Analysis of Effects of weather changes in California and its coastal line

Sihua Cai, Jaehyeon Park, Yuxi Jiang, Ankita Giri

June 6, 2022

1 Introduction

In this project, we observe climate changes in California and its coastal line by taking a look at various weather patterns. For the purpose of observing patterns and effects of weather change over years, we use publicly available data sources to scrape data. To streamline our analysis of weather effects we choose only certain variables and do five different analyses among those variables. The following are the analysis questions that we observe:

- Is there a relation between water temperature change and Fishery in California Coastline?
- Is there a relation between average water level, precipitation and the temperature change in California water?
- How did Droughts in California change over time?
- What are the effects of Ocean Acidification in the coastline of Calilfornia?
- Is there a relation between precipitation and salinity of ocean water?

The above analysis questions pertains to assessing climate change in both land and water. The use of data for the project is primarily accessed through NOAA API, other sources from the NOAA data discovery portal and through California Department of Fish and Wildlife.

2 Data Scraping

2.1 Fishery and Water Temperature along the California Coastal Line

A. Data set Collection Sources:

Water Temperature:

Using NOAA CO-OPS API for Data Retrieval, the water temperature data was accessed. NOAA CO-OPS data retrieval can be used to retrieve observations and predictions. The CO-OPS stations record 19 variables that are water related. For the purpose of the analysis, water temperature data from the year range 2000-2014 for four stations in each of Northern, Mid- Northern, Central and Southern California was selected. The selected location of the four stations are: Crescent City, North Spit, Monterey, San Luis and San Diego.

The stations chosen for Northern California are Crescent City and North Spit and merged for the missing years. Crescent city lacked values for the years 2005, 2006, 2012 so the years were taken from North Spit and merged.

Fishery:

The California Department of Fish and Wildlife has an extensive report of fish catches from the year 2000-2019. California is divided into nine areas for the purpose of reporting fisheries statistics. Each statistical area is named for a major port within its boundaries. [1] The fish catches are reported to twelve fishing port areas every year. The ports are: Eureka Area, San Francisco, Monterey, Santa Barbara, Los Angeles, Sacramento Delta, Bodega Bay, Fort Bragg, Inland Waters, Morro Bay, San Diego. All the nearby reported fish catches data is included in each port. The port's location ranges from Northern to Southern California along the coastal line and also includes Inland Waters and the Sacramento Delta Area which is not along California's pacific coastal line. Each port reports a monthly total landings of fishes(reported catch of fish) in pounds. Some fishes are landed in the port in eviscerated condition or beheaded. Hence, the data is collected according to the species of the fishes and the total summarized weight in the reported port. [1]

For the purpose of analysis, four fishery ports have been selected by dividing the California Coastal Line to four regions. The selected ports are: Eureka Area (Northern), Monterey Area (Mid- Northern), Morro Bay Area (Central), San Diego Area (Southern).

B. Data Scraping and Final Data set Explanation:

Fishery:

The fishery data has been accessed through the reported fish catch landings by the California Department of Fish and Wildlife. The reports are in a PDF format in the California Department of Fish and Wildlife website online here. Due to lack of API for accessing data , several data scraping methods have been used to automate the scraping method.

Scraping of the URLs:

Using the python library BeautifulSoup, all the href instances were accessed from the above mentioned website including the URLs to the PDF. The PDF links are then written into text files for each port which was later used to convert the PDF data into tabular format. Each link inside the text file contains data for the year ranging from 2000-2019.

Scraping of data from PDF to Tabular format:

Using tabula-py which reads PDF in a tabular format, fishery commercial landings data for each fishing port was extracted. Other functions created for extracting and cleaning the data was used. Tabula only extracts data from the first page of the PDF unless otherwise specified. Most PDF have at least 3 pages of data therefore the first three pages of the PDF was scraped.

The final combined fishery data set has 5799 rows and 4 columns. The rows are the species of the fishes. The four coulmns are: 'Species', 'Total', 'port', 'Year'.

- Species: column is for the reported species of the fish
- Total: column is the total lbs of the fish reported in a specific port in a particular year
- port: column is the port in which the total lbs of the specific fish reported in that year
- Year: column is the year in which the total lbs of the specific fish was reported

A total of 280 unique species of fishes are in the fishery data frame but some species may have been reported multiple time under different naming patterns. The year range is from 2000-2019. However, some fishery ports are missing data for certain years. Port 'Morro Bay' is missing data for years 2014, 2015, 2016, 2017 and Port 'Eureka' is missing data for years 2005, 2007, 2011, 2014.

Water Temperature:

The CO-OPS API has a limit of extracting data for only 31 days at a time so several functions were created to extract all the data for the year ranges from each station at once.

The final data frame extracted has columns:

- dateval: which contains date of the recorded water temperature in the format yearmonthday
- watertemp: mean water temperature in a day
- station: station location from which the data was extracted

- year: year the data was extracted
- month: month in letters from which the data was extracted

2.2 Average water level and climate change in California

A. Dataset Collection Sources:
We extract the historical record of water level heights from the tides and currents dataset on the CO-OPS website(Center for Operational Oceanographic Products and Services). Since there are a large number of Californian stations to choose from, we only select five of them to represent different geographical regions of California. These stations are San Francisco, Redwood, Santa Barbara, Los Angeles, San Diego from North to South, respectively. For each station, we scraped the available water level data from 2000 to 2021 and converted it into a Pandas dataframe.

Then, in order to gather the climate data in California, we accessed the NCEI website(National Centers for Environmental Information) and specifically chose the GSOM(Global Summary of the Month) dataset for our use. We specifically looked at the precipitation and temperature variabels which serve as a representation of climate in California. Once again, we picked the same set of stations for comparison.

B. Data Scraping and method explanation:
To simplify the process and improve efficiency, we created several functions for data scraping use. The first function, `url_to_dataframe()`, takes URL as the parameter, then requests the json file that contains all the necessary information, and converts it into a dataframe. The second function, `scrape_data_annual()`, takes parameters of station ID(`s_id`), Year, data product(`p`) and datum(`d`). Since the maximum time period to collect data from is one month, we manually created urls for each month and combined them together. The third funcion, `scrape_data_all()`, takes parameters of station ID, start year(`sy`), end year(`ey`), data product(`p`), datum(`d`). In this function, we scratch all the data from the user-selected year range, and combine them into one whole dataframe. Then, we calculated the average water level each year and reduced it into a table with only 22 rows, where each row represents a year in 2000-2021 and each column represents the mean water level in that specific year. After that, we apply the function into the five stations and get the corresponding information for each.

Again, we created three functions to scrape and process the data. The first function, `scrape_data2()`, takes parameters of station ID(`c_id`), start year(`sy`), end year(`ey`) and data product(`p`), then generates the json file from the url and converts it into a dataframe. The second function, `combine_data()`, takes parameters of station ID(`c_id`), data product(`p`), then combines all the available data from 2000 to 2021 and stores it in a dataframe. The third function, `plot_city()`, takes parameters of dataframe(`df`), county name(`c`), then plots the trends of three lines in one graph.

Using similar techniques, we converted the json file into a pandas dataframe and computed the annual precipitation and temperature on average between 2000 and 2021. After that, we merged all the dataframes for each station by index of year. At this point, we have all the information about the mean water level, precipitation and temperature information available in one table. The final dataframe for each station has 3 columns and 22 rows where each row represents a year from 2000 to 2021, and each column represents the average water level; precipitation; temperature in that year. Then, we visualized the trend of these variables across a twenty year period to observe if there are any notable patterns or potential correlations.

2.3 Drought condition in California

A. Dataset Collection Sources:
Drought in California has been a on-going problem for decades long due to the increasing temperature and decreasing rainfall. However, people often underestimate the harm that it could cause. We are interested in investigating the recent trend of drought statistics in California for the past twenty years. Our main data source is from the U.S. drought monitor website where the comprehensive drought data is provided. It maps out the drought data across different parts of the U.S. territory and references from local observers and other reliable online sources to update information each week. NOAA is one of the three organizations that take part in this project. According to the official definition, there are five categories of drought by severity. [2]

- D0: Abnormally Dry;
- D1: Moderately Drought;
- D2: Severe Drought;
- D3: Extreme Drought;
- D4: Exceptional Drought.

B. Data Scraping and method explanation:
To simplify the data scraping procedure, we created a function `drought_level()` that takes 2 parameters, year(`y`) and Drought Level(`l`), then requests the json file and computes the average percentage of areas that fall in each category and number of days in drought each year. Lastly, we store the information in one table. By applying the function to each year's drought data, we eventually collected 21 tables and combined them into our final dataframe. It has 21 rows and 10 columns, where each row represents a year from 2000 to 2001, each column represents the average area percent or number of days that corresponds to one drought level in that year. Then, to visualize the distribution of each drought

level in each year and a general trend of drought conditions in California, we created two functions to plot the data. The first function, `lineplot_drought()`, takes the dataframe(`df`) as a parameter and draws the line plots of the two variables(Number of Days and Avg Area Percent) in a twenty-year period. The second function, `barplot_drought()`, also takes the dataframe(`df`) as parameter and draws the histogram plots that shows the relative distribution of each drought level in each year.

Lastly, we also accessed the 'ConsecutiveNonConsecutiveStatistics' dataset on the website and scraped the record data of the number of nonconsecutive weeks in drought across all the California counties. Then, By sorting the values in ascending/descending order, we found the top three counties that suffer the most from drought and the top three countries that suffer the least.

2.4 Ocean Acidification in California

A. Dataset Collection Sources:
Ocean Acidification is one of the most significant environmental issues in the world. There is little change in acidity in the ocean, but a slight change in acidity can cause a big difference in the marine ecosystem. We are interested in investing in the recent trend, cycle, or irregularities between ocean temperature and ocean acidity for the last ten years. We extract the time-series dataset of Ocean Acidification on the `NOAA` website (link). There are only a few time-series acidification datasets on the `NOAA` website in California: The other dataset only includes from 2013 to 2016, so we chose this dataset.

B. Data Scraping and method explanation:
This dataset does not have an API, so we use several functions by ourselves for data scraping. First, we used the `get_urls` function to input URLs from the website. Next, use the `open` code to save CSV files, import CSV files, and make dataframes. Dataframes had PH variables, but PH is a log function, so even if there is a difference in acidity, there is no significant difference in PH in this Dataframe. Furthermore, there is so much NaN data, except for `xCO2` variables, so it is not meaningful. About the temperature part, there are `Licor temp`(Temperature of the Infrared Licor 820 in degrees Celsius) and `SST`(Sea Surface Temperature) variables. However, `SST` has so much NaN data, so we used only `Licor temp` in this analysis. Therefore, we used Dataframe (ACID) with `Date`, `Licor Temp`, and `xCO2` (partial pressure of carbon dioxide) variables. In this process, there was an issue using `skiprow`. For the Mac environment, we used `skiprow = 4`, and for the Windows environment, we used `skiprow = 8`. Finally, we use the `pandasql` package to make our Dataframe into `SQL` query. We used `Groupby` to classify by `year`, `month` and `day`. Each dataset has a different month and day every year, so we used the average value of `xCO2` and `water temperature`.

First of all, we visualized time series by year. However, we only have ten years, so visualization by year is pretty simple. Therefore, we made another visualization by month to find more specific trends and prediction. For visualization, we used `Tableau`.

2.5 Salinity of Ocean Water and other Ocean Water Variables in Relation to Precipitation

A. Dataset Collection Sources:
For this part of the analysis, we obtained data from two primary sources, the National Oceanic and Atmospheric Administration (NOAA) and one of its sub-branches, National Centers for Environmental Information's (NCEI) Global Temperature and Salinity Programme. From the NOAA dataset, we once again scraped the measurements of precipitation from 5 California coastal cities. From the NCEI dataset, which consists of data points collected from ship gauges, we scraped the following oceanic water and gaseous measurements:

- Salinity: ratio of salt to water in the ocean measured in Practical Salinity Units
- Mole fraction of CO2 in parts per million, measurements taken from ship equilibrator and the air
- Temperature of water, measurements taken from ship equilibrator and ship thermosalinograph in degrees Centigrade
- Barometric pressure in hectopascals
- Fugacity of CO2 in air, sea water, and equilibrators in microatmospheres

Unfortunately, this dataset is very limited in terms of its coverage. Among all the public datasets in NOAA and NCEI and even other environmental organizations, this is the only one that contains any information on ocean water salinity. It only covers a range from March 18, 2000 to November 13, 2000, with some values missing in between that date range, especially in the months of May and September. However, we will still use this dataset as we believe it contains important information that pertains to much of ocean biodiversity, as many of these variables are descriptors of the aquatic environment that many organisms live in.

B. Data Scraping and method explanation:

To extract the data, we used the requests library in Python and cURL requests in the command line.

For the scraping of precipitation data from the NOAA dataset, we used the `requests.get()` function along with custom functions to optimize the creation of a comprehensive list of API request URLs. Since we wanted to query multiple stations in multiple months, we stored the API URL filter fields and values in a Python dictionary called `url_param`. That way, the values could easily be modified to access a portion of the dataset of our interest through the custom `change_station()` function, which we used to cycle between 5 coastal stations: Eureka, Santa Cruz, Santa Maria, Los Angeles, and San Diego. We also created functions to set and reset dates. After the setting the values in `url_param` to our desired parameters, the custom `join_url` function can be used to concatenate the base URL with the most current version of `url_param` for the full API request URL, which was then easily passed to the `requests.get()` function along with the NOAA access token. The return value was then converted into a json format and parsed into a Pandas dataframe. We then formatted the data in a table format to contain 5 columns for each station's daily precipitation summary, where each row was for the date.

We had to take a slightly different approach for the NCEI dataset. We collected the data through cURL requests in the Command Line and saved the output as a csv file. Because of the finnicky nature of this specific data access tool, we could not filter for specific variables in the request, so all data cleaning and filtering were done with Python in Jupyter Notebooks. Some of those tasks included removing irrelevant or repetitive variables and replacing NA values (represented as -999.99 in the original data) with NaN values that the Python libraries would have an easier time working with. We also normalized all values in both datasets so that we could plot the variables against one another later on.

3 Analysis and Results

3.1 Fishery and Water Temperature along the California Coastal Line

Water temperature change is known to impact fisheries. The migration patterns of the fishes change according to temperature changes. Reported fish catches increase more during warm temperature years for some fishes while some fishes are reported less and vice versa. [3]

Our data include fishery information from the years 2000-2019 with some years missing for a few ports. The water temperature data however is only from the years 2000-2014 (also missing some values for some regions). This has posed challenges in finding succinct patterns among water temperature change and fishery. The utilized way to visualize the patterns for fishery in the report is to identify the years when water temperature trends are notable and take a look at the fishery data for the particular year.

Temperature:

Each port has its own mean temperature patterns which are not consistent among all the ports. A mean temperature rise in the year 2014 was the most notable pattern however among all ports. For this reason, the temperature change is observed in each month per year for every port. Table 1 shows that the temperature trends among all the ports for the year range 2000-2014. One most notable pattern recorded temperature is seen to be higher for the year 2014.

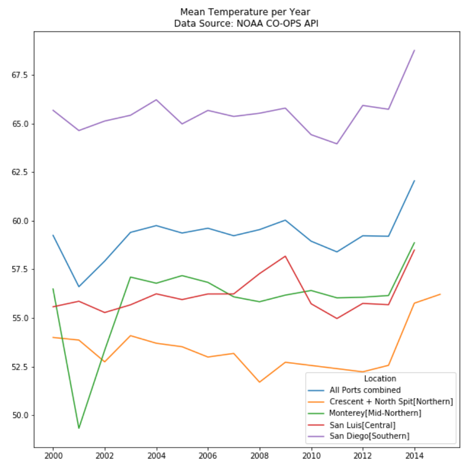

Figure 1: Mean Temperature Per Year in each California region

California Region	Station	Year	Some Visible Pattern
Northern	Crescent+ North Spit	2014;2004	Highest recorded temperature in summer months/ Overall general increase in temperature in 2014; High temperature in summer month than other previous year
Mid- Northern	Monterey	2001; 2002; 2007 2014;	Record low temperature in April; Record low in March, April; High temperature in 2007 summer; Overall General Increase in Temperature(2014)
Central	San Luis	2008-2009; 2010	Overall high temperature in 2008-2009;Record high than in winter temperature;
Southern	San Diego	2007;2014	Record high August Temperature; Overall general high temperature recorded in 2014

Table 1: Temperature Trends in Each Port (Year Range: 2000-2014)

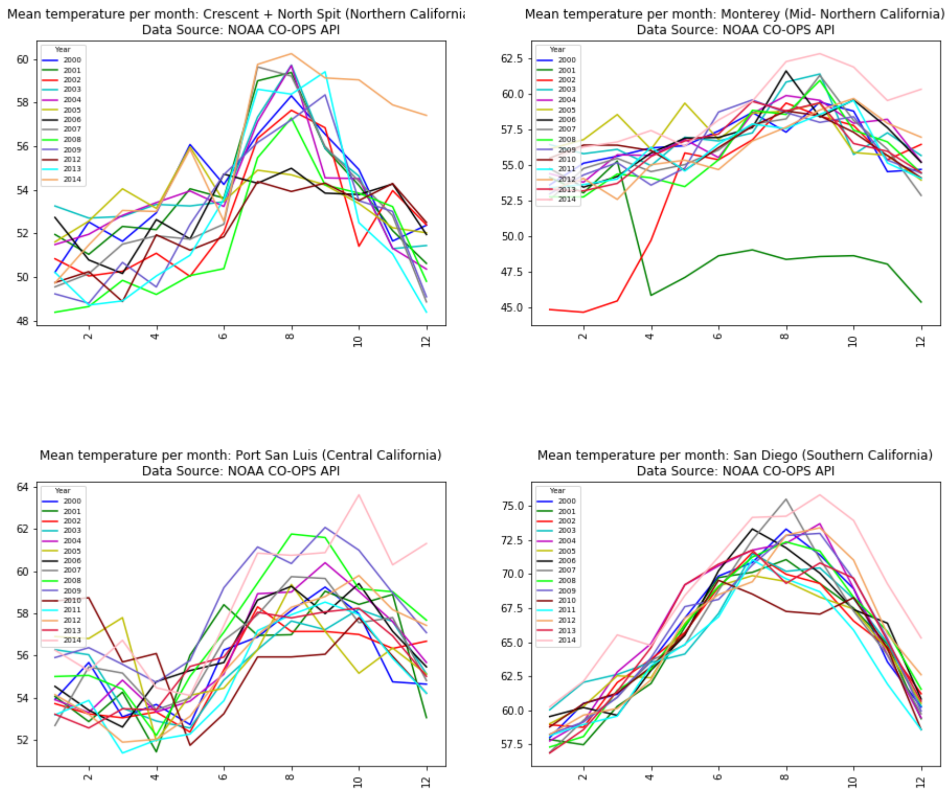

Table 2: Mean Temperature per month in each California Region (Year Range: 2000-2014)

Fishery:

Due to a large number of features, 280 unique species of fishes count and year ranging from 2000-2019 for the fishery data, various methods needed to be applied for generating interpretable results. First, a look at fishes that are present in all the years for each port is observed. Then, a view of cluster (using hierarchical clustering- complete linkage) years among the species is observed to see if there is any notable pattern among the species that are present in every year. Then, a total of 2 species of fishes are present in all the ports so we take a look at the reported fish catch for those fishes in each port and observe trends spanning the year range 2000-2019. Lastly, we observe trends in new species that were different from the previous year. The analysis was performed for each California Region.

Fishery and Water Temperature

Fishes that are present in all the years in each port:

As a form of exploratory analysis, we use a complete-linkage agglomerative hierarchical clustering method. The goal is to observe which years are in clusters in terms of reported fish catches and see if the temperature trends in those years are similar. Since the number of clusters or patterns is unknown in advance, the method is appropriate. We then view the Dendrogram to look at clusters. The sample index of the years is used in the dendrogram. 0 is for the year 2019, 1 for 2018, and the final index value 18 is for the year 2000.

Southern California:

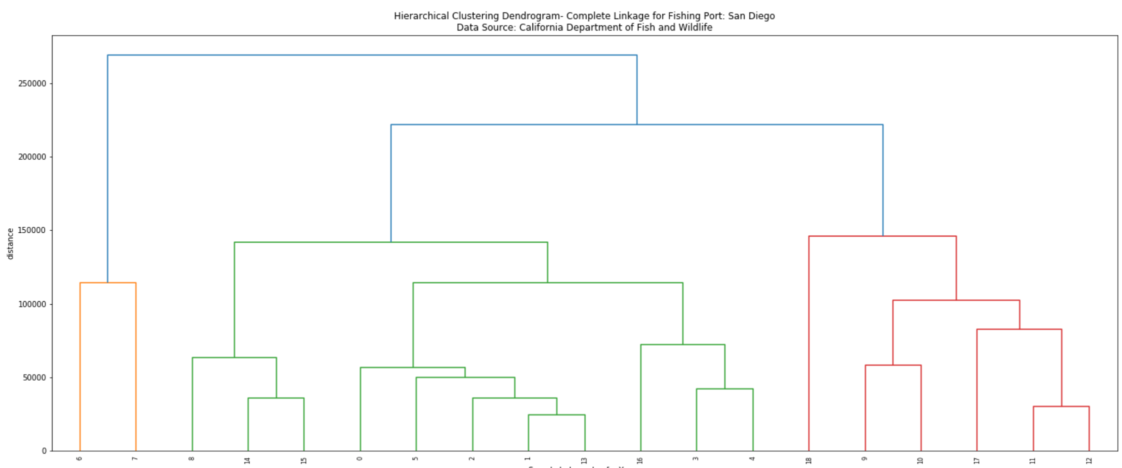

Figure 2: Dendrogram Showing Cluster years for Southern California (Fishery Data)

Notable Relation between temperature trend and the observed clusters:

In the cluster year 2010,2004,2003 the observed mean temperature trend is similar. All the three years have a high temperature increase than its previous years.

In the cluster year 2000,2009,2008,2008,2001,2007, 2006 the mean temperatures are relatively the same (can be seen flatter in a curve). Note: All the mentioned temperature trends are also observed through the plot of Mean Temperature per year above.

Central California:

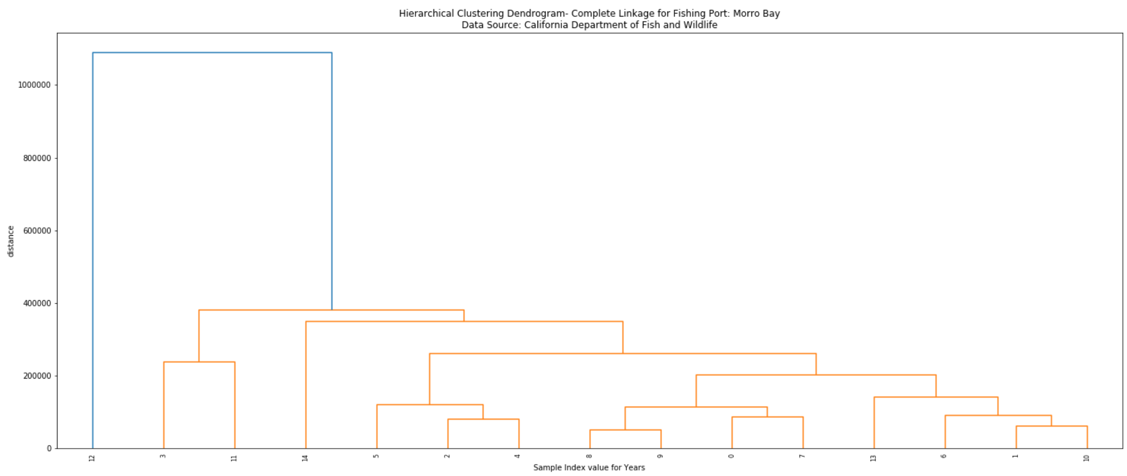

Figure 3: Dendrogram Showing Cluster years for Central California (Fishery Data)

Notable Relation between temperature trend and the observed clusters:

Cluster years 2000,2009,2012,2010 are sudden peak years for Central California.

Cluster years 2011, 2003 are sudden lower temperature points in Central California.

Mid- Northern California:

Notable Relation between temperature trend and the observed clusters:

Cluster years 2014, 2001 are sharp increase and sharp decrease in mean temperature respectively.

Northern California:

Notable Relation between temperature trend and the observed clusters:

No notable cluster relation with temperature in Northern California.

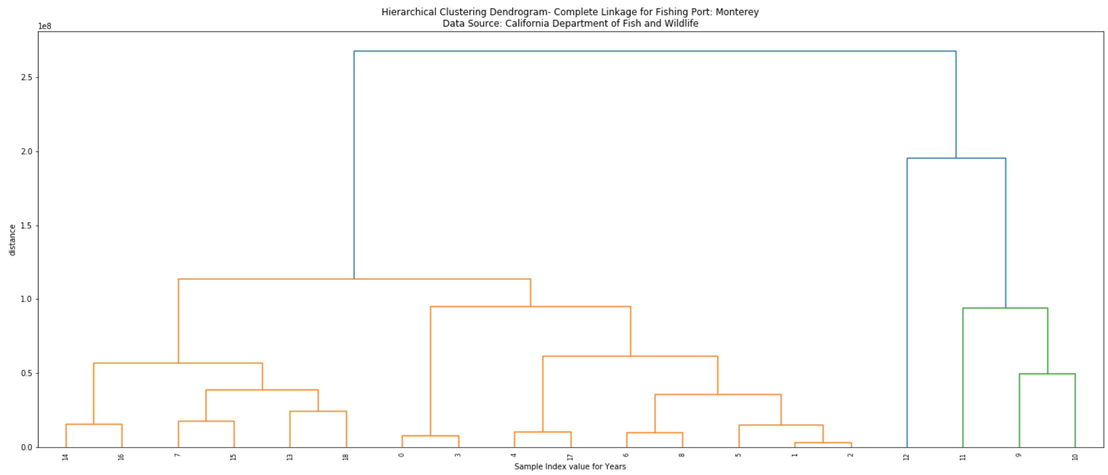

Figure 4: Dendrogram Showing Cluster years for Mid- Northern California (Fishery Data)

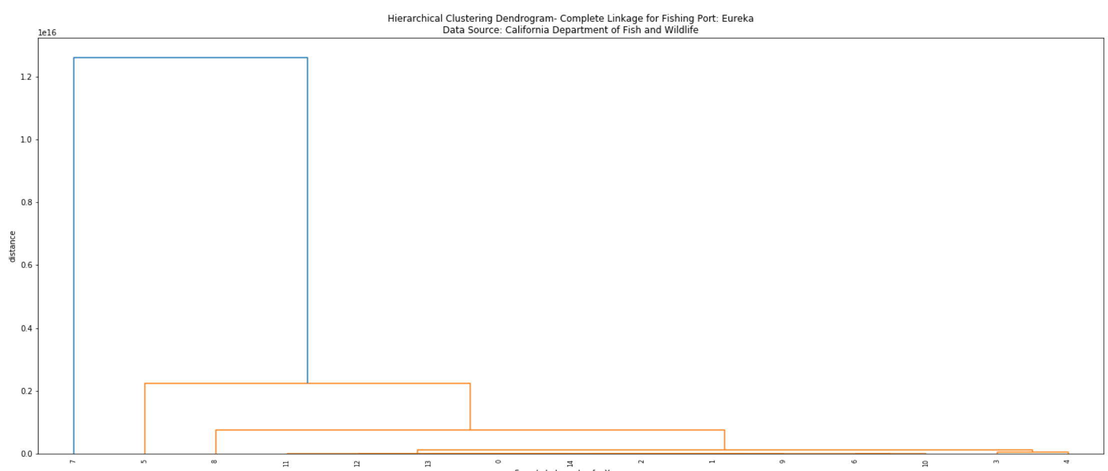

Figure 5: Dendrogram Showing Cluster years for Northern California (Fishery Data)

Limitations of the Cluster Analysis:

One of the biggest shortcoming in the hierarchical cluster analysis and comparing it with the temperature trend is that the stations and the port locations are not the same but are rather in the vicinity. This may have confounded the results. Also, some regions had missing values for the ports so this also the cluster results might have been different if the years were included. Complete Linkage hierarchical clustering is also highly sensitive to outliers as it measures dissimilarity value through euclidean distance. This could also have impacted clusters as seen in the dendrograms.

Species of Fishes that are present in all the port in the year span 2000-2019)

A total of 2 species of fishes are present in all the ports so we take a look at the reported fish catch for those fishes in each port and observe trends spanning the year range 2000-2019.

Lingcod:

For species Lingcod, larger value of the total reported catch is noted for the year 2003 which is also a peak temperature for Northern California(Port Eureka). However, its reported catch in the Northern California is seen to be decreasing in other high temperature years. In Central California, the years 2014 and beyond see a large increase in the fish catch reported. The years see an increase in mean temperature. The the fish catch is increased in the years 2003,2004 which are peak mean temperature years for Southern California. For the species Lingcod, there is no clear pattern with the mean temperature and the reported catches in a particular year.

Cabezon:

For species Cabezon, it is observed that with the increase in years and change in mean temperature, the reported fish catches has decreased.

Limitations:

For both Lingcod and Cabezon, the relation with mean temperature change is difficult to assess as there could have been many confounding variables that changed the reported catches. Although Cabezon can be seen to have a decrease in reported fish catch with increasing years. It could have been attributed to general increase in temperature per year.

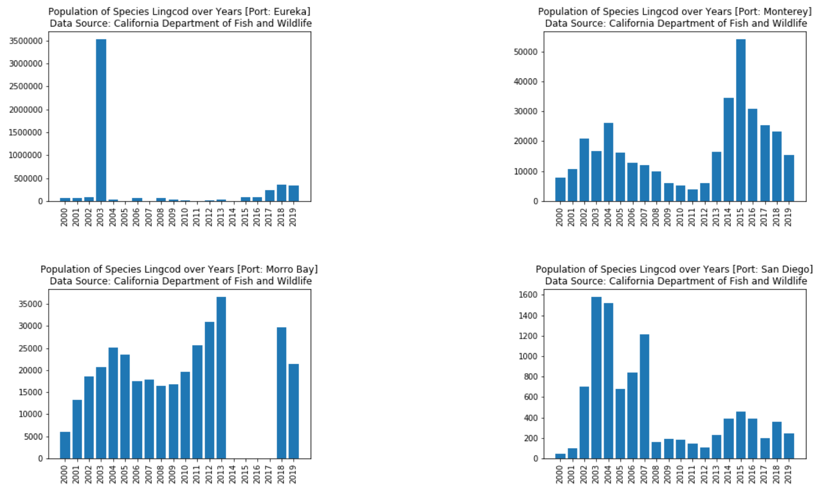

Table 3: Fish species:Lingcod in each year in each port (2000-2019)

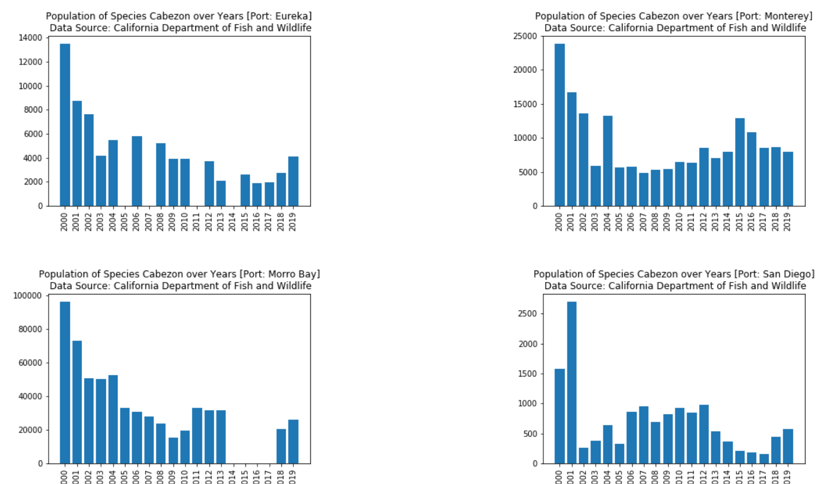

Table 4: Fish Species:Cabezon in each year in each port (2000-2019)

Reported New Species of Fishes Every Year:

In the year 2014, Morro port observed a significant amount of increase reported catch of newer species than previous year. This could have caused due to fish migration in that year as in 2014 a significant high mean temperature is noted than the previous years in all the ports. Except for Morro Bay, there isn't any significant pattern in the reported fishes per year. This could be because with water temperature changes, fishes migrate to a more suitable temperature areas.

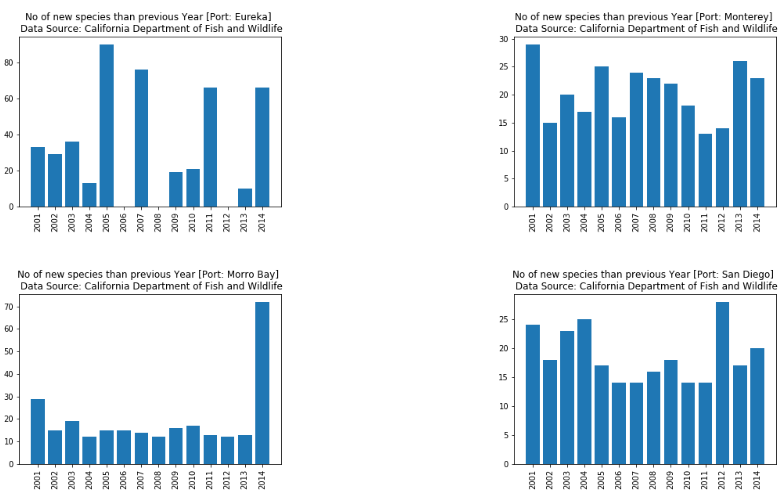

Table 5: Reported New Fish Species every year

3.2 Average water level and climate change in California

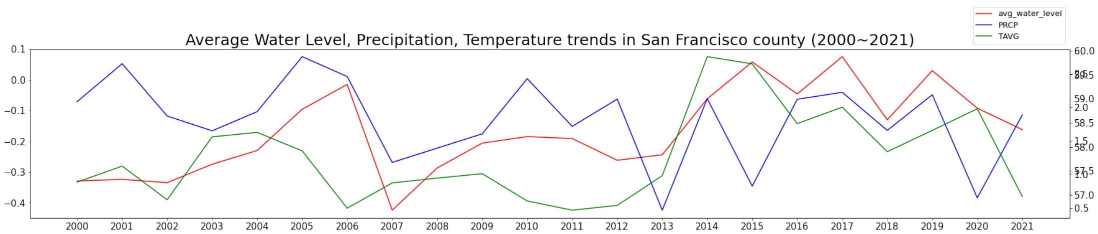

Figure 6: Plot of Avg Water Level/Precipitation/Temperature in San Francisco(2000-2021)

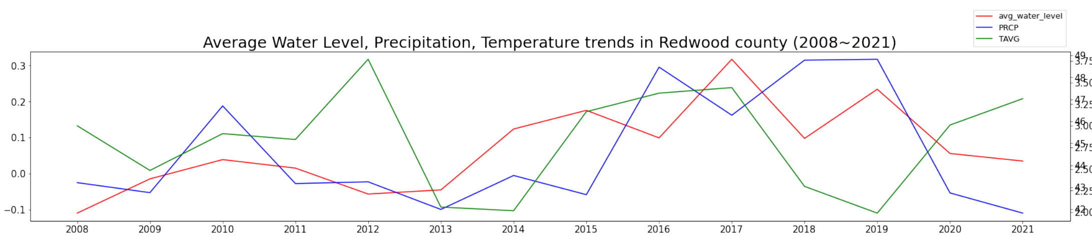

Figure 7: Plot of Avg Water Level/Precipitation/Temperature in Redwood(2000-2021)

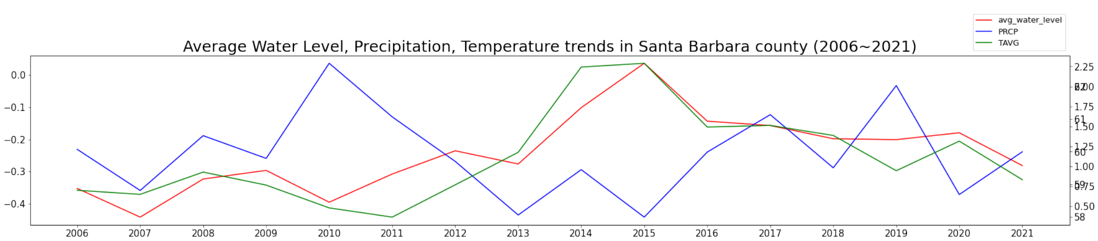

Figure 8: Plot of Avg Water Level/Precipitation/Temperature in Santa Barbara(2000-2021)

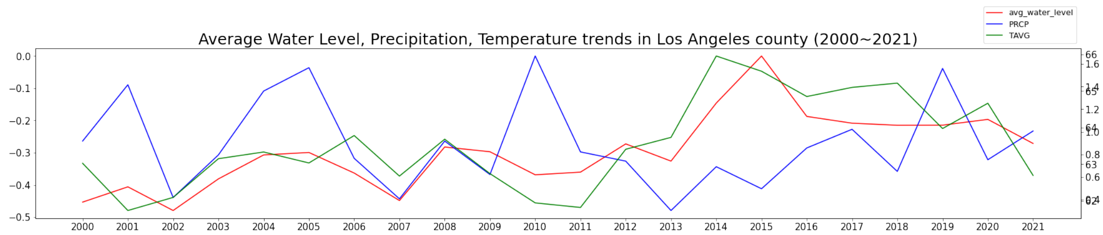

Figure 9: Plot of Avg Water Level/Precipitation/Temperature in Los Angeles(2000-2021)

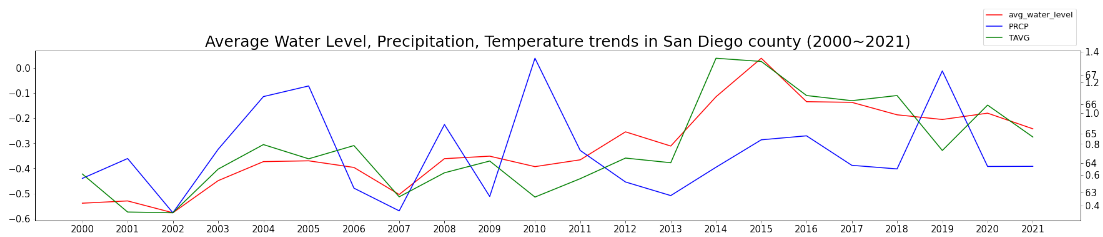

Figure 10: Plot of Avg Water Level/Precipitation/Temperature in San Diego(2000-2021)

From the above figures, we observe that the trend of average water level and temperature are positively correlated in all five counties, and the average precipitation and temperature are inversely correlated except for San Diego county. This finding matches the result of correlation matrix as shown above. In general, there is an upward trend in average water level. In 2001, 2005, 2010, 2019, all the counties' average precipitation reached a peak. In 2002, 2007, 2013, 2020, all the counties' average precipitation suddenly dropped. By contrast, the trend of average temperature varies with each county. More specifically, the temperature pattern in Redwood is exactly the reverse of the temperature patterns in other counties. For example, In 2014, 2015 and 2021, all other counties' average temperature reached the highest point while Redwood county is at its lowest. Overall, there is an alternating pattern of average precipitation and temperature for all counties.

3.3 Drought condition in California

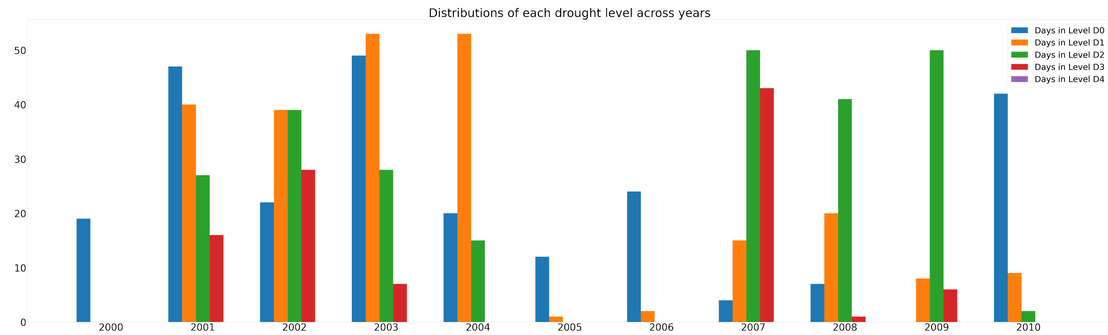

Figure 11: Distribution of Different Drought Levels(2000-2010)

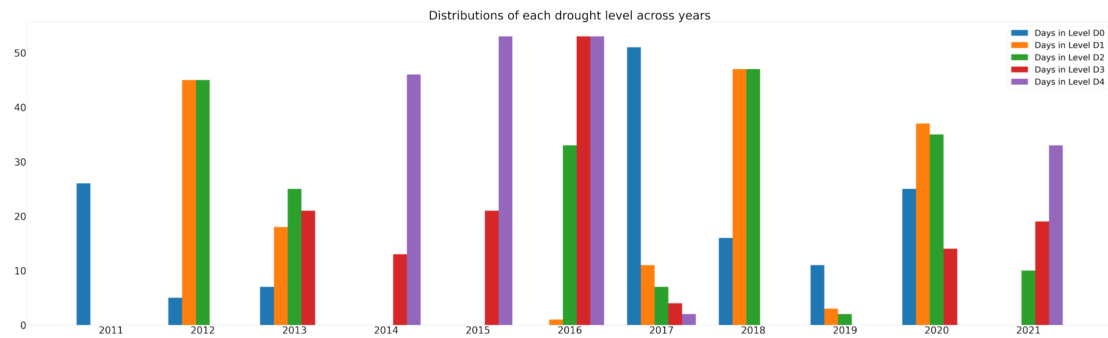

Figure 12: Distribution of Different Drought Levels(2011-2021)

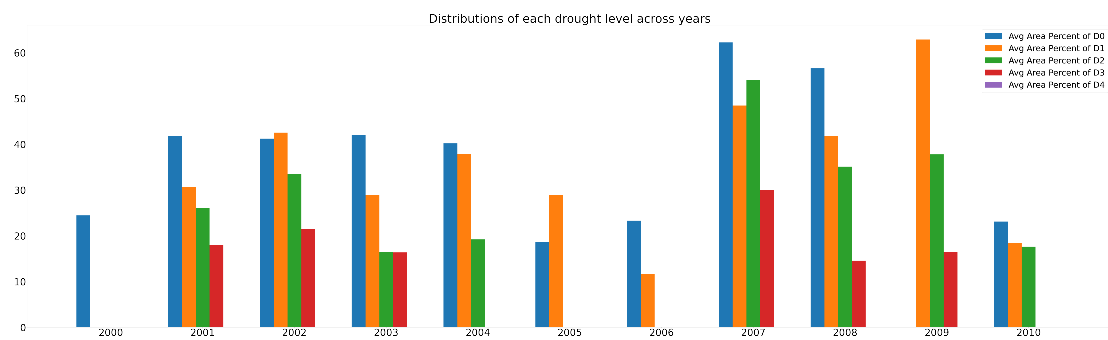

Figure 13: Distribution of Different Drought Levels(2000-2010)

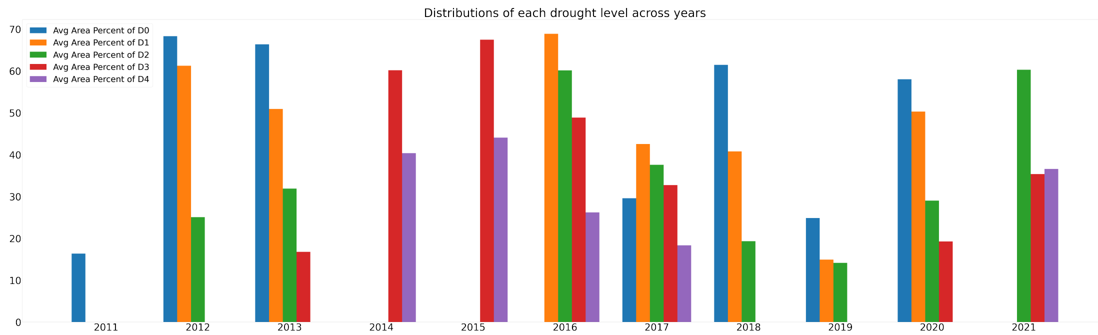

Figure 14: Distribution of Different Drought Levels(2011-2021)

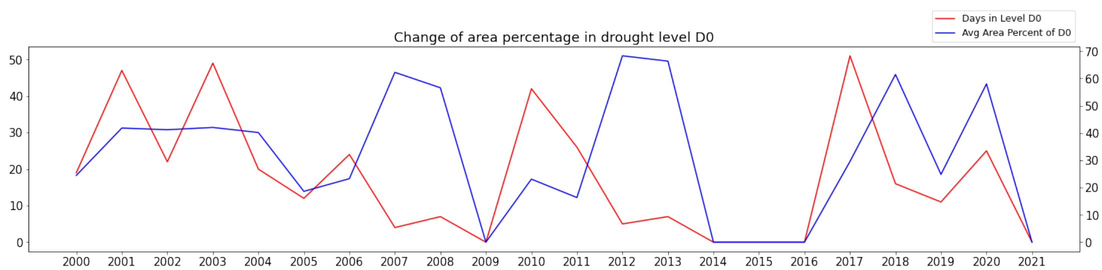

Figure 15: Trends of Drought Level D0(2000-2021)

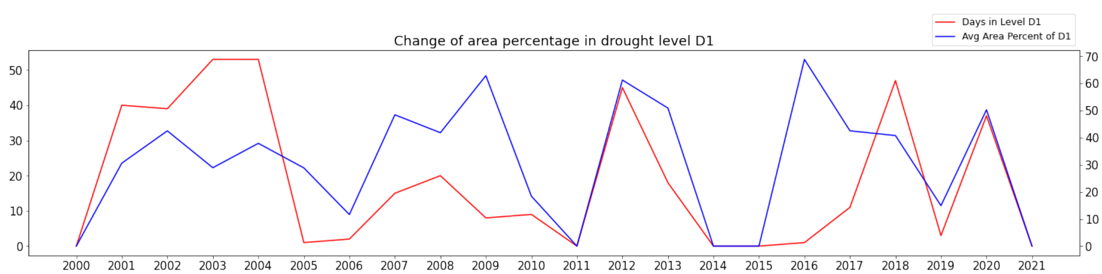

Figure 16: Trends of Drought Level D1(2000-2021)

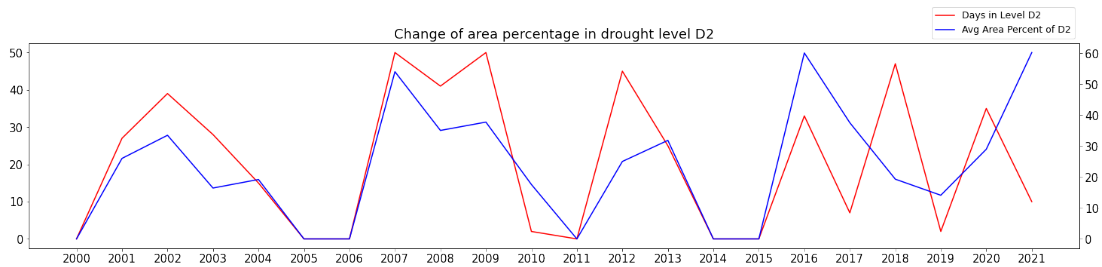

Figure 17: Trends of Drought Level D2(2000-2021)

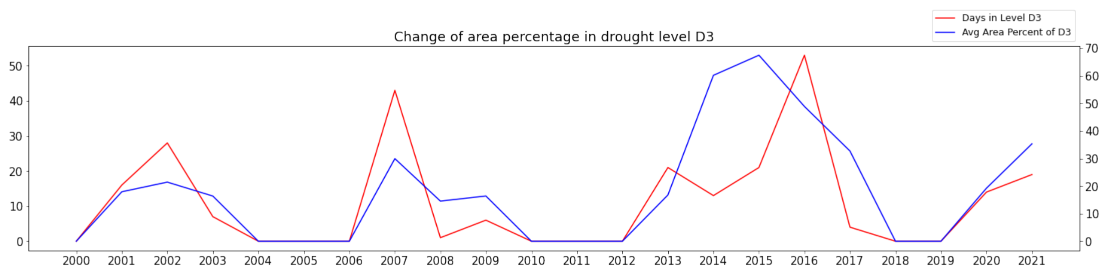

Figure 18: Trends of Drought Level D3(2000-2021)

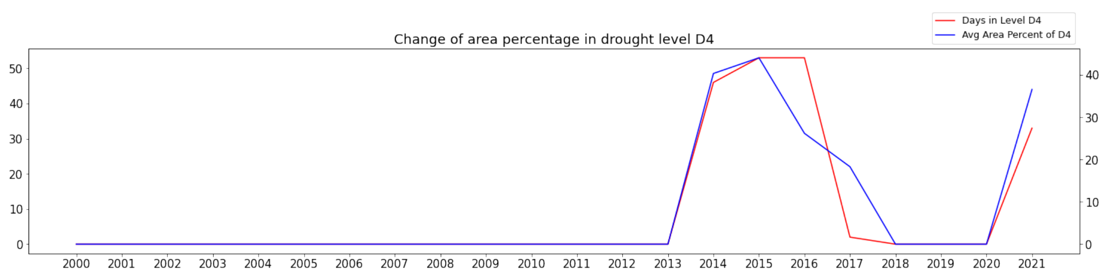

Figure 19: Trends of Drought Level D4(2000-2021)

From the bar plots, we observe that in the period of 2000-2010, drought is less severe where lower level drought takes the most part of it both in terms of the area percentage or number of days. Extreme or exceptional drought(D3D4) are not as common as other types. By contrast, between 2011 and 2021, we can see that the red and purple bars appear more often and go up a lot, which indicates that the drought conditions are gradually worsening in California. Also note that from 2014-2016, there are around 50 days of exceptional drought in California, which signals a disastrous drought in that period of time. Also from the line plots, it's also easy to see an upward trend of D3 and D4 levels of drought, especially from 2014 to 2016. The D0 and D1 levels of drought were becoming less frequent in recent years.

After sorting the values of NonConsecutiveWeeks statistics, we conclude that Imperial County, Riverside County, San Bernardino County have the longest duration of drought in California. San Francisco County, Humboldt County, Del Norte County have the shortest duration of drought in California. Interestingly, the former three are all in the Southern part of California, and the latter three are all in the Northern part of California. Whether a geographical difference could impact the drought condition could be something to look further into.

3.4 Ocean Acidification in California

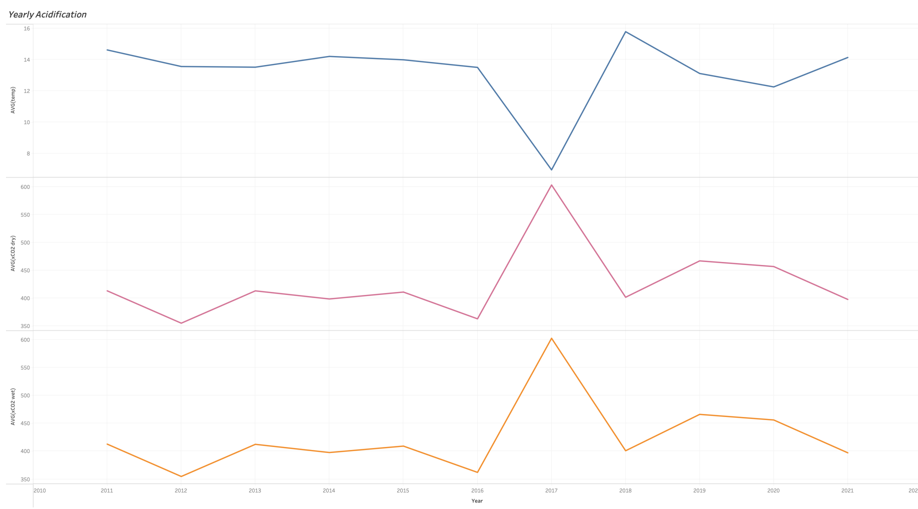

Figure 20: Yearly Acidification

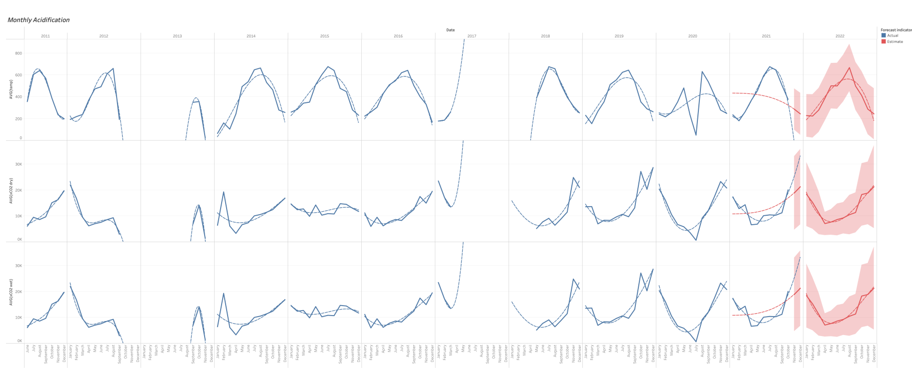

Figure 21: Monthly Acidification

According to the visualization by year, we could observe that data from 2017 looks weird. There is relatively more minor data, and we used average value, so data from 2017 has a pretty different value. Except for 2017, all variables have a more considerable value from 2018 to 2021 than from 2011 to 2015. It is pretty weird since gas solubility is typically low if the temperature is high. Therefore, we can find that as time goes by, the temperature and gas solubility of carbon dioxide get larger. It means that global warming might produce much carbon dioxide, and then the gas solubility of carbon dioxide is getting larger even though the water temperature is also more significant.

According to the visualization by month, we could observe more specifically. We used Tableau to make the line of trend and prediction graph for 2022. This plot shows that it had a yearly cycle. From January to August, the water temperature rises, and the value of the `xCO2` is getting less. The water temperature decreases from August to December, but the value of the `xCO2` increases. It showed that water temperature is one of the reasons for acidification. We could not find the visualization by year. The trend line mostly looks similar except for 2013 and 2017 since both years have fewer datasets. Sometimes it also has irregularities in 2020 that the water temperature is lowest in August and the value of the `xCO2` is highest in August. This cycle and trend are also shown in the prediction.

3.5 Analysis of Salinity of Ocean Water Other Ocean Water Variables

The primary focus of this analysis is to observe whether there are patterns in the yearlong range provided. We would like to explore the idea that precipitation has the potential to affect a location's salinity. In theory, the precipitation that enters the body of water would temporarily dilute the water and decrease the salinity of the water.

Furthermore, we will also spend some time observing the interrelational correlations between the various ocean water variables, especially focusing on salinity.

In Figure 22 shown below, we can see the 5 normalized variables of CO2 density, water temperature, barometric pressure, water salinity, and fugacity of CO2 in air for the year of 2000.

We can also take a look at salinity plotted against the 4 other variables individually to reduce the clutter for more clarity in Figure 23.

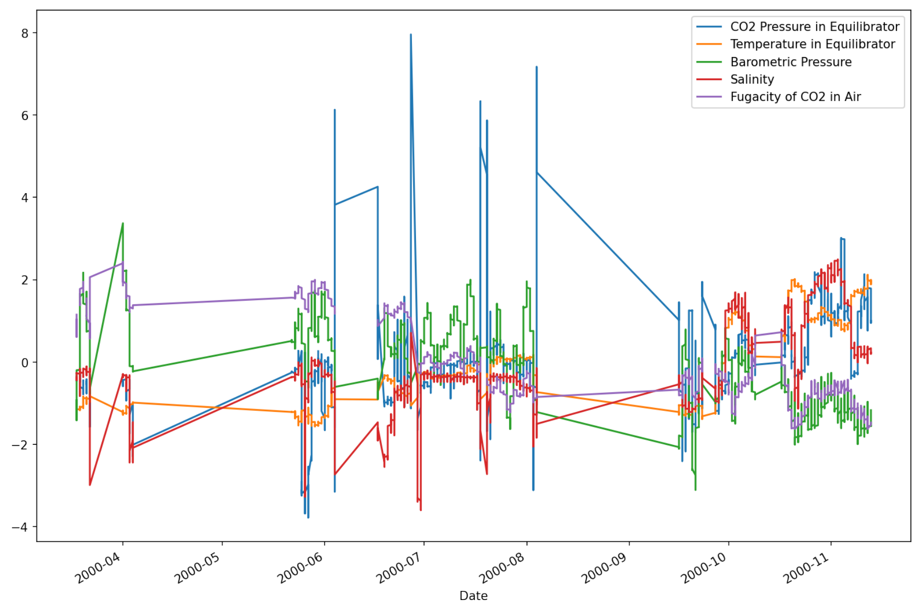

Figure 22: Sea Water Variables Plot against one another

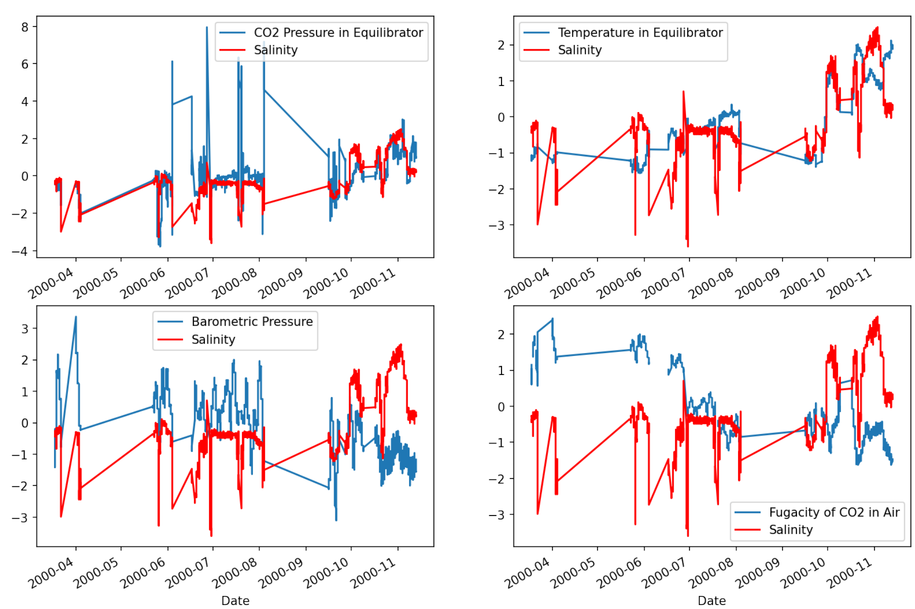

Figure 23: Sea Water Variables against Salinity

Some observations:

There seems to be an inverse relationship between salinity and CO2 pressure. For a period in early to mid-June 2000, when there was a spike in CO2 pressure, salinity then decreased. A similar pattern occurred in August to October 2000, where a increase in salinity can be observed alongside a decrease in CO2 pressure. Similar occurrences can also be seen on smaller scales across the graph. This phenomenon may have to do with the effect that high pressure has on the evaporation and accumulation process in the water cycle. With a higher water content being evaporated from the body of water entering the atmosphere, salinity increases in the water left behind.

Let's also take a look at the correlation matrix of the 5 variables plotted to confirm our findings in the above plots.

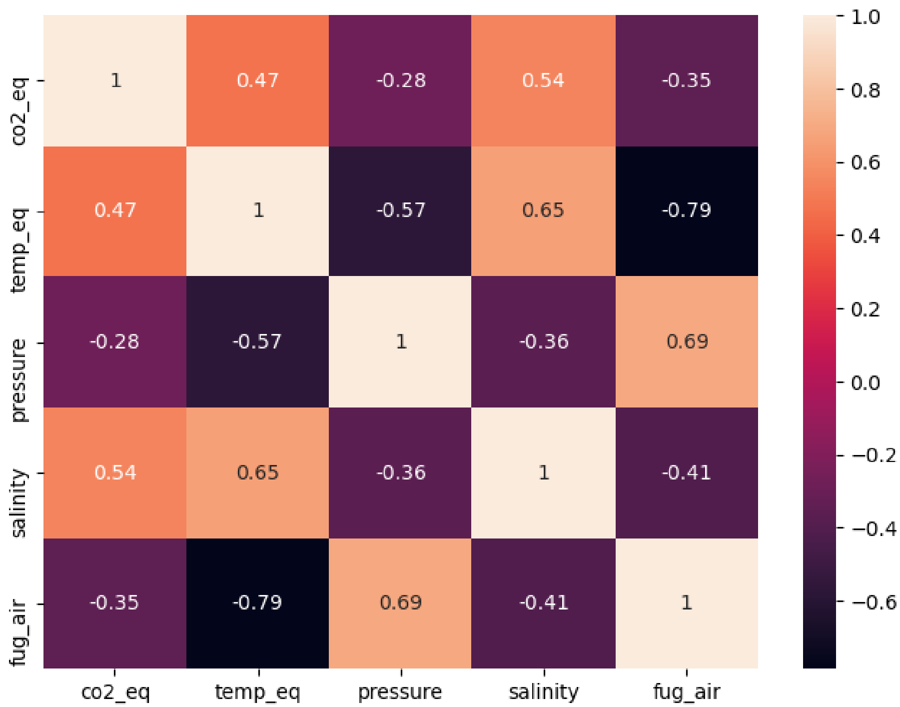

Figure 24: Sea Water Variables Correlation Matrix

If we keep our focus on the correlations that salinity shares with the other 4 variables, we find that CO2 pressure (`co2_eq`, CO2 pressure measured from the ship's equilibrator) and water temperature (`temp_eq`, water temperature measured from the ship's equilibrator) have the highest absolute correlation with salinity.

To briefly entertain our early speculation that salinity would be affected by precipitation, we have also plotted precipitation and salinity to see possible correlations.

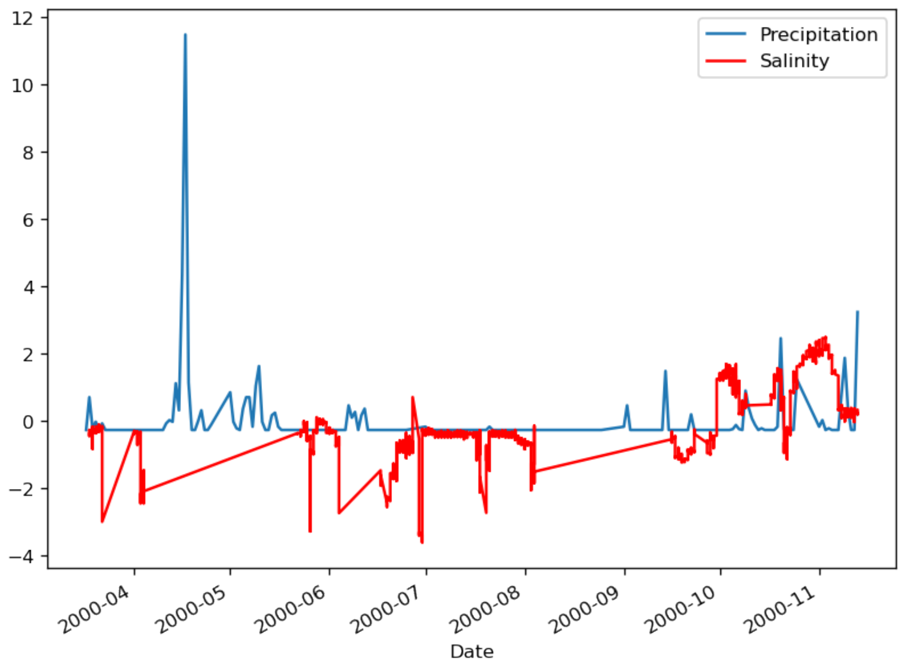

Figure 25: Salinity, Precipitation plot

As we can see from the graph, salinity and precipitation seem to not mutually affect one another as we originally thought they might.

4 Conclusion

From the fishery and water temperature analysis, it can be concluded that increase or decrease in the water temperature does create some noticeable patterns in the reported fish catches. The most noticeable relation between water temperature and the reported fish catch data were seen among the clusters. Similar mean water temperature trend years also were seen as clusters in the fishery data. However, many confounding variables could have been present attributing to those patterns.Fishery and water temperature analysis had many limitations. The number of recorded species features was large(280 species) and not all species were present in all the years. Also, the water temperature data had some missing values for certain regions. Interpretability was limited for some analysis like cluster analysis or observing the species reported catch in each port per year.

From the analysis of the average water level and climate change in California, we conclude that average water level in California fluctuated periodically but generally went up to some degree during the past twenty years. In the selected five counties, a relatively strong positive correlation is present between average water level and temperature in all five counties. However, there is no sufficient evidence indicating a casual relationship between the two variables. On the other hand, a weaker negative correlation between the average water level and precipitation is observed in all other counties except for San Diego. In time periods 2001-2002, 2005-2007, 2010-2013, 2019-2020, the average precipitation first reached the peak but quickly dropped. In addition, the climate and average water level in Redwood county shows an entirely different pattern from other counties, but due to the limited data sources, we haven't been able to find the reason why.

From the analysis of drought condition in California, it's easy to see that mild drought occurred more frequently during the 2000-2010 time period. D3 or D4 level drought were relatively uncommon and had a limited effect in terms of area percentage covered. By contrast, when we get to the 2011-2021 time period, the droughts are getting worse in California. D3 and D4 level droughts were taking a greater portion of the California regions and lasted significantly longer than before. Especially during 2014-2016 and 2021, California had experienced a period of exceptionally severe drought. In more recent years, D0 and D1 level droughts showed a downward trend while D2, D3 and D4 level droughts all showed an upward trend. This can be a dangerous sign. Further more, we discovered that Southern parts of California(Imperial County, Riverside County, San Bernardino County) suffered the most from drought while Northern parts of California(San Francisco County, Humboldt County, Del Norte County) suffered the least from drought. However, we still need to conduct more research to study the impact of geographical location on a region's drought severity.

From the ocean acidification analysis in California, we can conclude that water temperature might not affect ocean acidification a lot. When we see the yearly acidification plot (Figure 20), except in 2017, both water temperature and xCO2 gradually increase. If water temperature was the main reason for ocean acidification, xCO2 should decrease since gas solubility increases when the temperature is low. Therefore, it can be concluded that increasing carbon dioxide is not because of water temperature but because of the increasing amount of carbon acid in the air, such as global warming. In addition, we can find some patterns on the monthly acidification plot. As temperature increases in summer and decreases in winter, water temperature also has the same pattern. Because of this pattern, the value of xCO2 decreases in summer and increases in winter. Therefore, we can find that water temperature has a concave trend curve, but xCO2 has a convex trend curve except in 2013 and 2017 (there are just a few datasets). We can also see the plot of prediction in 2022, which is similar to the trend curve. There are some missing data, especially in 2013 and 2017. Since we used mean value in this project, missing data could affect a peculiar trends curve and show distorted prediction.

In the final analysis topic, rather than the result that we expected for precipitation and salinity, we are able to observe a correlation between air pressure (specifically CO2 pressure) and precipitation. We found that salinity had a correlation coefficient of 0.65 and 0.54 with water temperature and air pressure, respectively. We speculate that this is in line with the water cycle processes, as higher temperatures and higher air pressures lead to increased levels of evaporation, thus decreasing water volume while salt levels remain constant, effectively increasing salinity levels.
Of course, the greatest limitation of this analysis was the short and inconsistent coverage. Much more could have been achieved with this dataset had it been more extensive and covered a date range longer than what we were given. It would be wise to take the conclusions from this particular analysis with a bit less weight, because its short range does not give it as much credibility to draw definite conclusions from.

References

[1] CA Department of Fish and Wildlife. *Statistical Areas, Species Definition, Condition and Value*. 2019. URL: https://nrm.dfg.ca.gov/FileHandler.ashx?DocumentID=178007&inline.

[2] *What is the USDM*. URL: https://droughtmonitor.unl.edu/About/WhatistheUSDM.aspx.

[3] PFEL Cilmate Marine Fisheries. *Temporary Changes in the Distribution of Warm Water Fish During El Niño Ocean Warming Events*. URL: https://www.pfeg.noaa.gov/research/climatemarine/cmffish/cmffishery2.html.