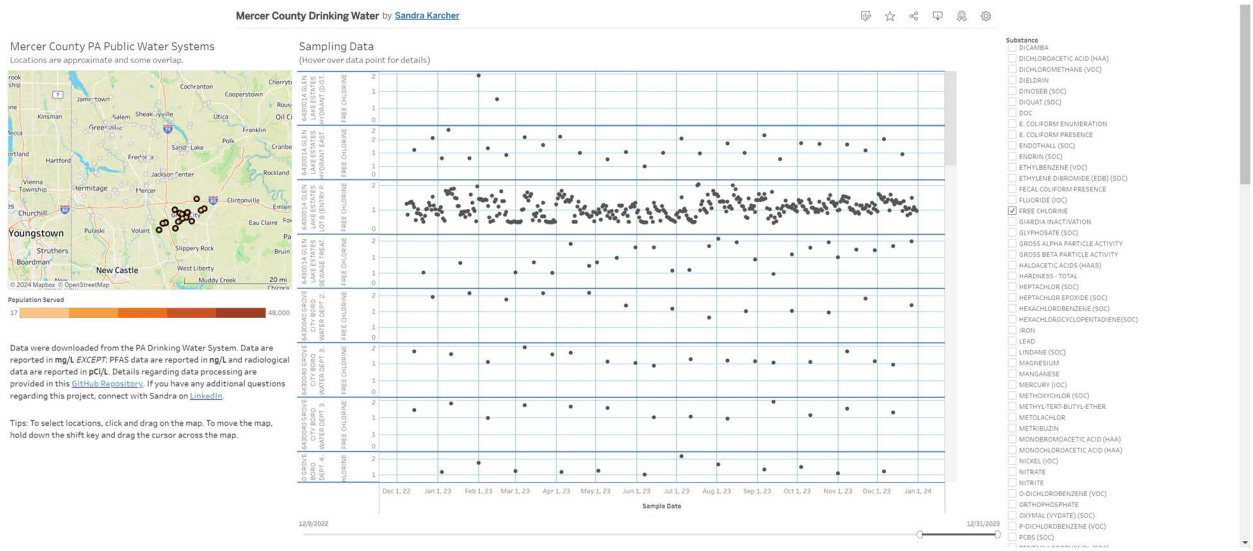

Figure 3: Viz Showing One Graph for Each Location and Substance

(https://public.tableau.com/views/MercerCountyDrinkingWater/Dashboard1?:language=en-US&:sid=&:display_count=n&:origin=viz_share_link)

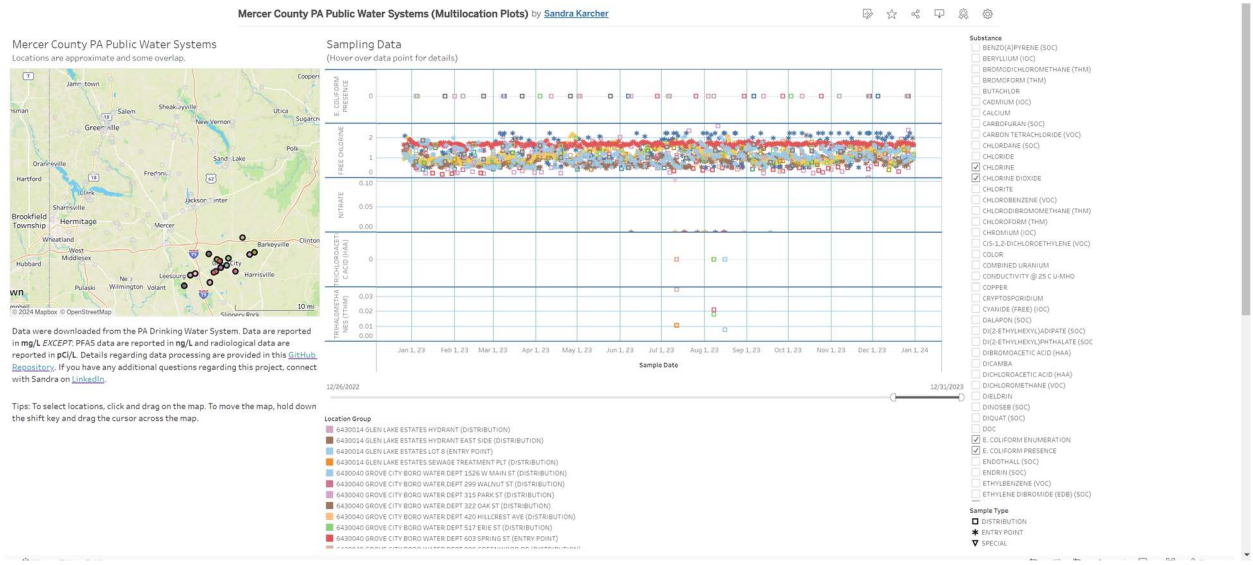

Figure 4: Viz Showing Samples from Multiple Locations on the Same Graph

(https://public.tableau.com/views/MercerCountyPAPublicWaterSystemsMultilocationPlots/Dashboard?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link)

The interactive dashboard allows the user to drag over the locations to select data from the corresponding sampling locations. Selected locations are displayed more prominently on the map. Data from the selected locations are plotted with the date on the x-axis and the amount of substance on the y-axis. An adjustable timeline is available at the bottom of the plots, allowing the user to zoom in/out to a timeframe of interest.

When first working with Tableau, it can be a little challenging to figure out how to write expressions and choose how to add columns and rows to a worksheet, but these are common tasks in Tableau and resources are highly available for helping new users learn how to accomplish them. There is a series of videos provided by Salesforce that can serve as a refresher or jump start and I highly recommend them for new users (https://public.tableau.com/app/learn/how-to-videos).

In general, I was able to find additional resources using a simple search. For example, when trying to figure out how to connect the map with the sampling data, in Google search, I entered: *tableau use a map to select data in another sheet*. Scrolling through the returned results, I found, *Dashboard with interaction between two sheets* (https://community.tableau.com/s/question/0D54T00000C5k9gSAB/dashboard-with-interaction-between-two-sheets). From this article, I learned that I needed an "Action" to make the connection and was able to perform additional searches to find the information needed to get the connection working. As another example, to learn how to put samples from multiple locations on the same plot, in Google search, I entered: *tableau how put more than one thing on a plot*. Scrolling through the returned results, I found, *Add Axes for Multiple Measures in Views* (https://help.tableau.com/current/pro/desktop/en-us/multiple_measures.htm), which included the information I needed.

A few notes about using Tableau Public:

- When I added multiple normalized separate files, Tableau joined them into one flat file, so I found it easiest to add one flat file as the data source.
- Tableau has an undo button, which is nice.
- I added the latitude and longitude of the sample locations to the data source file. Tableau Public could not place locations on a map using an address.
- One point of frustration was that, when searching for resources, it is common to share twbx files. These do not work in Tableau public.

Overall, I found Tableau Public fairly intuitive to use. The series of videos provided on the Tableau Public website helped me get started quickly. When I could not figure something out by trial and error, I was able to find the resources I needed to accomplish my project goals.

My Experience Using ArcGIS Online (Free Version)

ArcGIS Online offers free and paid options. I thought it would be interesting to compare working with Tableau Public (which is free) to working with the free version of ArcGIS Online. Based on my experience looking for training resources, many of the tutorials for learning how to use ArcGIS Online are targeted towards the paid versions. At times, this made it difficult to tease apart user error from the limitations of working with the free version of the software. After reviewing multiple online resources (and with a lot of trial and error), I determined that I could not accomplish my project goals using the free version of ArcGIS Online. To create charts in ArcGIS Online, a feature layer is required. Creating a feature layer from a csv file is not supported in the free version of ArcGIS Online.[1]

To confirm this functionality was not an option using my free account, I logged into my account, clicked on Content, then clicked on New Item. The only options available to me are URL and Application. There is no option to create a hosted feature layer.

I was able to add a csv file (with latitude and longitude fields) directly to a Map Viewer file. Once added, I was able to plot the sample locations on the map and add a chart to the location pop-up, however, the chart could only show a single value for each attribute in the underlying table, it could not show data as a function of a date field. I could not find an alternative way to generate time series graphs from a csv file or find a way to transform the csv file into a feature layer.

To further examine the limitations of the free version of ArcGIS Online, I created a new map and added an existing feature layer (Figure 5).

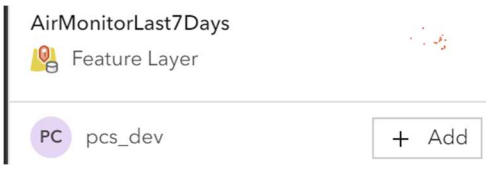

Figure 5: Feature Layer Used for Exploring ArcGIS Online

[1] Feature Layer References:

- Feature layers and tables saved in a group layer are supported if they are hosted in ArcGIS Online. Creating charts is not supported for CSV layers, feature layers with location sharing, hosted imagery layers, or map image layers (https://doc.arcgis.com/en/arcgis-online/create-maps/configure-charts-mv.htm)
- Requirements for making a feature layer: ArcGIS organizational account with a Publisher, Facilitator, or Administrator role (https://learn.arcgis.com/en/projects/create-a-layer-and-add-features/)
- Tutorial: Import data to create a feature layer: Under Data management tools for ArcGIS Online, create a hosted feature layer (https://developers.arcgis.com/documentation/portal-and-data-services/data-services/tutorials/tools/import-data-to-create-a-feature-layer/)
- In https://www.youtube.com/watch?v=l2nhhOw7HW0, at 0:30, the screen shows several more options than are available when using the free version of ArcGIS Online.

Using the existing feature layer, I was able to add a chart to the pop-up, but just as when using a csv file, the chart only showed single values for attributes in the feature layer's table. Unlike with the csv, when using the feature layer, I was able to add time series charts. I made six charts: five line and one bar. The line charts were generated using the RecordedDate field for the x-axis and a measured value for the y-axis; the chart was split by location. For the bar chart, the RecordedDate field was selected for the category and the measured values were selected for the y-axis. The bar chart does not support the split by location.

A few notes about using ArcGIS Online Map Viewer (free version):

- The line charts automatically connect the data with lines. The formatting options for the symbols on the line chart appear to be limited (https://community.esri.com/t5/arcgis-pro-questions/how-to-remove-markers-from-line-graph/td-p/1259790).
- The number of points (or bars) that can be displayed on a chart is limited. If not using aggregation, the dataset may need to be prefiltered. With this dataset, a filter of two locations was applied when exploring charting without aggregation. The filter was removed when the charts were aggregated to display means.
- Each chart was created individually. It would be preferable to have a way to automatically create multiple charts in a grid. This functionality is available in ArcGIS Pro (https://www.esri.com/arcgis-blog/products/arcgis-pro/analytics/grid-layout-charts/), but does not appear to be available in any version of ArcGIS Online.
    - The underlying structure of the data in the feature layer uses a separate column for each constituent. The water dataset in the csv file is structured so that the substance is indicated in a single column with an associated measured value in another column. This structural difference would affect the way the charts are constructed.

The free version of ArcGIS Online includes the option to create a web app and offers several starter templates. For this project, I used the Chart Viewer template. This template allows display of a map and up to three charts. A user can limit the data to display in a chart by zooming into a targeted area of the map, then clicking the icon to display data by extent in the sidebar of a chart. The user can select which charts to display by selecting charts from the panel that lists all the available charts (Figure 6).

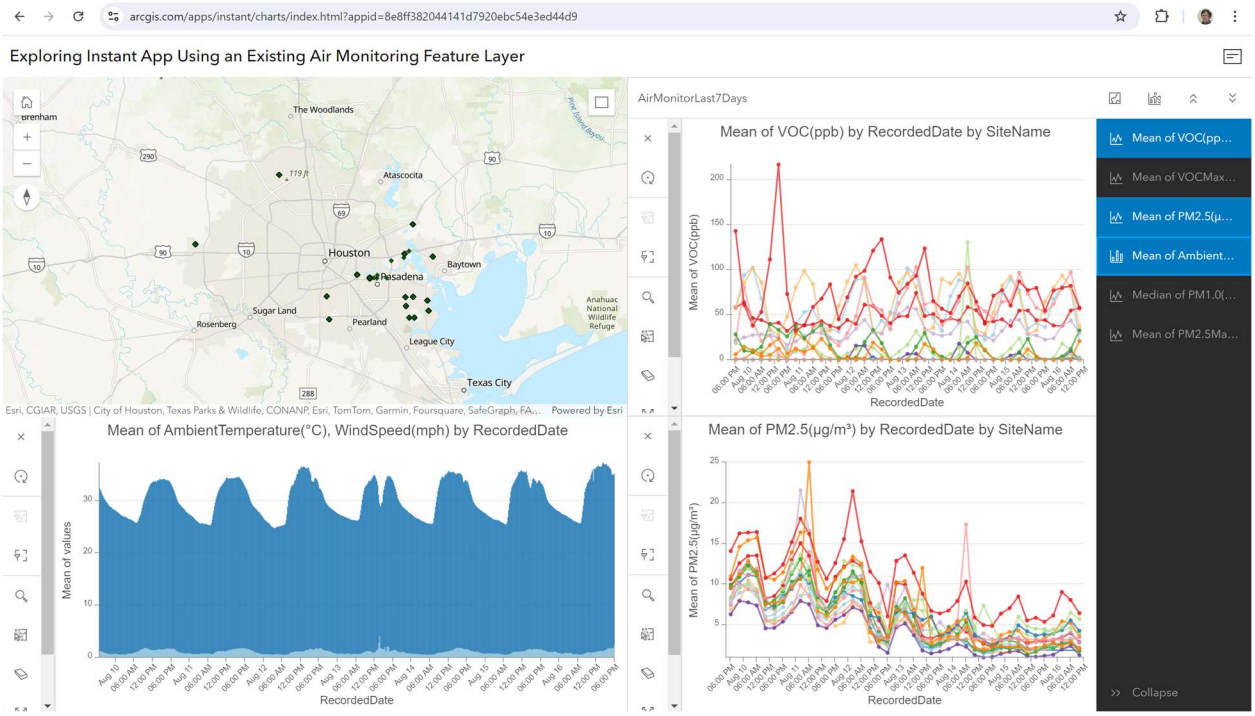

Figure 6: Exploring ArcGIS Online (Free Version) Using a Feature Layer

(https://www.arcgis.com/apps/instant/charts/index.html?appid=8e8ff382044141d7920ebc54e3ed44d9)

There are still several issues with this app that are not optimal. These include:

- A legend can be shown on the charts, but the order of the charts changes with the chart selection and putting a legend on every chart would waste a lot of space.
- It would be preferable to have an easy way to make the color of the symbols on the map match the charts when they are split by location.
- When clicking on the locations in the map, the pop-up is sometimes displayed off the screen, requiring panning to see the information in the pop-up.

The free version of ArcGIS Online offers many options for sharing geographic data; however, uploading data via a csv file limits options for graphing. Using an existing feature layer opens access to more functionally; however, there were still limitations that made it difficult to present the data the way I wanted to.

I invite you to email me at SandraKarcher44@gmail.com with comments, questions, and/or corrections.