DSE 241 – Data Visualization

Submitted by Alejandro Hohmann
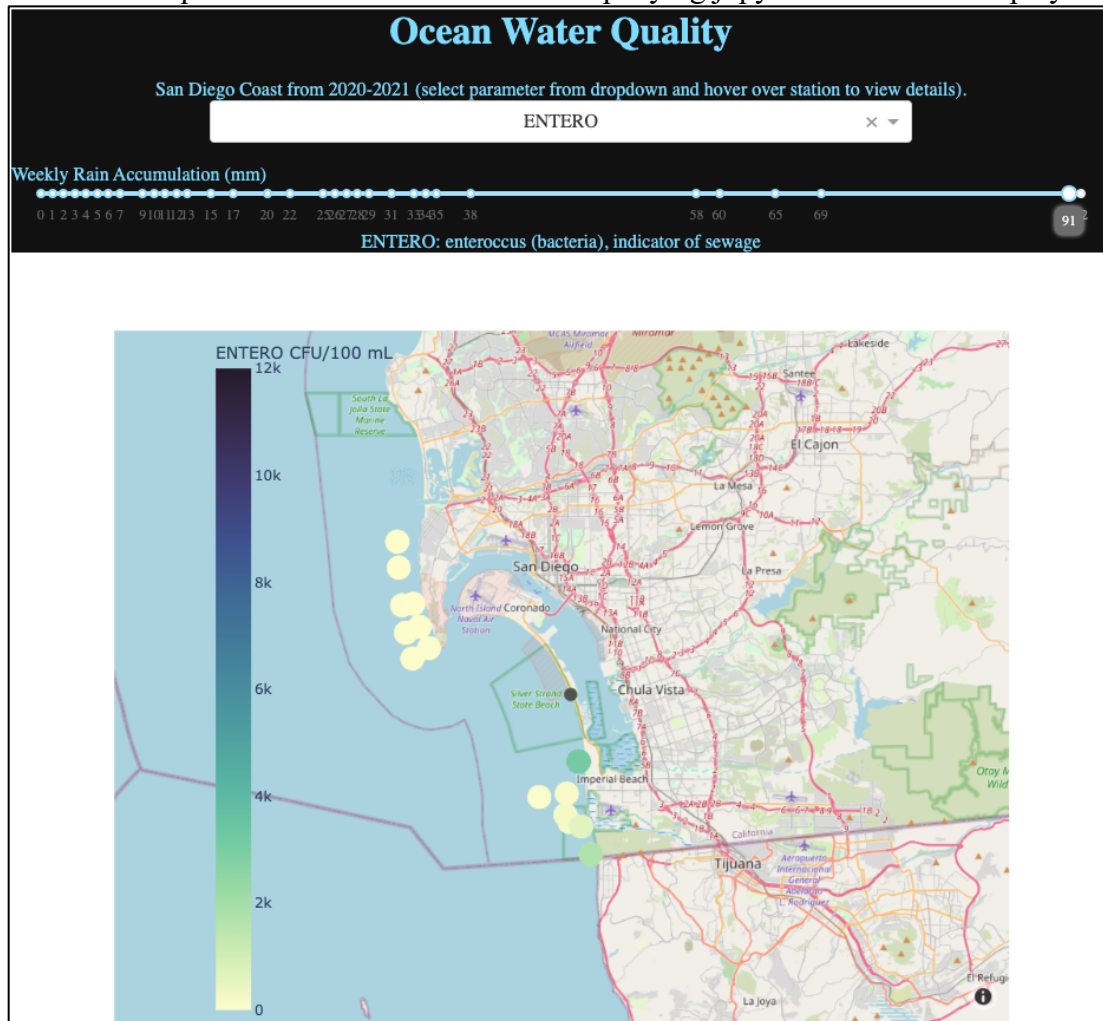
Final Project: *San Diego Ocean Water Quality*

Report Contents
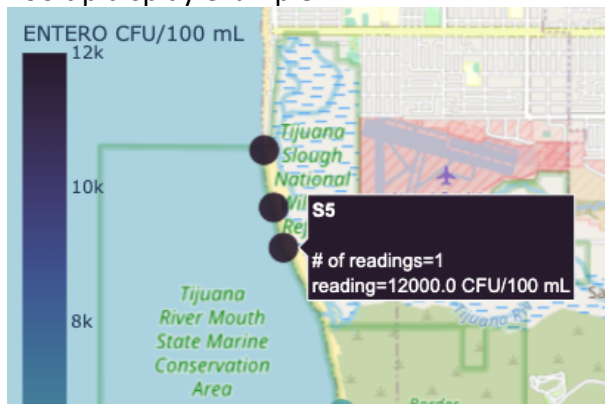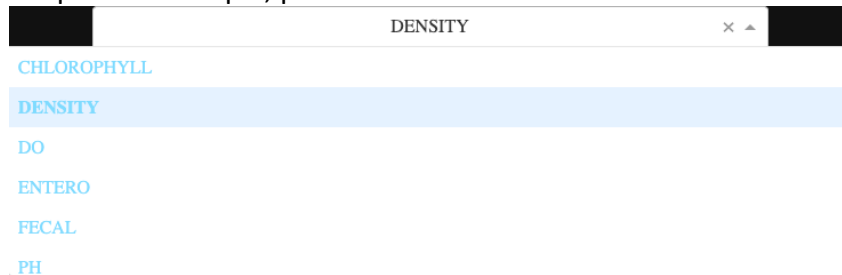
The Viz

Actual viz requires interaction. Run the accompanying jupyter notebook to display in browser.
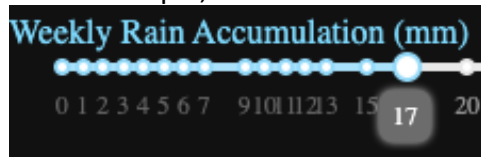

Tooltip display example:

Dropdown example, parameter selection:

Slider example, rain amount selection:

Motivation

This visualization is intended to edify users on:
- which San Diego Ocean Shore locations have higher bacteria levels
- how bacteria levels relate to rain

Datasets and Augmentation

Datasets:
- Measurements – samples of ocean water
  - Source: San Diego Ocean Monitoring Program
  - 104 sites
  - 15,000 samples
  - 10 parameters
    - Chlorophyll
    - Density
    - Entero
    - Fecal
    - Ph
    - Salinity
    - Temperature
    - Total (unclear from dataset what this measures)
    - XMS
- Locations – latitude and longitude of measurement stations
  - Source: San Diego Ocean Monitoring Program
- Rain – daily rain totals (mm) measured at San Diego International Airport
  - Source: NOAA using `meteostat` python package

Augmentation
- merge Measurements dataset with Locations dataset on *station*

- calculate weekly rain accumulation
- merge Measurements dataset with Rain dataset, requires datatype manipulation
- calculate absolute maximum for each parameter for consistent colorscale
- create and merge feature dictionary for plotting context
  - dataset does not include definitions of the parameters so sourced definitions from a [separate report](#) from the Ocean Monitoring Program
- create string labels with measurement scale for tooltip (e.g. degrees Celsius; parts per 100/mL, etc)
- bin the rain accumulations for easier comprehension (e.g. 1.8mm = 2mm)
- calculate number of measurements at a site (required at runtime due to filtering)
- rename/map column names to be friendly for user / plot
- set css parameters for chart elements to be user-friendly
- utilize mapbox layer that has detailed coastline polygon

Script created so that if new data were to be added then lat/long, min/max, dates ranges, etc would all automatically adjust to any new values included in the raw data.

Tasks

The user is required to:
- select a parameter (e.g. FECAL; DISSOLVED OXYGEN)
- select a rain accumulation
- zoom in/out to desired area (autoscaling initializes with all points)
  - due to a bug in Dash, the graph window size resets to a small size when new parameter is selected but adjusts to appropriate scale when the browser window size is changed
- hover over station to view details

The chart initializes with Temperature and zer0 rain so there is data present even if the user does not interact.

Expressiveness of design

The color channel of the scatter chart and the mark size of the size channel of the marks do an excellent job of immediately highlighting differences between the stations. The viz is straightforward and easy to orient as a user. A layperson or an expert could intuitively use the tool and make inferences with the encoded information.

The mapbox layer is rich with context. In some cases it contains too much information (e.g. the highways include exit numbers which are probably irrelevant for this tool), however configuring/disabling detail layers with this free version are available.

Effectiveness of the solution

Color Channel – measurement scale (e.g. temperature ranging from 0° C to 25° C)
- indicates how high/low a reading is for a particular parameter
- encoded on a continuous scale with lighter color indicating lower value and darker color indicating higher value
- this information is encoded in the marks themselves as well as a colorbar on the viz for reference
- the user can very quickly interpret that lighter color means lower value and darker color means higher value
- because most of these measurements are related to bacteria, in general, higher value is bad. The higher/darker marks stand out on the viz and the continuous color scale does a great job of quickly conveying a lot of information

Size Channel – mark size of stations
- indicates how many readings there are at a particular station
- though it's not the most important channel for this viz's purpose, the meaning of the size channel is not explicitly stated anywhere on the chart which may confuse the user
- the number of measurements is, however, included in the tooltip. Generally there are more readings closer to shore than further from shore. By hovering over points and comparing the information encode in the tooltip, a user may ascertain its meaning.

Text Channel – tooltip
- this is important for users to get more detail on the measurements. While the color channel does a great job of differentiating low/high values and the size channel the number of readings, a user who wants more detail can hover over various points to get exact details and further draw conclusions if they wish.

Speed & Scalability
- The solution is fast. Interacting with the fields changes the results with near zer0 latency. It is also scalable. New data, both station measurements and measurement types, can easily be added and the chart would dynamically encode the additional data without additional configuration required.

Interaction

Dropdown
- the user can select the parameter they want to view (e.g. SALINITY; CHLOROPHYLL)
  - initialized with TEMP
- selecting a parameter dynamically changes the color channel and size channel with the data filtered to just these metrics. It also provides a brief description of the parameter so that the user understands what measure they are viewing.

Slider
- the user can select the rain accumulation by adjusting the slider from left-to-right, ranging from zer0 mm of rain to the maximum amount of rain in the data (92 mm)
  - initialized with zer0 mm

- selecting a parameter dynamically changes the color channel and size channel with the data filtered to just measurements with this corresponding rain total metrics.

Hover
- the user can hover over specific stations to get details for the filtered data
- originally the data included lat/lon in the tooltip but removed for elegance. Could be added back to the solution depending on the audience.

Zoom
- the user can zoom to any part of the map
- the mapbox labels/details dynamically change with zoom, so as a user zooms in to the map they can see a neighborhood on the land (e.g. Ocean Beach, Imperial Beach, etc) which provides context on the measurement station.

Conclusions

This implementation is effective at conveying different types of bacteria in measurement areas. It is intuitive and straightforward to use. The solution perfectly addresses its original motivation. By interacting with the parameters, it highlights that there is generally an increase in bacteria levels as rain accumulation increases, especially closer to the border. A user can toggle to FECAL and ENTERO, for example, and quickly see the darker marks, indicating higher levels of bacteria, closer to the border stations and darker marks when there is more rain selected in the slider. Conversely, there are lighter markers, indicating lower levels of bacteria, when the station is further from the border and there is little-to-no rain. The solution is fast and scalable and could be used as a tool for a professional audience.