In [1]:
```python
# libraries
import pandas as pd
```

Analysis of the Southern California Coastline's Ocean Data

Tingting Liang, Mac Beggs, Apollo Hodges

Author contributions

Mac worked on exploring the differences in the swell data. Apollo worked on finding correlation between variables. Tingting worked on the summarizing the results.

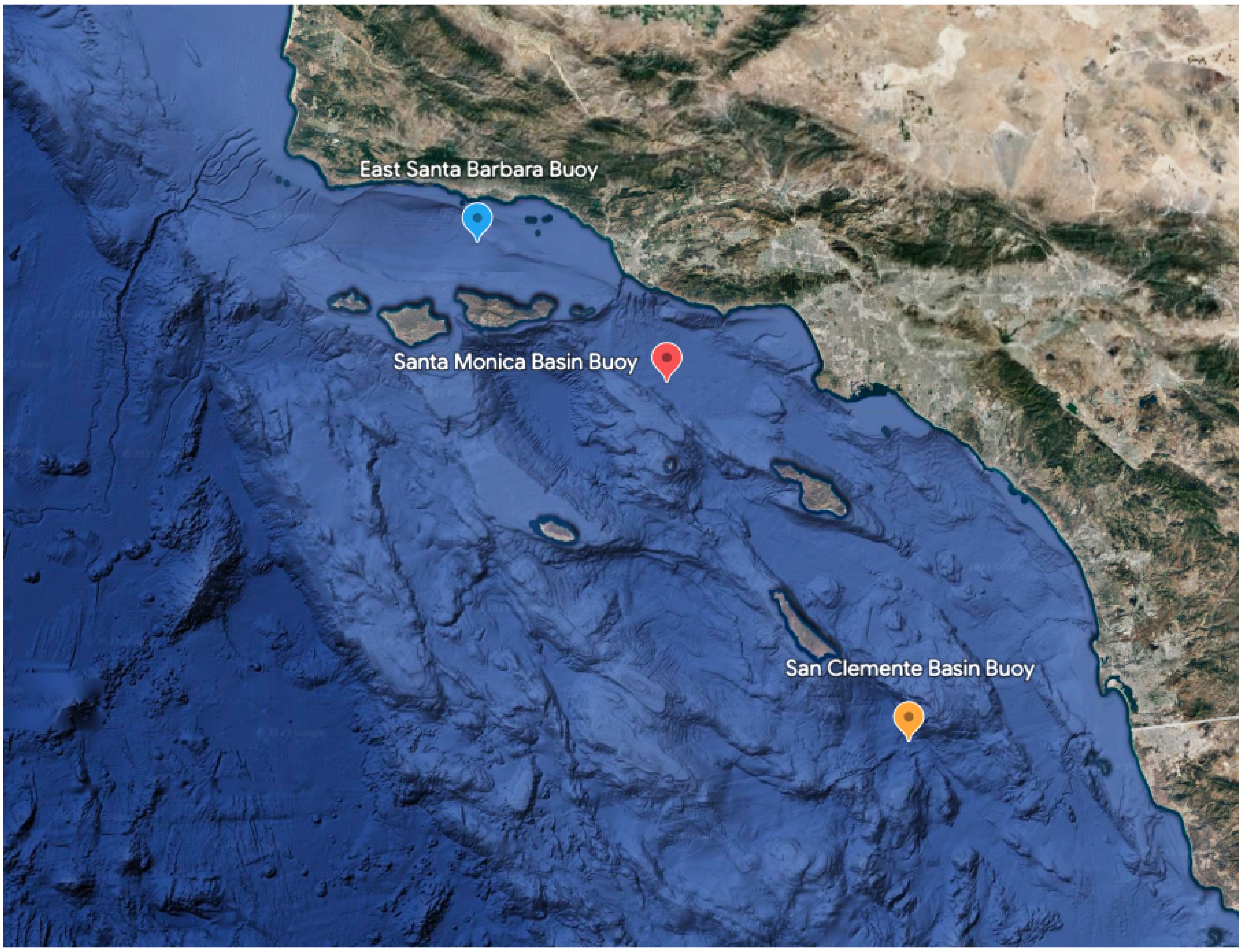
Abstract

For this project, we are exploring the 2022 ocean data from three locations (East Santa Barbara, Santa Monica Basin, San Clemente Basin) along the Southern California Coastline. Since there are a chain of Channel Islands near the coast of Southern California, we want to anaylze how the Channel Islands will affect how the basins are receiving wind and wave. The aims of this project are to find the differences of ocean wind and surface wave data among East Santa Barbara, Santa Monica basin and San Clemente basin particularly during certain seasons and time. We found that these locations receive swells differently in different seasons. During the fall and winter, the most energetic swells of Santa Monica Basin and San Clemente Basin came primarily from the west and in Spring and Summer the most energetic swells came primarily from the south southwest, whereas the East Santa Barbara Buoy only receives swells from the west all year long. We also found some interesting correlation between wind speed and averge wave period, although higher wind speeds generally means lower average wave periods, extremely high wind speed that goes above 10 m/s would instead increase the average wave period causing longer swells.

Introduction

Background

Living less than a mile away from the great Pacific Ocean, we were interested in how the water and waves differ along the coast of California. We will use the National Oceanic and Atmospheric Administration (NOAA)'s Standard meterological data taken from the National Data Buoy Centers at three locations, East Santa Barbara, Santa Monica basin and San Clemente basin. The dataset contains basic information about surface waves and ocean wind. We will explore the difference in the waves and wind from these three locations along the coast and analyze what may cause those differences. The locations are marked on the map below:



Aims

The aims of this project is the find the differences of ocean wind and surface wave data among East Santa Barbara, Santa Monica basin and San Clemente basin. In particular, we would like to explore any differences in the locations during certain seasons or times. Furthermore, we would like to find out if these locations receive swells and/or winds from the different directions. We are interested to learn if there is any correlation between wind direction and wind speed. We are also curious to find any correlation between swell height or swell period and swell direction.

Because waves are created by strong winds blowing over large areas in the ocean, we predict that wind speed will be a strong factor in predicting wave height. We also predict that the San Clemente Buoy will yeild strong winds and stronger swells than the East Santa Barbara and Santa Monica Basin buoys, because it is more exposed to the ocean's storms, whereas the Santa Barbara and Santa Monica buoys are more protected by the Channel Islands. We also predict that the East Santa Barbara buoy will not capture any swells from the south, because Santa Cruz Island, San Miguel Island, and Santa Rosa Island are directly beneath the buoy, which would block any swell coming up from the south. We also predict that the Santa Monica Basin buoy will not capture any swells from the west, because the same chain of islands that are blocking the Santa Barbara buoy from the south are directly west of the buoy.

Materials and methods

Datasets

The data is collected by the National Ocean and Atmospheric Administration (NOAA)'s National Data Buoy Center. The Santa Monica Basin data is from Station 46025 located at 33.755 N 119.045 W (https://www.ndbc.noaa.gov/station_history.php?station=46025). The East Santa Barbara data is from Station 46053 located at 34.241 N 119.839 W (https://www.ndbc.noaa.gov/station_page.php?station=46053). The San Clemente Basin data is from Station 46086 located at 32.499 N 118.052 W (https://www.ndbc.noaa.gov/station_history.php?station=46086). We merged the three datasets into one and tidied it into a dataset named `df`.

The data was taken from ocean buoys that averages and reports its observations over various time frames; the majority of the data is reported hourly or every 20 minutes.

- The **observational unit** is ***Location***.
- The **variables** are ***#YY, MM, DD, hh, mm, WDIR, GST, WVHT, DPD, APD, MWD, PRES, ATMP, WTMP***.

The table below provides variable descriptions and units for each column in the dataframe.

Name	Variable description	Type	Units of measurement
Location	Location the data is collected (East Santa Barbara or Santa Monica Basin)	Categorical	None
#YY	Year	Numeric	Year
MM	Month	Numeric	Month
DD	Day	Numeric	Day
hh	Hour	Numeric	hour
mm	Minute	Numeric	minute
WDIR	Wind direction (the direction the wind is coming from in degrees clockwise from true N) during the same period used for WSPD.	Numeric	degT
WSPD	Wind speed averaged over an eight-minute period for buoys and a two-minute period for land stations.	Numeric	m/s
GST	Peak 5 or 8 second gust speed measured during the eight-minute or two-minute period.	Numeric	m/s
WVHT	Significant wave height is calculated as the average of the highest one-third of all of the wave heights during the 20-minute sampling period.	Numeric	meters
DPD	Dominant wave period is the period with the maximum wave energy.	Numeric	sec
APD	Average wave period of all waves during the 20-minute period.	Numeric	sec
MWD	The direction from which the waves at the dominant period (DPD) are coming. The units are degrees from true North, increasing clockwise, with North as zero degrees and East as 90 degrees.	Numeric	degT
PRES	Sea level pressure	Numeric	hPa
ATMP	Air temperature	Numeric	degC
WTMP	Sea surface temperature	Numeric	degC

The first few rows of the data are shown below:

In [2]:
```
df = pd.read_csv('tidy-data.csv')
df.head()
```

Out[2]:

	Location	#YY	MM	DD	hh	mm	WDIR	WSPD	GST	WVHT	DPD	APD	MWD	PRES	ATMP	WTMP	WDIR_binned	MWD_binned
0	East Santa Barbara	2022.0	1.0	1.0	0.0	40.0	269.0	10.2	12.7	1.91	7.14	5.12	266.0	1007.4	12.9	13.3	W	W
1	East Santa Barbara	2022.0	1.0	1.0	1.0	40.0	292.0	14.8	18.4	2.07	6.67	5.19	249.0	1006.7	12.7	13.3	WNW	WSW
2	East Santa Barbara	2022.0	1.0	1.0	2.0	40.0	274.0	11.1	14.0	2.02	7.14	5.37	261.0	1008.4	12.4	13.3	W	W
3	East Santa Barbara	2022.0	1.0	1.0	3.0	40.0	297.0	11.4	14.6	2.24	6.25	5.45	265.0	1009.1	12.4	13.3	WNW	W
4	East Santa Barbara	2022.0	1.0	1.0	4.0	40.0	291.0	10.1	14.4	2.13	9.09	5.58	254.0	1009.4	12.3	13.3	WNW	WSW

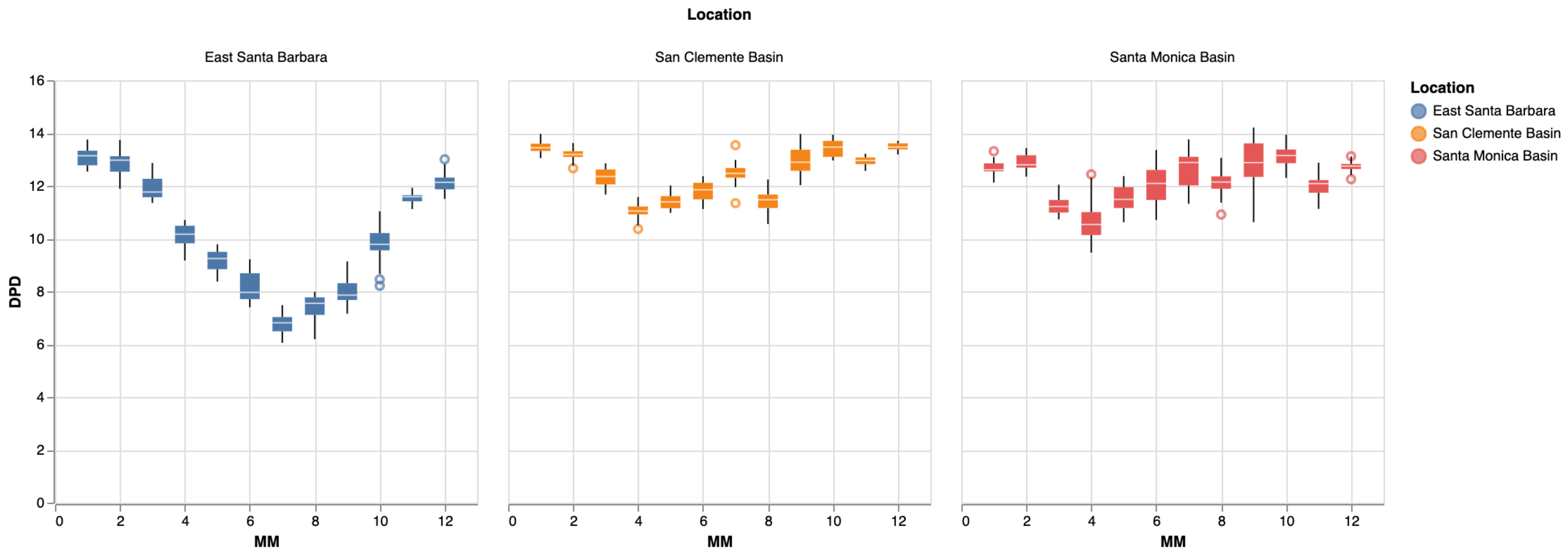
Methods

To find out how swells and winds differ among the three locations, we first grouped the data by months. Then for each month we took the mean of all variables at each hour. We plotted the hour's mean of each variable over 24 hours, giving us what an average day at each buoy looks like in a particular month. We plotted month against dominant wave period by location to see how different locations are receiving swells in different months. We plotted month against dominate wave direction by location to see from which direction the buoys are receiving the swells in different months. We plotted a heatmap to show the correlation between variables. We also plotted a scatterplot of wind speed against average wave period to see how wind speed is correlated to average wave period.

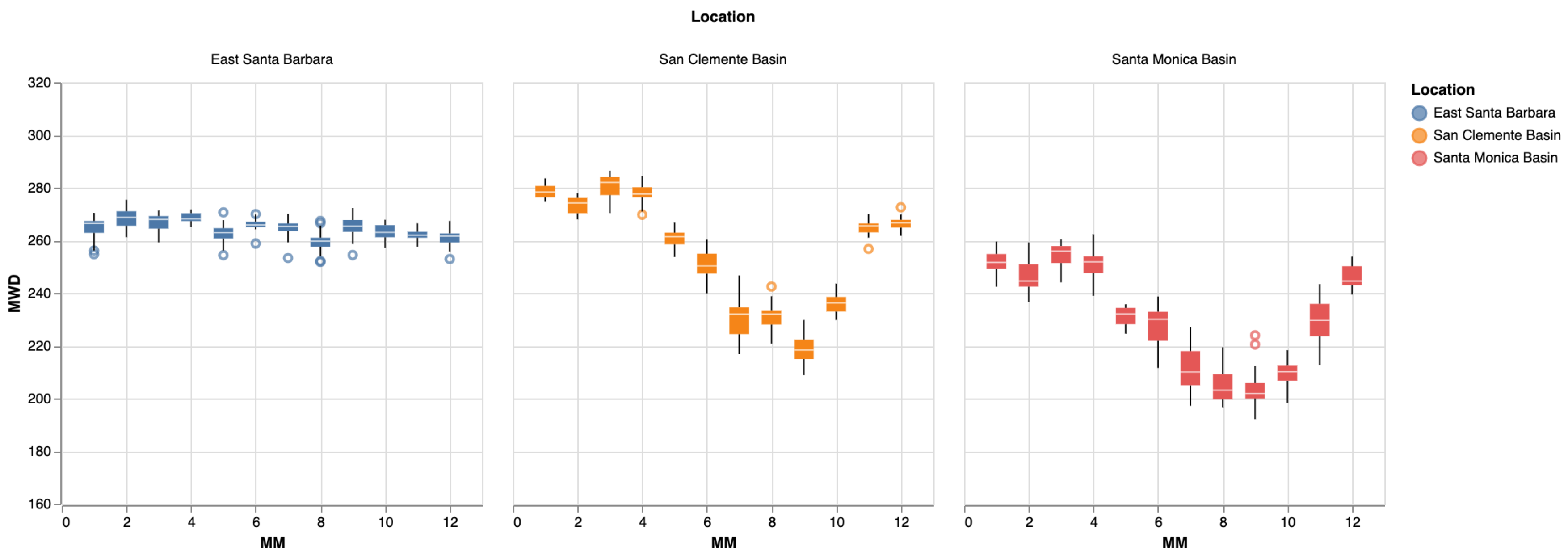
Results

Dominant Swell Periods (top) & Directions (bottom)

The plot below shows dominate swell periods in different months in the different locations.



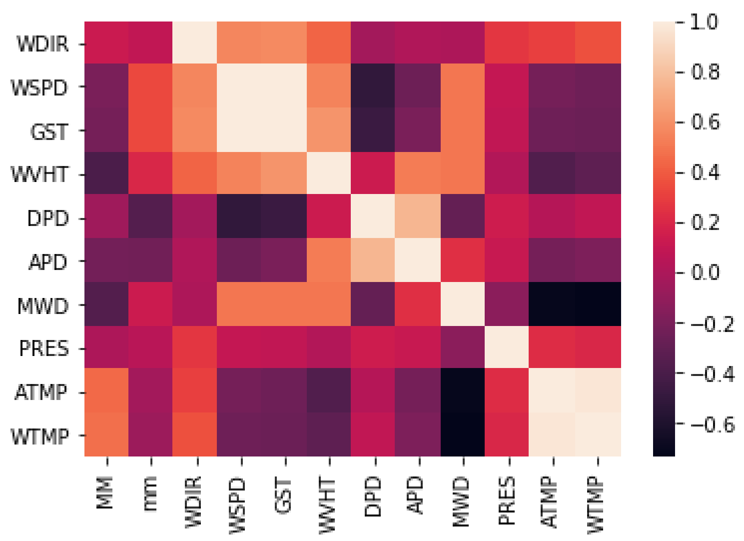
The graph below shows from which direction the buoys are receiving swells in different months. Santa Monica Basin and San Clemente Basin receive swells from south and west in different seasons while East Santa Barbara only receive swells from the west for the entire year.



180 degrees is south, and 270 degrees is west.

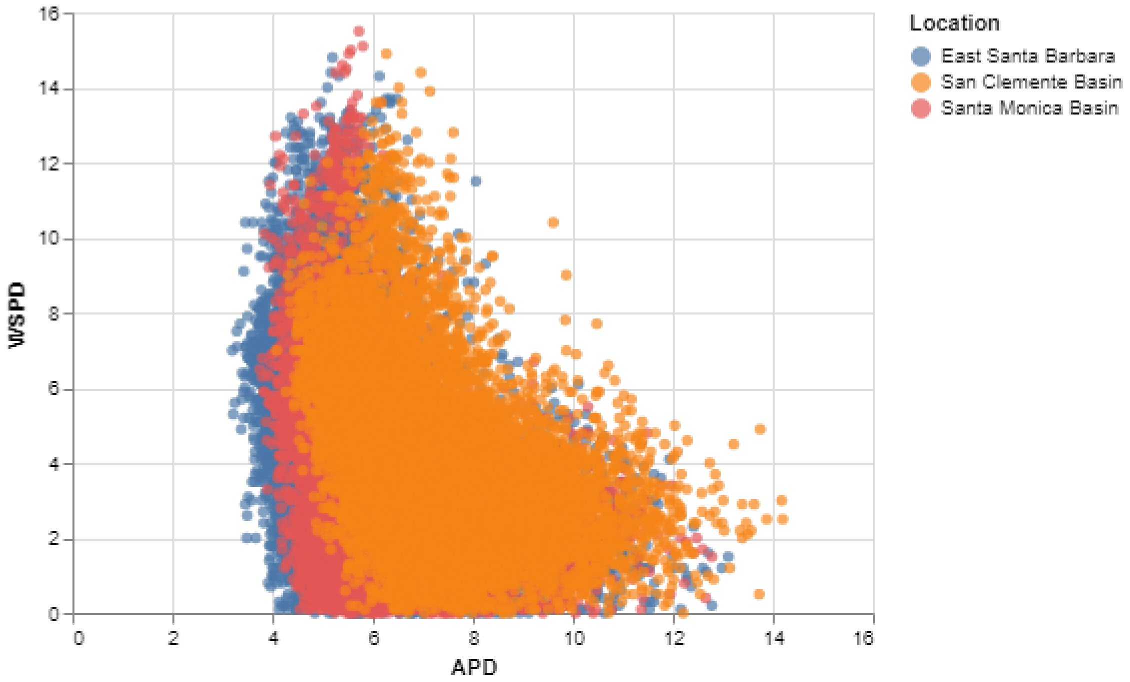
Correlation Plot

The heatmap below shows correlation between variables. Postive correlations in orange and negative correlations in purple. Strong correlations in dark tones, and weak correlations in light tones.



Wind Speed vs Average Wave Period

The plot below shows correlation between wind speed and average wave period. There is a overall negative correlation between wind speed and average wave period. However as wind speed is greater than 10 m/s, there's a weak positive correlation between wind speed and average wave period.



Discussion

In our exploratory data analysis, we were pleased to find out that at the San Clemente Basin and Santa Monica Basin swells came from different directions at different times of the year. During the fall and winter months the most energetic swells came primarily from the west, and in the Spring and Summer months the most energetic swells came primarily from the south southwest. As we expected, this was not the case for the East Santa Barbara buoy. Santa Barbara only received swells from the west the entire year. Therefore, our prediction that the East Santa Barbara buoy would not receive any swells from the south was correct! This brings us to the difference in the swells that the locations received. In the winter all three locations got their most dominant swells from the west. However, in the summer, the San Clemente Basin and Santa Monica Basin received its strongest swells from the south, while the southern hemisphere was in winter. During the summer, East Santa Barbara got its strongest swells from the west. The swell periods Santa Barbara received from the west were those it got from the west during the winter. In conclusion, the seasons do have a profound effect on the ocean waves at these three locations. Though, to our surprise the Santa Monica Basin buoy did receive swells from the west. When comparing the swells the Santa Monica Basin saw from the west to the other buoys' west swells, you can observe that the west swells at the Santa Monica buoy came at a slightly lower angle south. This slightly lower angle may be due to the chain of islands directly west of the buoy we cited in the "Aims". On the other hand, the San Clemente Basin experienced slightly higher angeled west swells than the other two buoys. This also may be due to the buoy's position to the Channel Islands. The buoy is less protected by the Channel Islands. The buoy is also less wedged in the curvature of Southern California's coastline.

We also looked at the correlations between the variables of our dataset. We noticed a particularly interesting trend between the Wind Speed vs Average Wave Period, and decided to dig deeper. Looking at high Wind Speeds (>10 m/s), we noticed that there is a positive correlation between it and Average Wave Period, despite the total correlation being negative. As wind speed decreases, the correlation between it and Average Wave Period does as well. Therefore, we can conclude that while higher Wind Speeds generally mean lower Average Wave Periods, extremely high Wind Speeds rebound and increase the Wave Period. This could be due to choppy or stormy seas distupting longer period swells in the water. The longer period swells seen at lower wind speeds would also support this hypothesis.

Some further steps we could take is taking the buoy data from previous years and use them to compare, or perhaps predict future years if there's a noticable trend.