

# **Blue Ocean Gear**

# Exploratory Data Analysis

University of Chicago - Data Science Clinic

Aurora Peng  
18th Oct, 2022

# Goals of EDA

We had five objectives for this analysis:

1. Identify and drop “bad” data prior to the subsequent analysis.
2. Describe the distribution of buoys and messages across time and by fishery.
3. Segment buoy messages into deployments according to their system status flags, and describe deployments across time and by fishery.
4. Compare buoys’ swing range across fisheries and identify outliers.
5. Examine water temperature by fishery.

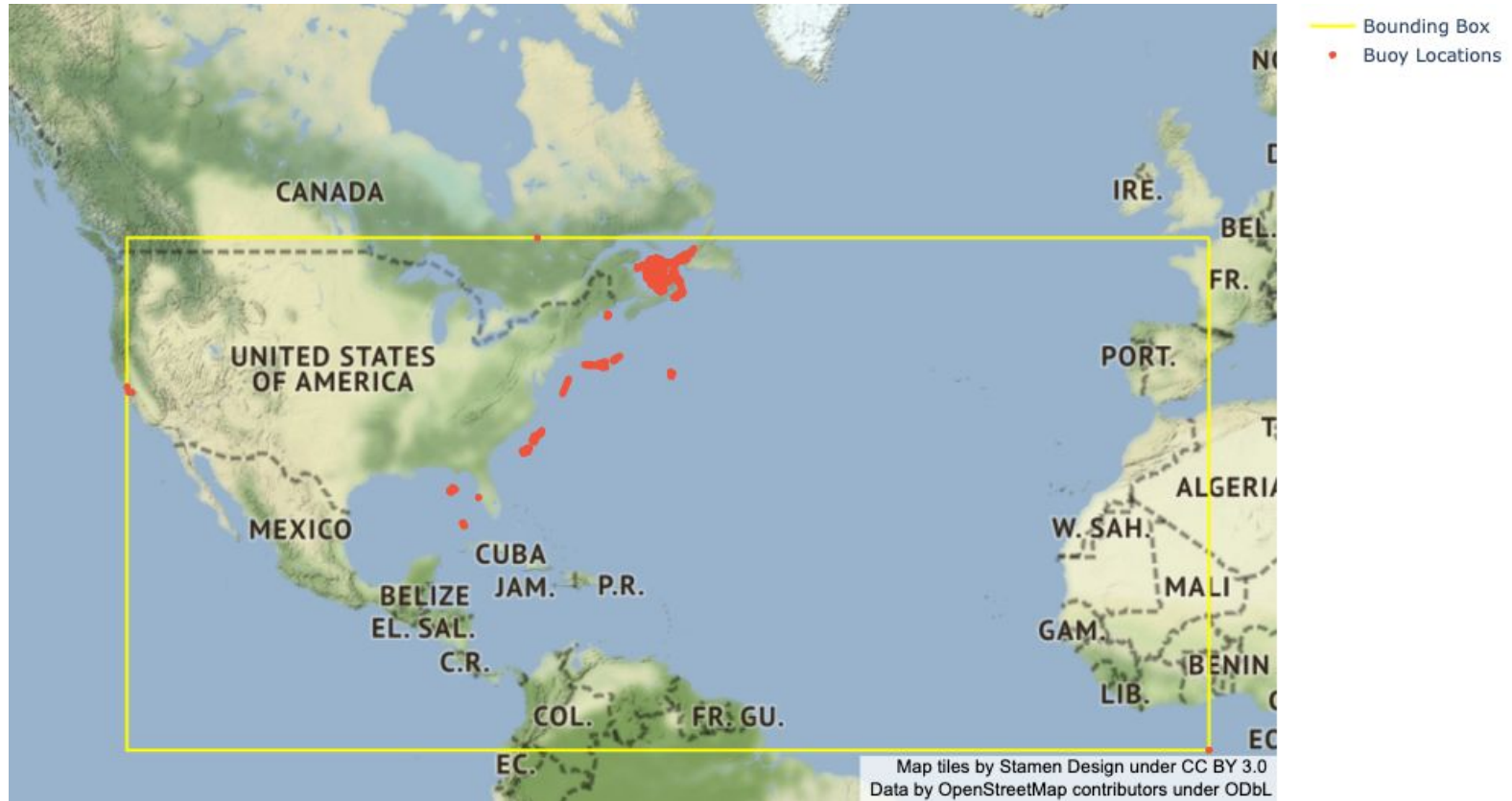
# Drop Bad Data

1. Bad Data - occur outside of expected geographical zone
  - a. Any buoy messages that were identified outside this range were dropped:  $31.885447 \leq \text{latitudes} \leq 49.237$  and  $-77.60128 \leq \text{longitude} \leq -58.327663$
  - b. This is the range for coastal areas near Maine, Massachusetts and New Brunswick
2. Began with 31398 records
3. Dropped 1589 records
4. Remaining analysis done on 29809 records

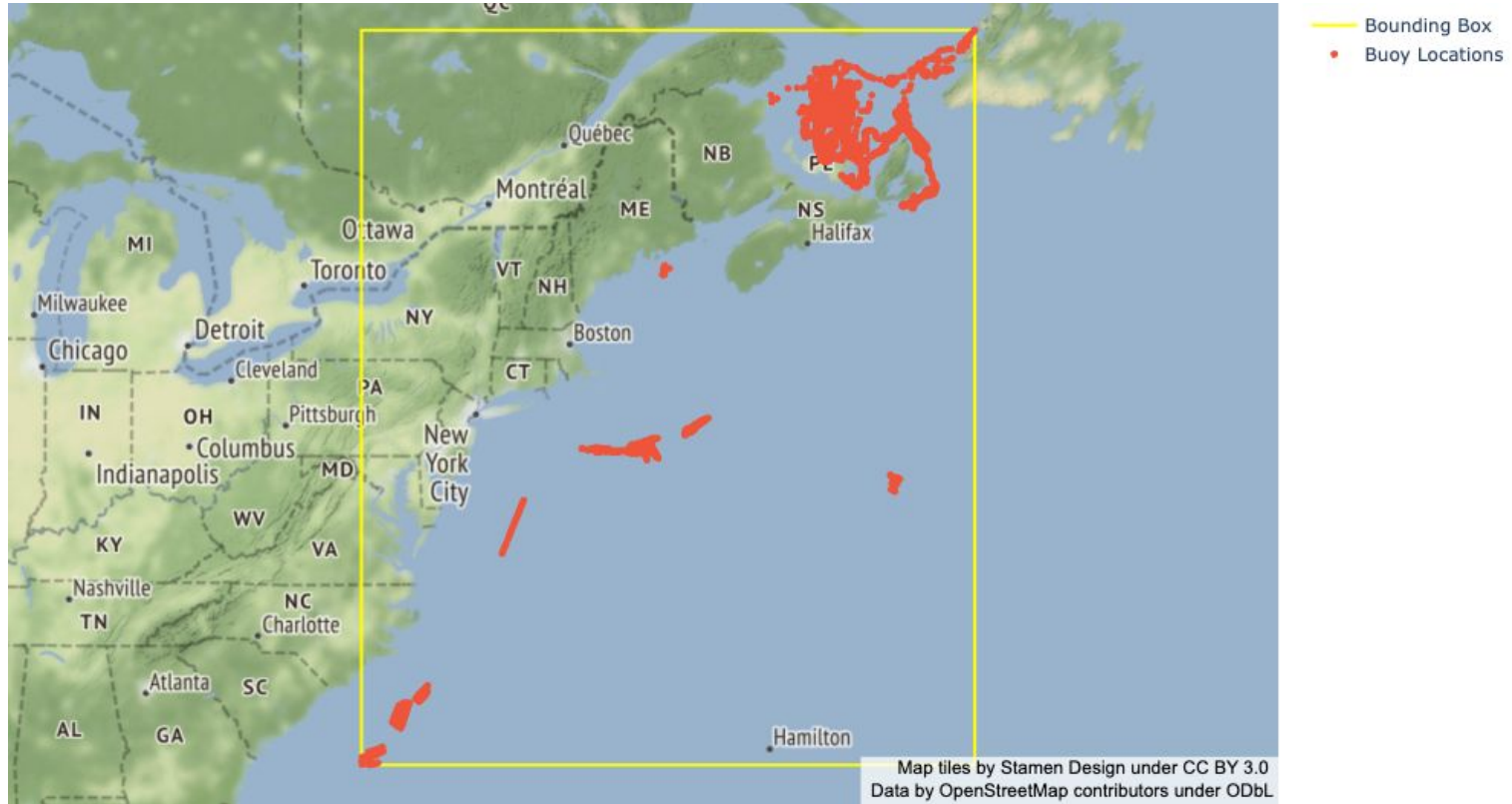
# Missing Values

1. 3 fields had all empty values:
  - a. fast\_update
  - b. last\_updated
  - c. long\_life
2. High correlations emerged among the below fields:
  - a. Cloud\_battery\_soc and battery\_soc = 0.94
  - b. Position\_delta and system\_status = 0.57
  - c. Water\_temperature\_q3 and fishery\_id = -0.59
  - d. Water\_temperature\_q3 and water\_temperature\_mean = 0.7
  - e. Fishery\_id and longitude = 0.81
  - f. Fishery\_id and latitude = 0.79

# Known Geography/Current Patterns - before dropping bad data



# Known Geography/Current Patterns - after dropping bad data



# Buoy Counts

29809 buoy messages

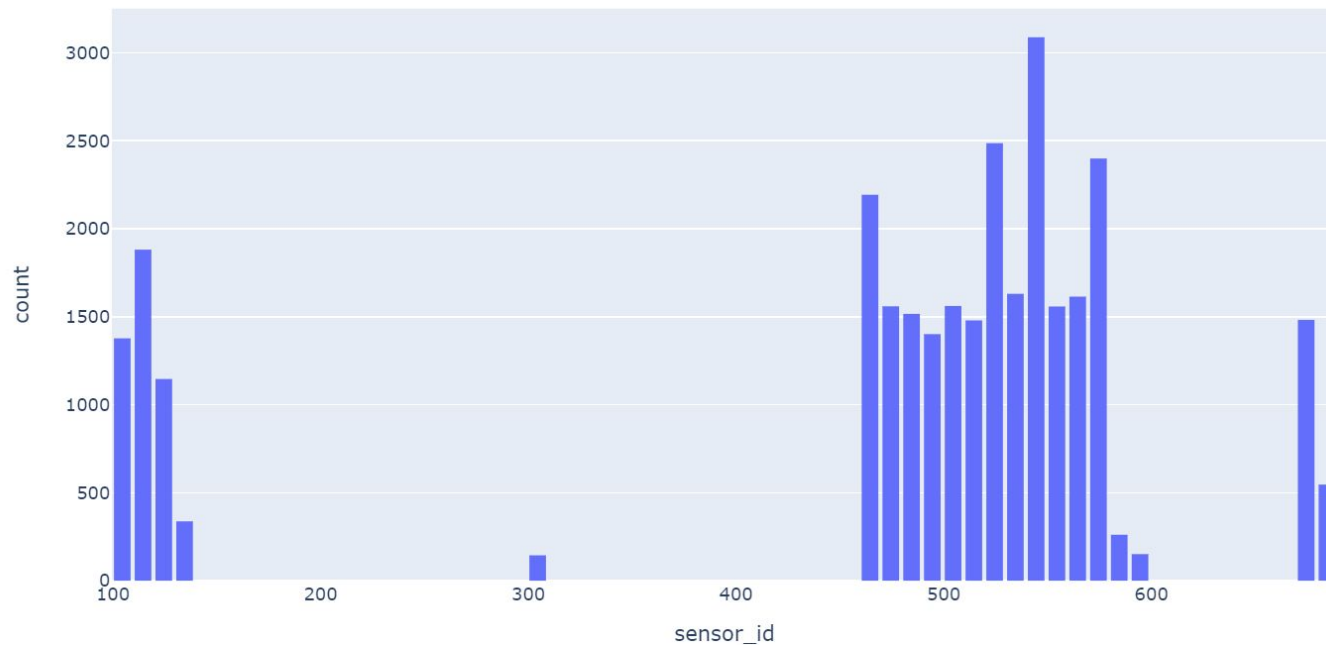
158 unique buoys

## Buoys by Fishery Id

```
{ '1': '587,589,588,590',  
  '2': '680,678,679,674',  
  '3': '103,104,107,109,108,105,114,110,111,113,115,112,116,117,118,119,120,121,122,124,123,125,126,127,128,130,129,131,132,307,  
460,462,461,463,464,466,467,465,468,469,470,471,472,473,474,475,476,478,477,479,480,481,482,483,484,485,488,486,487,489,490,49  
1,492,493,495,494,496,498,497,500,501,499,502,503,504,505,506,507,508,509,511,512,510,514,516,513,515,517,518,519,521,520,522,5  
23,524,526,525,527,528,531,530,529,532,534,533,536,537,535,539,538,540,541,542,543,544,545,547,546,548,549,550,552,551,553,554,  
555,556,557,559,558,560,561,562,565,564,563,567,566,570,568,569,574,573,572,571,576,575,578,577,579' }
```

# Message Counts

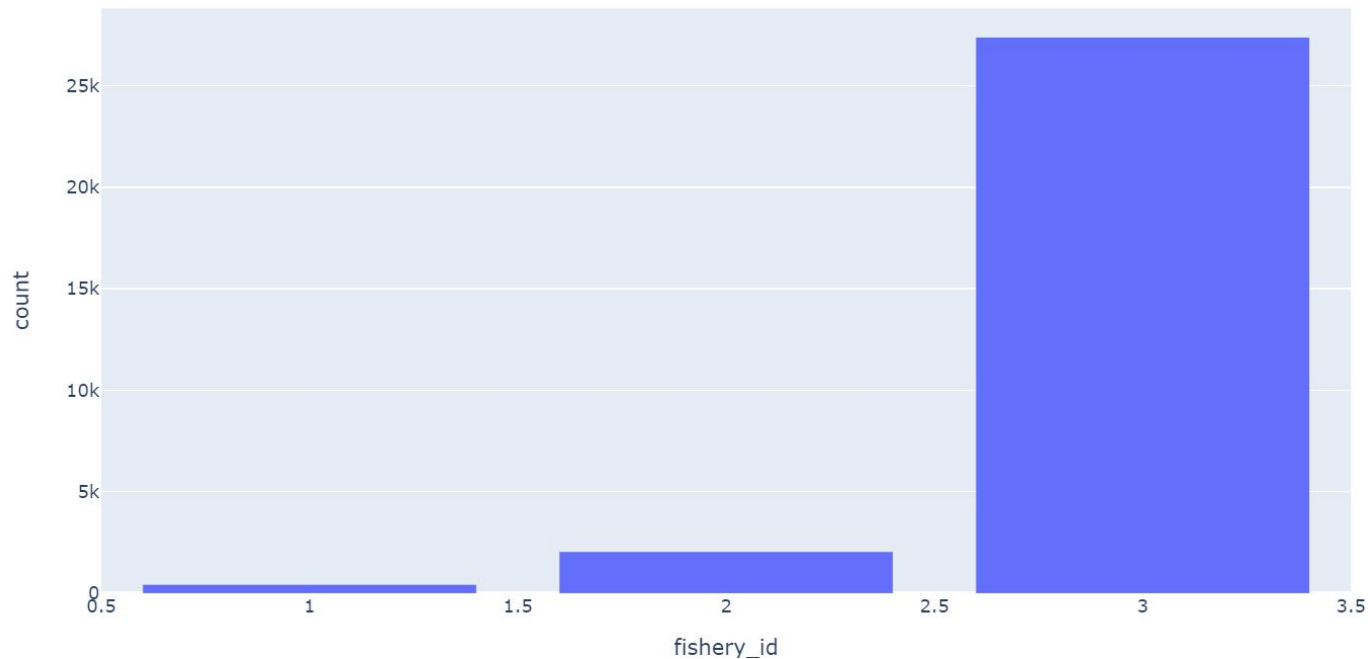
## Messages counts by Buoy





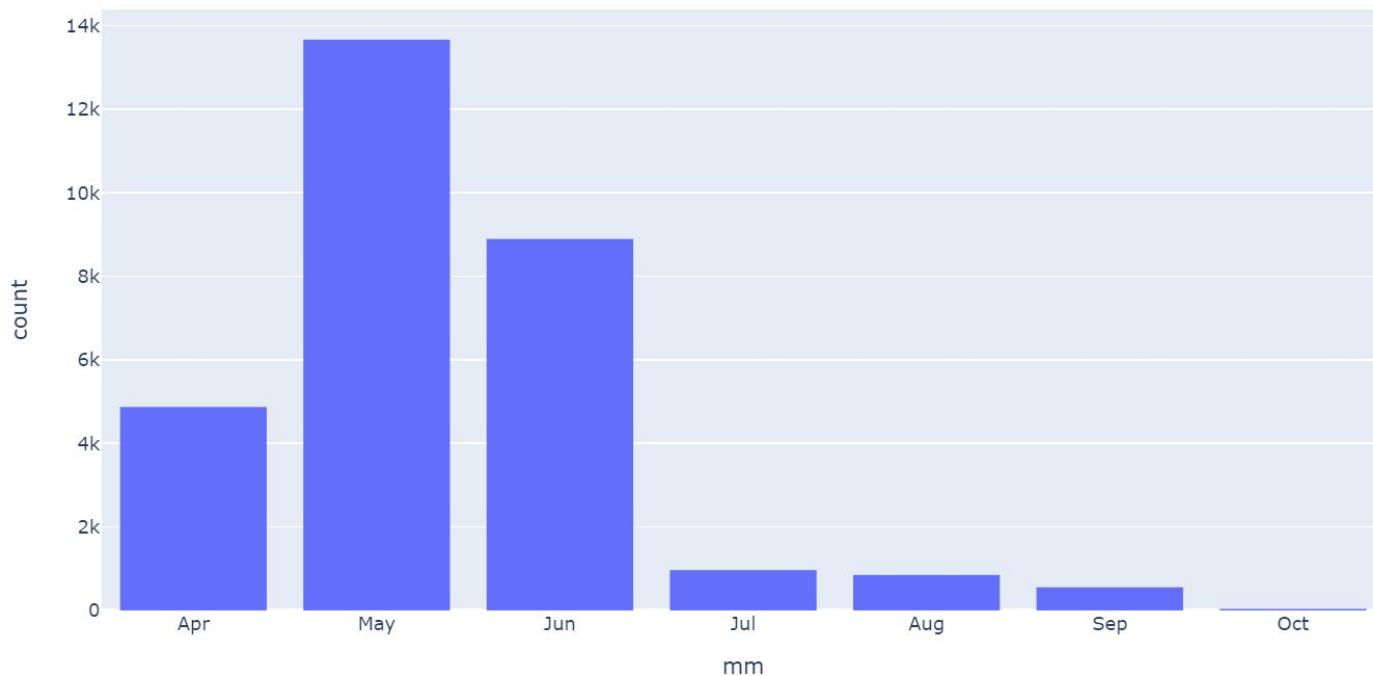
# Message Counts

## Messages counts by Fishery



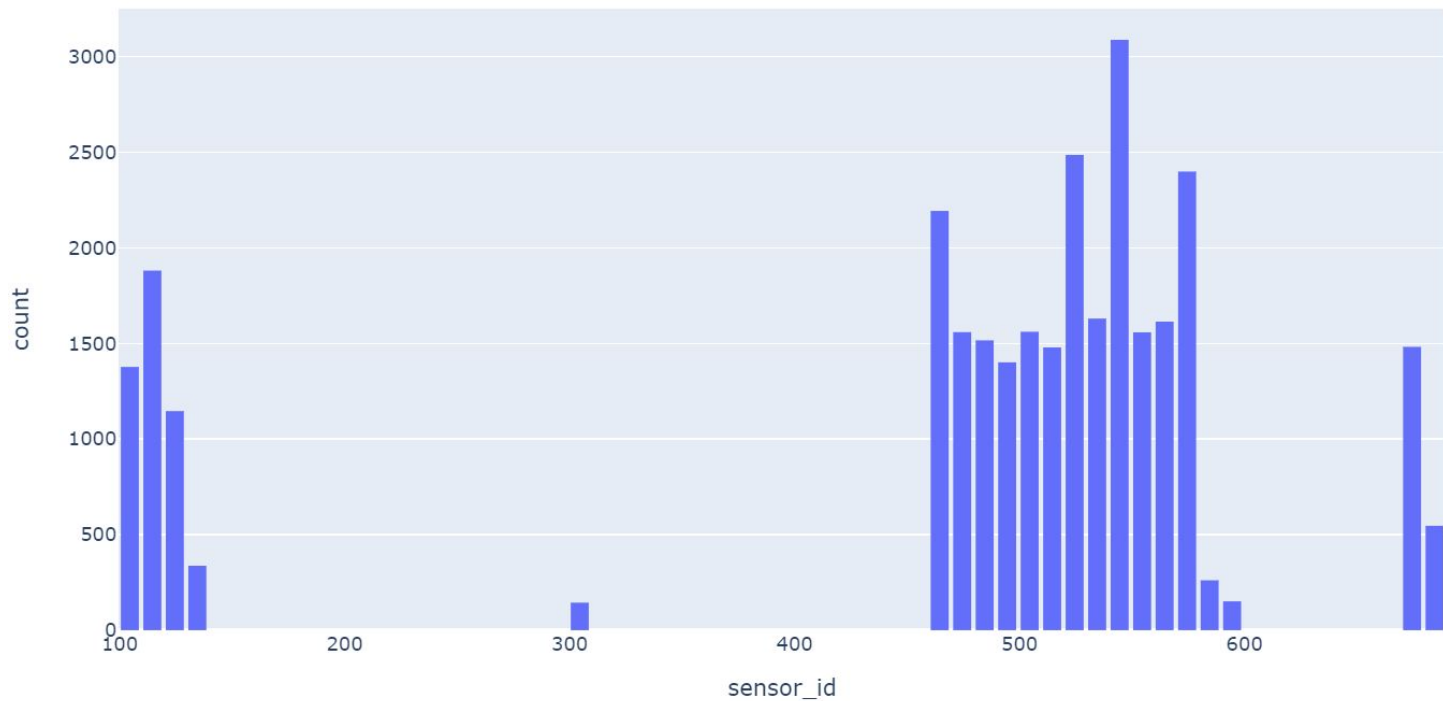
# Message Counts

**Buoy messages across time  
(12th April to 4th Oct 2022)**

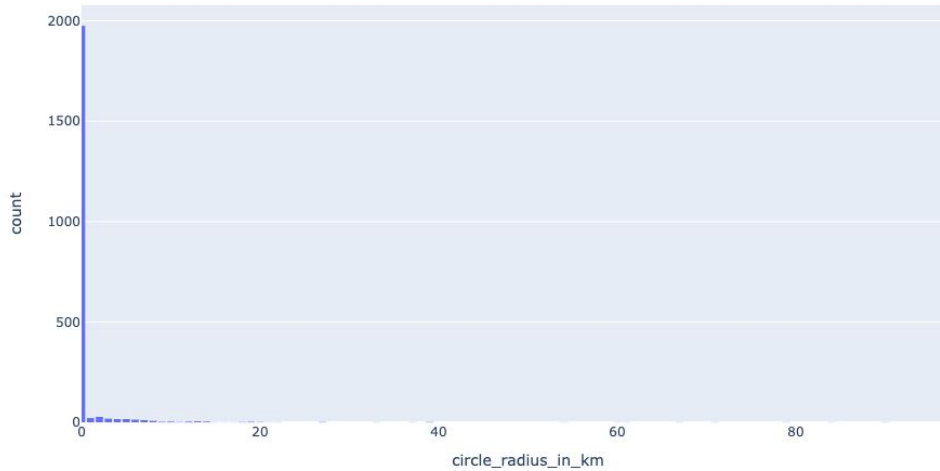


# Message Counts

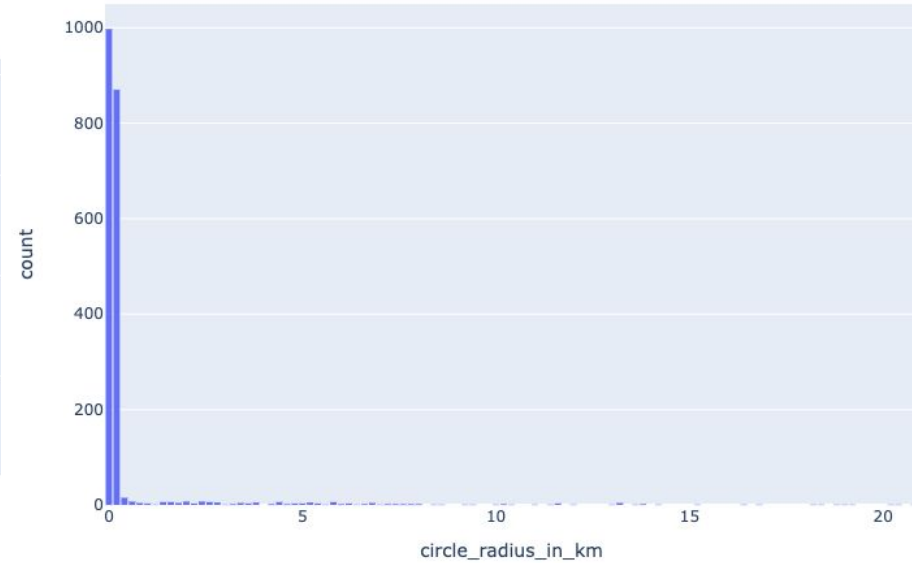
## Messages across Buoy IDs



# Buoy Swing Range



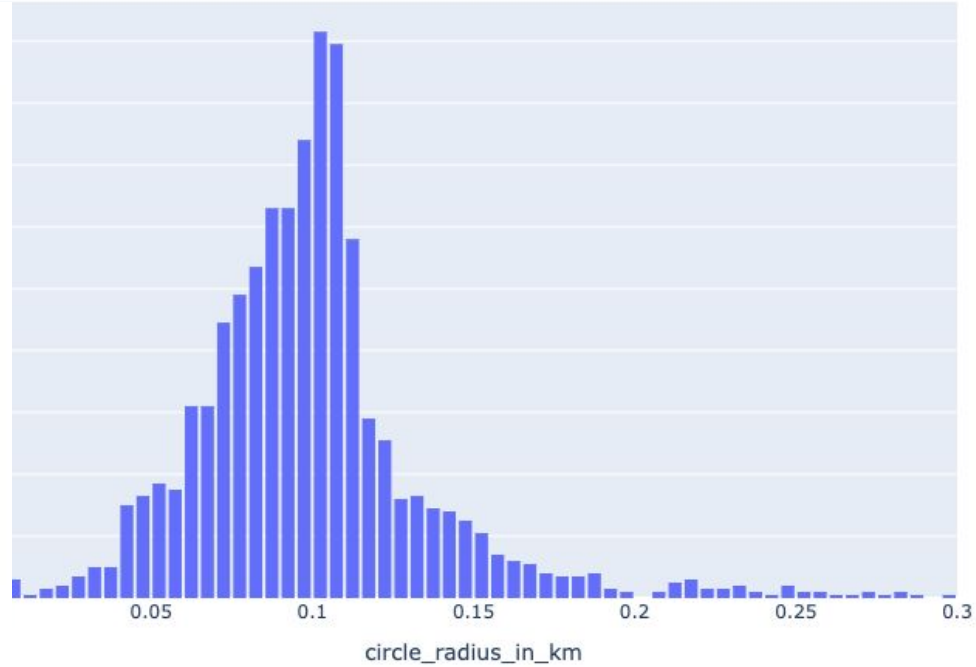
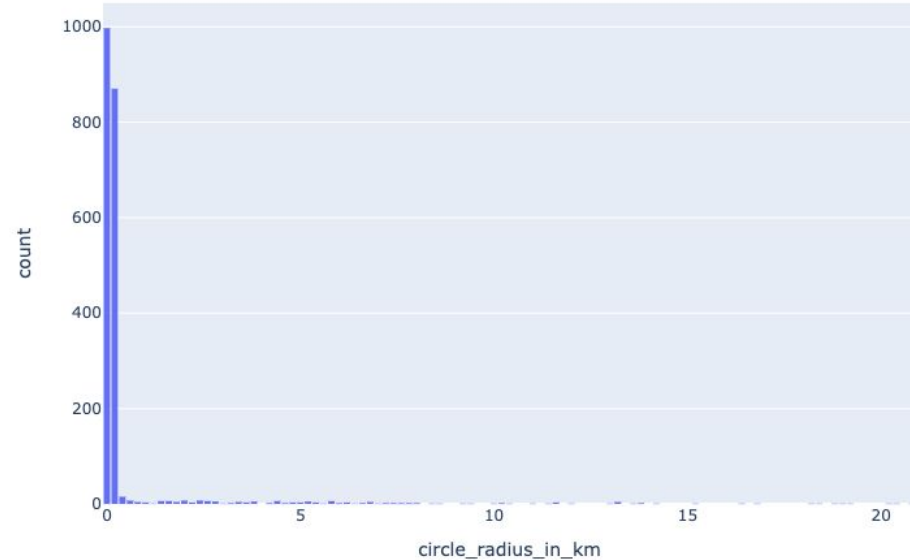
**Swing Range Across Deployments before filtering outliers**



**Swing Range Across Deployments after removing outliers 2 standard deviations away from mean**

# Potential outliers

Average radius size after removing outliers: 0.0996 km

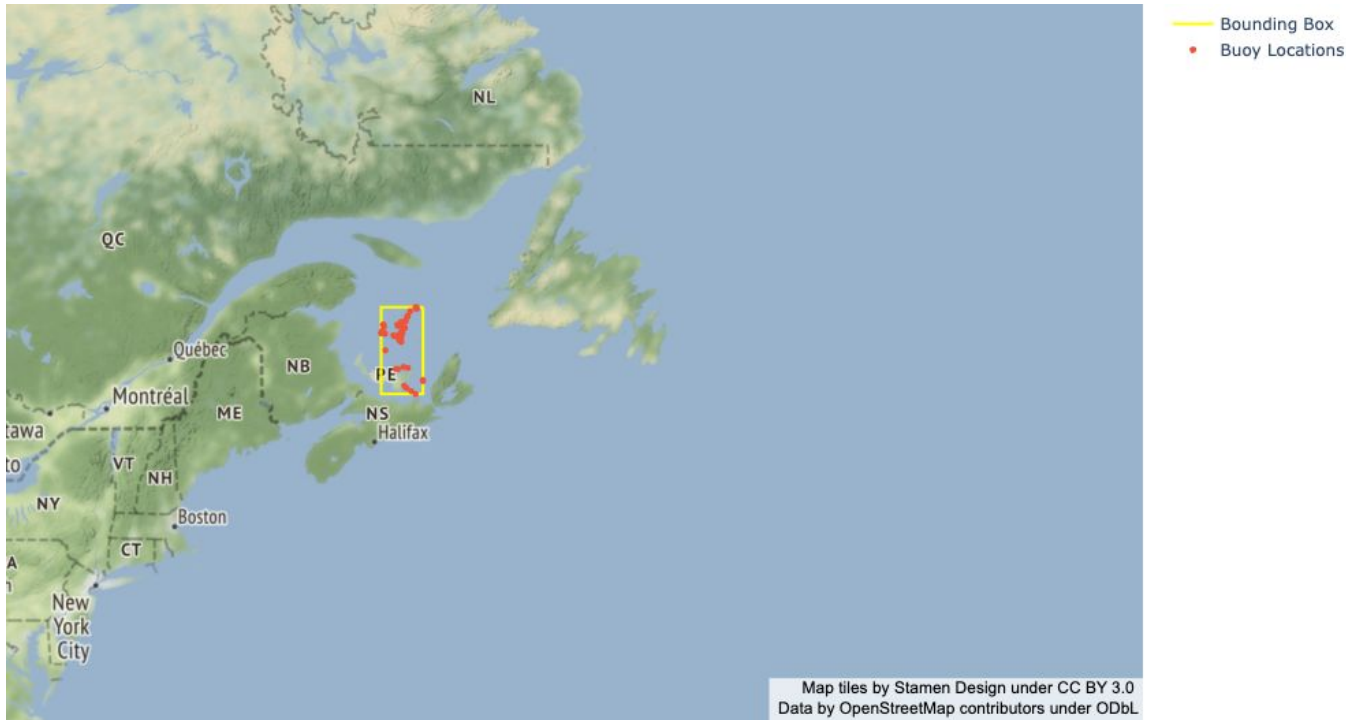


**Swing Range Across Deployments after removing outliers  
2 standard deviations away from mean**

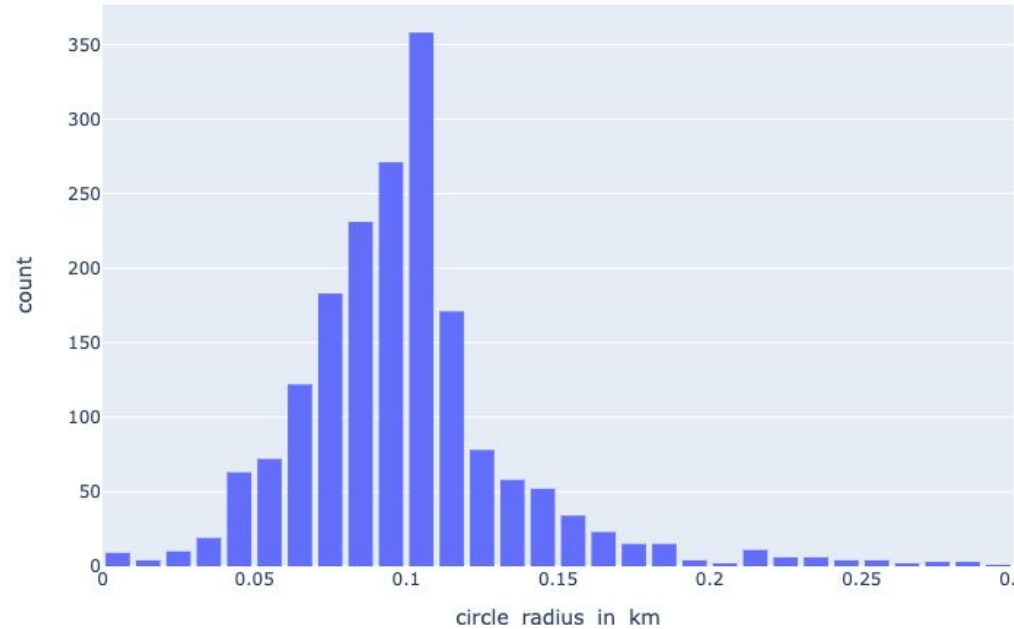
**Swing Range Across Deployments after removing outliers  
Whose circle\_radius\_in\_km > 0.3km**

# Potential Outliers in Map

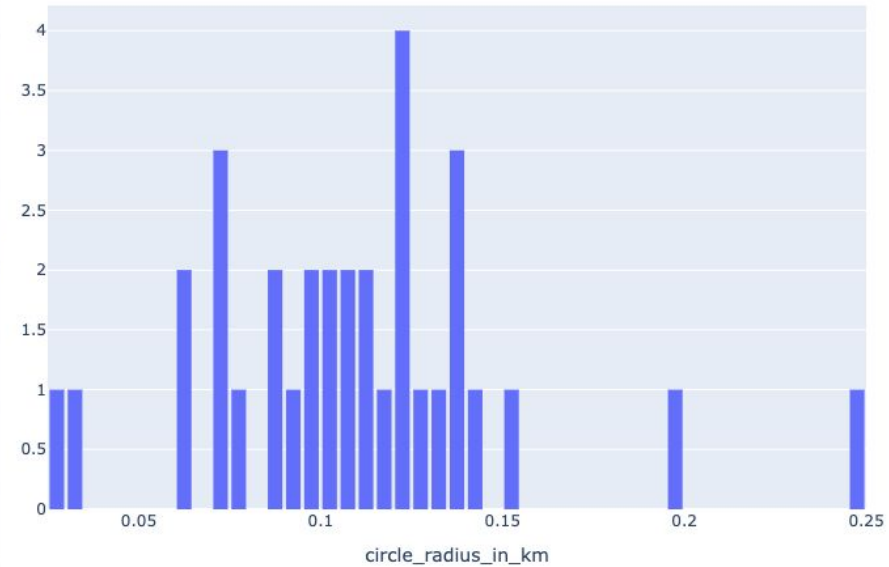
196 outliers in total



# Buoy Swing Range by Fishery

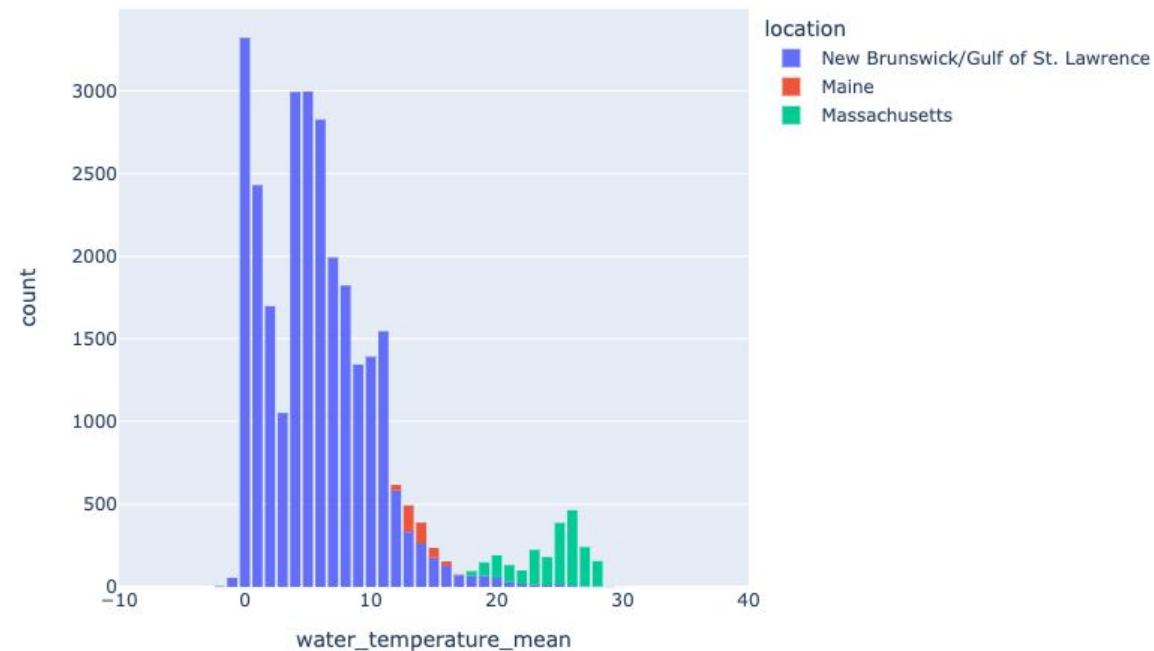


**Final Swing Range Across Deployments -  
New Brunswick/Gulf of St. Lawrence**

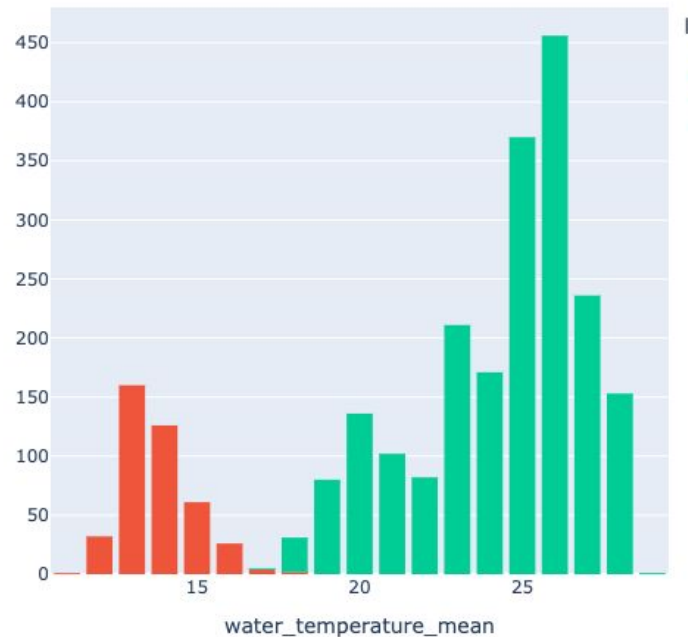


**Final Swing Range Across Deployments - Maine**

# Temperature by Fishery



Water temperature mean across three fishery locations



Water temperature mean in Maine v.s. Massachusetts



# Next Steps

- Start with the smallest enclosing circles developed during EDA as a baseline
- Run hundreds of thousands of simulations of secured buoys in the North Atlantic using [OceanParcels](#)
- Use simulation data to train random forests (RF), long short-term memory neural networks (LSTM), and/or transformers and then fine-tune models on actual BOG buoy trajectories
- Evaluate and compare model results