Jen-Li Chen's Website

View on GitHub

2020 Chesapeake Bay Hackathon *Hack the Bay*

1. Introduction

The Chesapeake Bay watershed is home to over 18 million people, spanning 6 states and Washington DC. The Bay is the largest estuary in the United States, providing 500 million pounds of seafood harvest each year. Declining water quality due to pollution and overharvesting in past decades led to poor ecosystem health and the near extinction of economically and culturally important aquatic species. In response to this decline, state and federal partners have invested heavily in restoration efforts, amounting to nearly $1.5 billion of state and federal funding in 2019 alone. Despite many years of costly efforts, warming temperatures, sea level rise, and increasing precipitation due to climate change have hindered the attainment of restoration goals. (Citation)

1. Data Collection

I chose **Water Temperature** as the indicator. I used the CMC and CBP water quality data set provided by Booz Allen Hamilton. For the water temperature, all of the filtered data sets for the analysis and visualization are from the CMC dadabase. (Link to the raw data)

1. Data Preprocessing

   - Year: I used Python pandas to trim date and time elements from the **Date** parameter, and filtered 2015 - 2020 as the time window I am interested in.
   - Season: I categorized May to August as **Summer** , and November to February as **Winter**.
2. Exploratory Data Analysis

   - Methodology:
   1. Stationality:

      Use AF, PACF and Augmented Dickey–Fuller test to check stationality. The water temperature is stationary time-series data. (Citation)

   2. Summary Statistics:

      After plotting water temperature within 2015 to 2020, I found the temperature range for the winter has something interesting to look at. Especially after 2017. For Example, the following scale of the winter water temperature reveals the increasing range:

2017 Winter 2018 Winter 2020 Winter

1. Geospatial Maps:

   I used the 5-year water temperature for the color scale, so I would be able to see whether a certain range of temperature increases in certain area. In 2020, there are several unusual high temperature points.

2015 Summer CMC 2016 Summer CMC 2017 Summer CMC 2018 Summer CMC 2019 Summer CMC 2020 Summer CMC 2015 Winter CMC 2016 Winter CMC 2017 Winter CMC 2018 Winter CMC 2019 Winter CMC 2020 Winter CMC

3. Data Visualization

   - I created a seasonal overview dashboard in Tableau, using year, season and state parameters to invite end users to take a look at the data set. Additionally, With Tableau's geospatial mapping feature, I was able to check on the land use in the above region, The Yellow Breeches Creek in Pennsylvania, which is a commonly fishing resort (Reference)
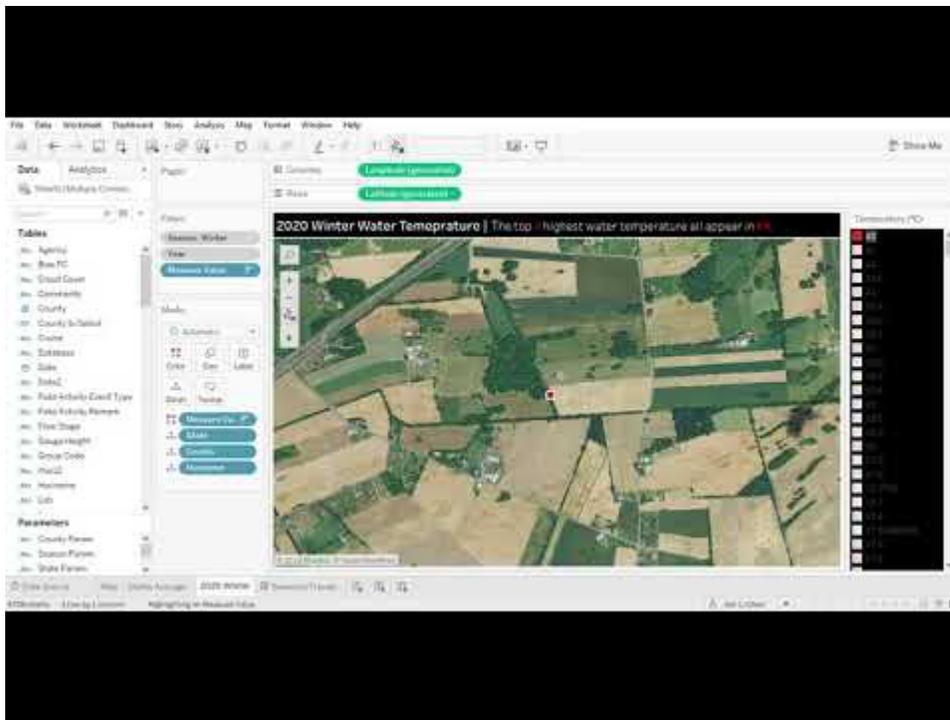
4. Challenges

   - Not all the states in the Chesapeake Bay area have comprehensive data in the 5-year window.
   - Most data sets are in large magnitude, and hope to be able to access data sets in the FTP in the near future

5. Recommendation

- Due to the time limit, I had not yet got time to dig into the residents occupation, conductivity, recreation and harvesting products in the Yellow Breeches Creek in Pennsylvania. From the satellite iamges, I saw most of the land use there is in agriculture. I assume agriculture and fishing might be the major land and water use there.
- I should have checked out on the data collection time in the specific region, as well as weather condition.
- I had not checked on the air temperature and should have mapped a confusion matrix to check on the correlation between the air temperature and the water temperature.

6. Conclusion

In conclusion, I was super thrilled to be able to touch on this topic and have a chance to look at the data sets, read the background documents, and work with my great teammates in the 2020 Hack the Bay Hackathon. Since the state and federal restoration efforts and funding have poured in since 2019, in the summer in 2020, the water temperature did not show high deviation from the previous year, though the winter temperature had several outliers. The top 2 highest degree are in Pennsylvania, and third highest is in Maryland. However, the highest one is 43 degree Celsius, while the second and the third are 26 and 24 degree Celsius. On the "Yellow Breeches Creek, Pennsylvania Fly Fishing Reports & Conditions" website, it did mention about high water temperature, but it stated the water quality is good. As a result, I propose the 43 degree Celsius in the winter might not be a typo from data collection. I look forward to expand this case study further in the days to come.

Jen-Li-Chen.github.io is maintained by jenli810006995.

This page was generated by GitHub Pages.