



Background

The Ocean
Cleanup Project
Introduction

Severe Situation

Research
Goal



Severe Situation

\$ 6-19 Billion

1.4 Billion Pound

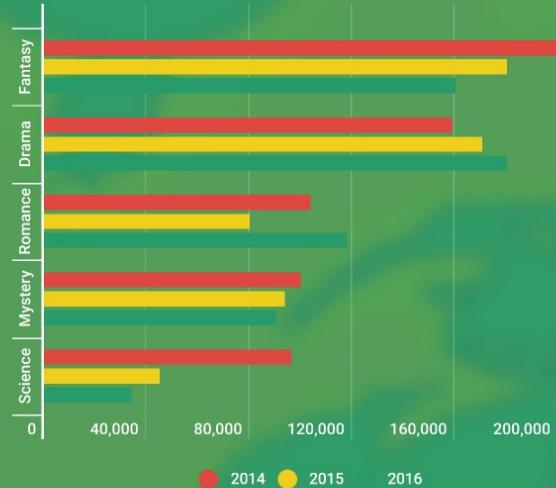
Environment

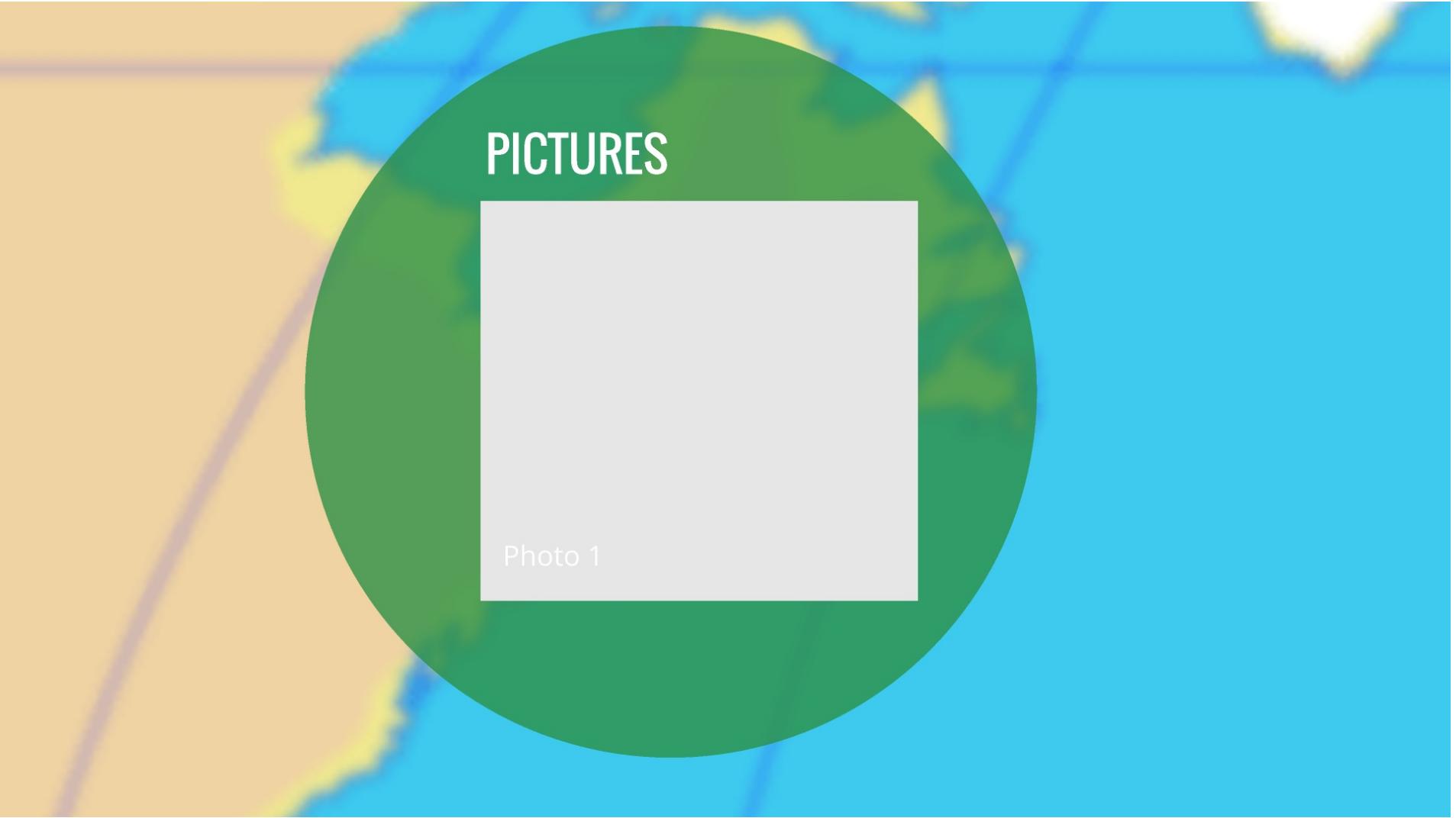
Economy

Health



CHARTS





PICTURES

Photo 1



Click to edit text



Who is the Ocean CleanSwell?



NON-PROFIT
FOUNDATION

ACHIEVEMENT

MISSION





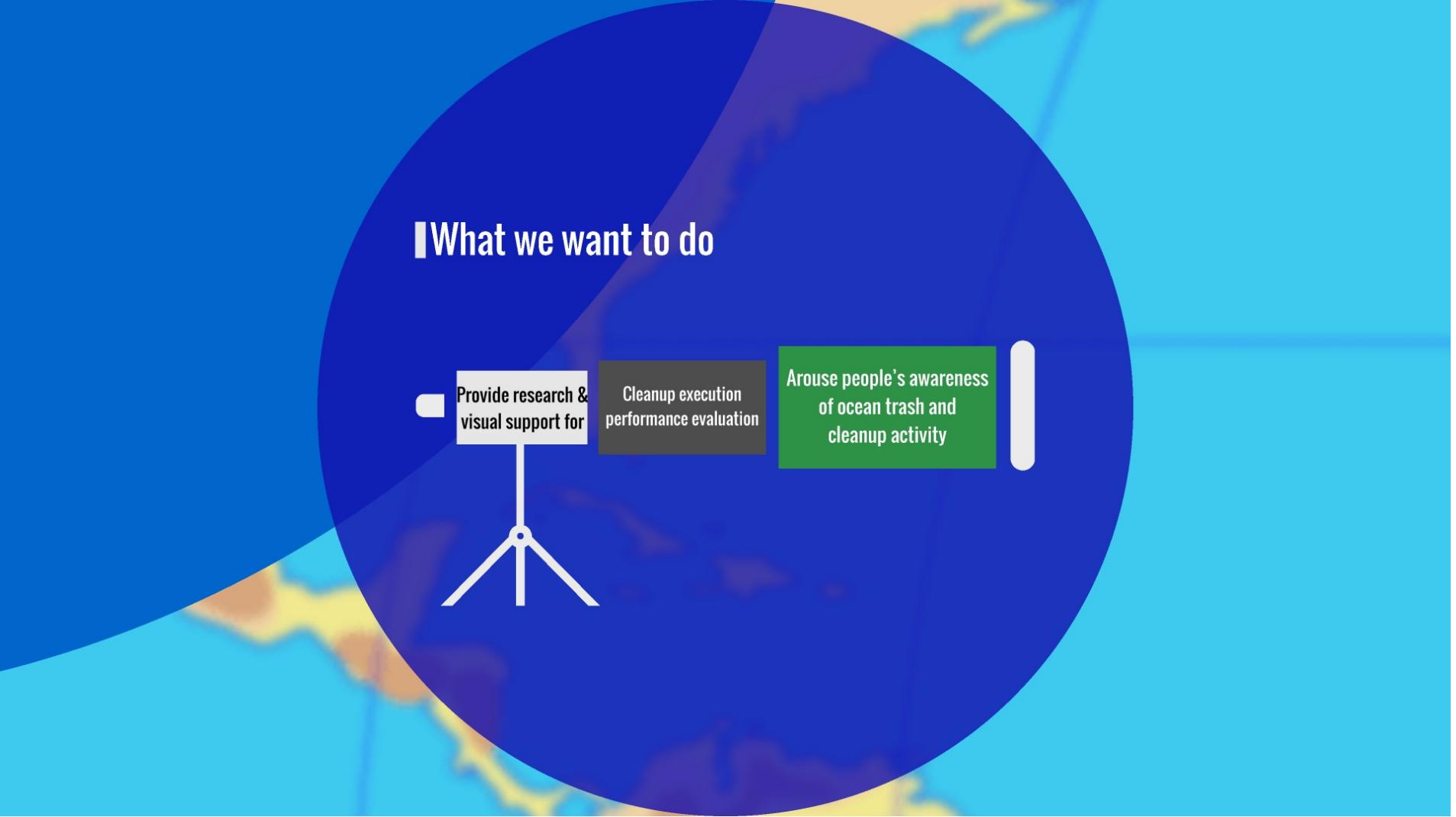
ACHIEVEMENT

For over three decades, more than 12 million volunteers have collected over 220 million pounds of trash.



MISSION

Create SCIENCE-BASED solutions for a healthy ocean AND the wildlife and communities that depend on it.

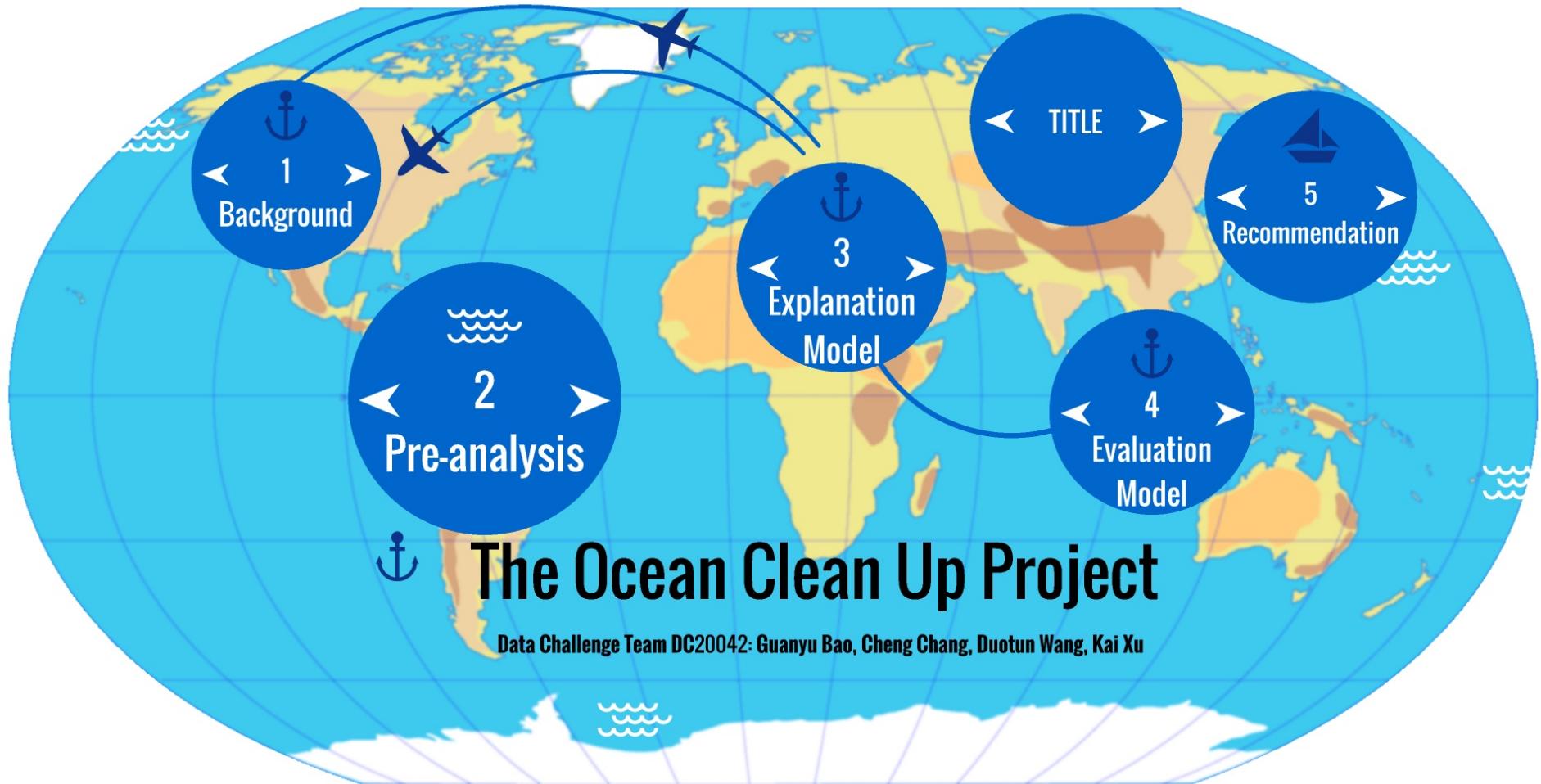


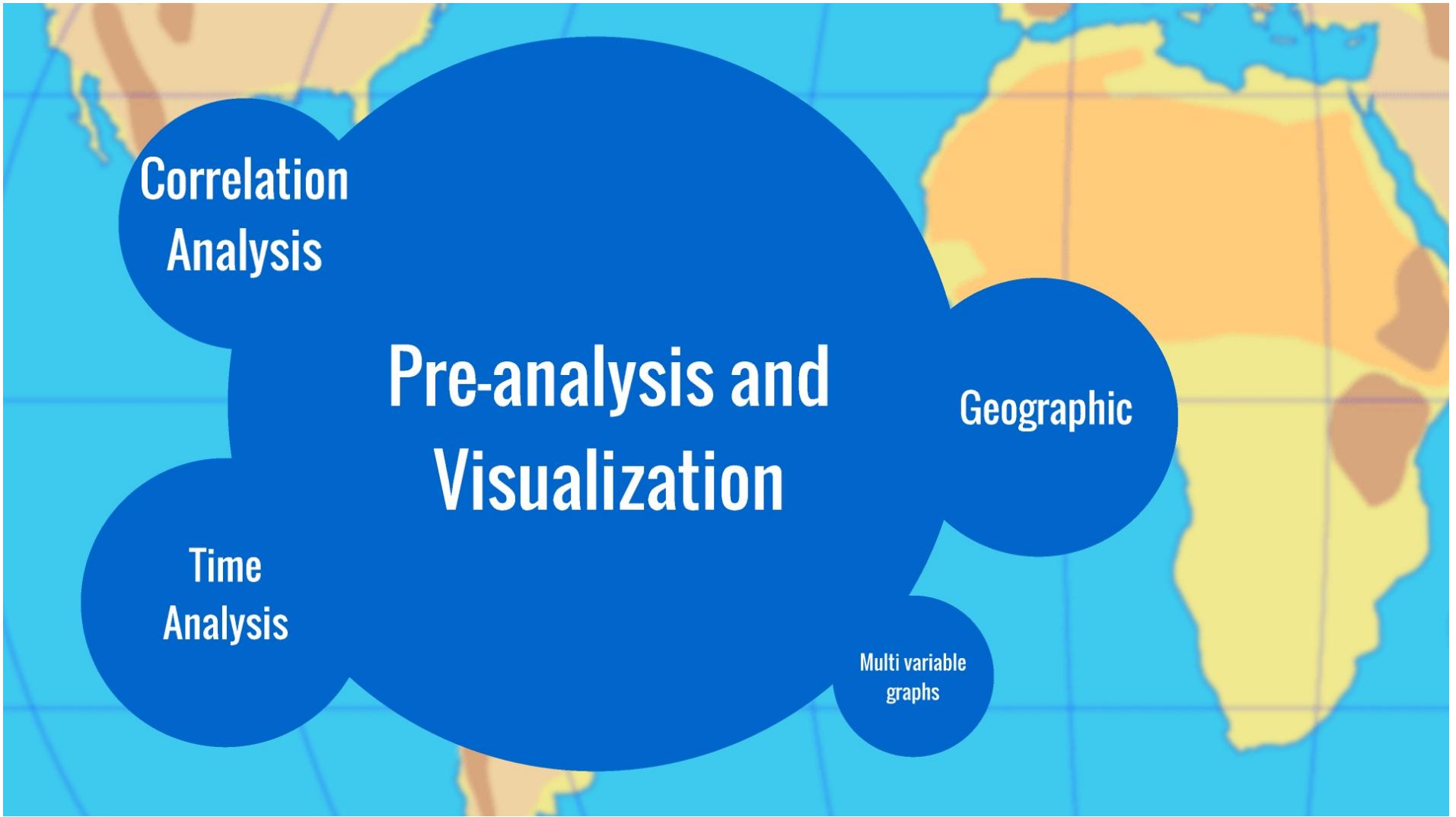
I What we want to do

Provide research &
visual support for

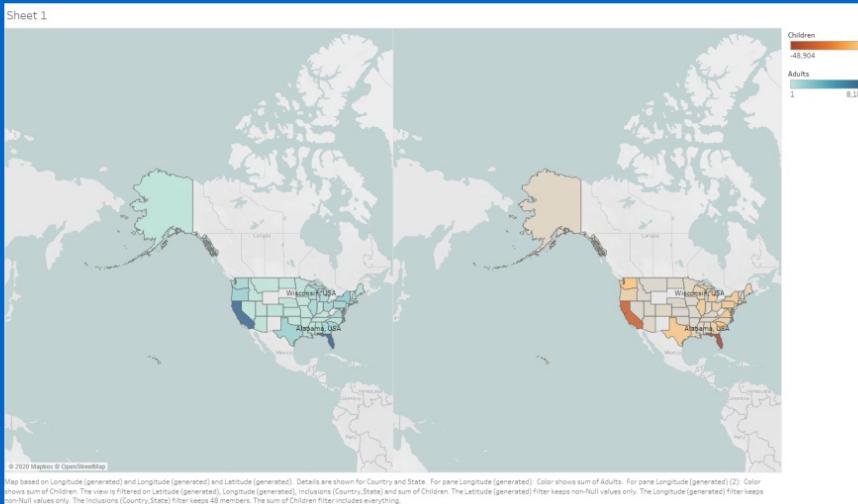
Cleanup execution
performance evaluation

Arouse people's awareness
of ocean trash and
cleanup activity





Adults and Children

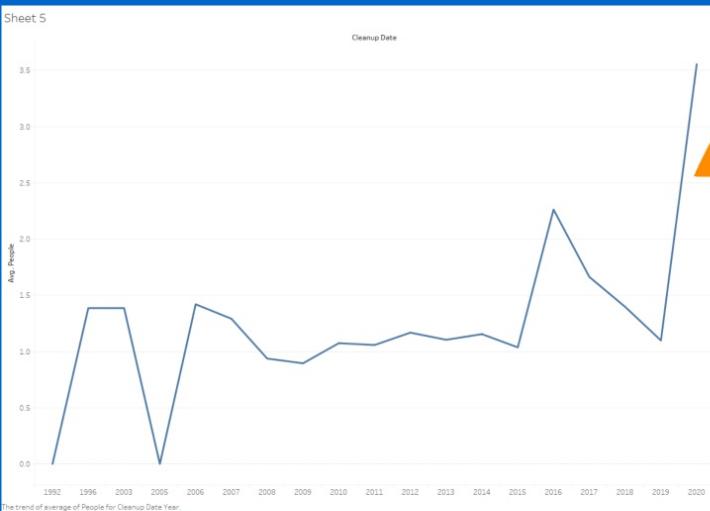


• Strengthen the education for children about the cleanup.



Click to edit text

Activeness through time





Multi-variable graphs

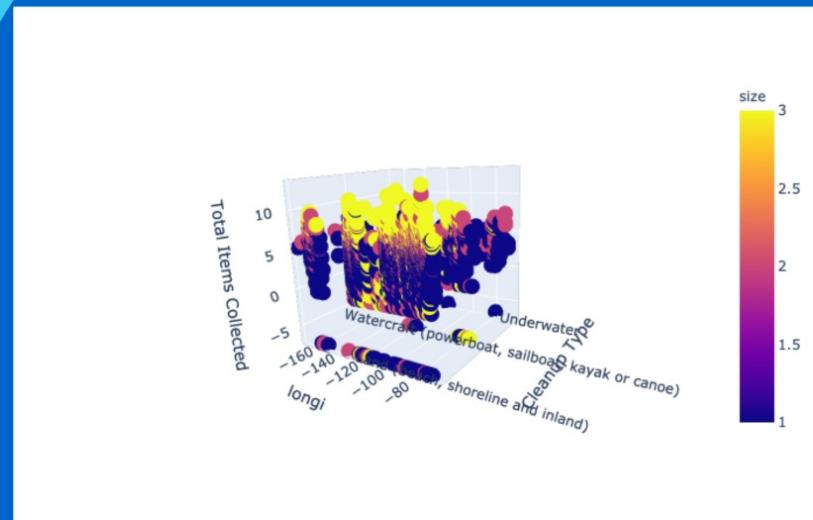
Variables considered:

'Clean up Type',
'Total Items Collected',
'plastic','recycle','toxic','size'"longi','lati'

Size and "Total Item"

Plastic and Recycle

Size and "Total Item"

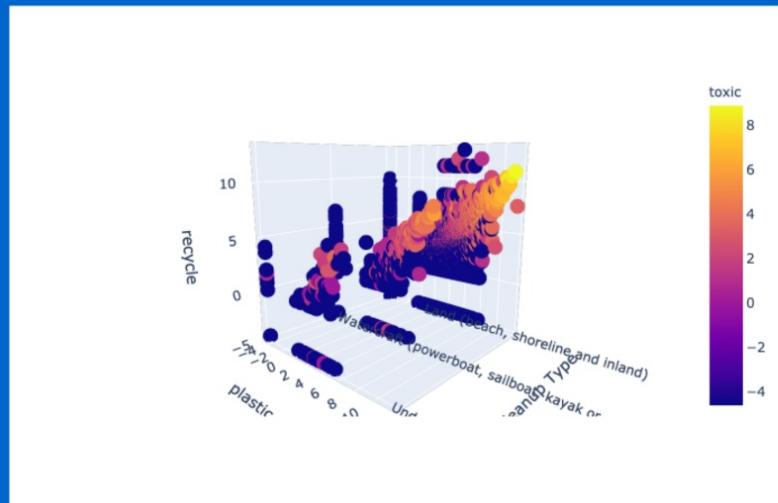


- Size 3 (people > 50) tends to collect more trash.



Click to edit text

Plastic and Recycle

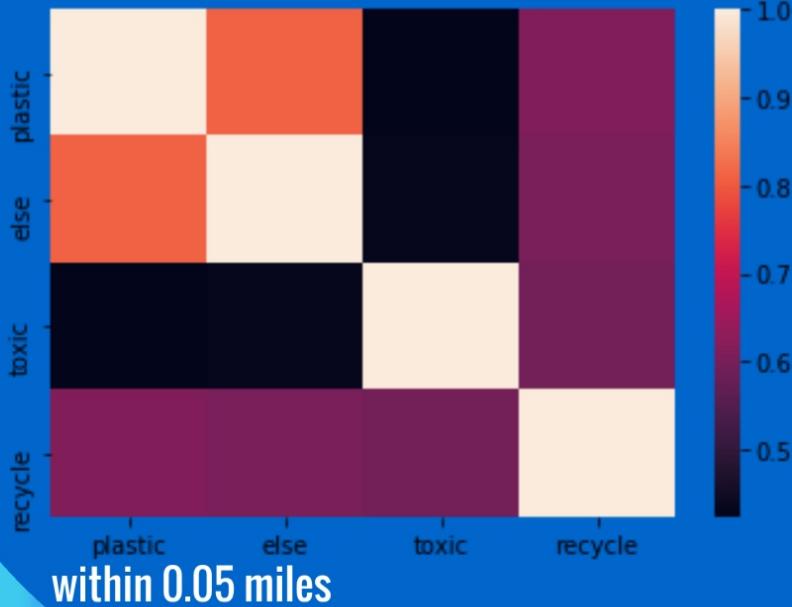


- Plastic and recycle trash always show up together



Click to edit text

Correlation Analysis

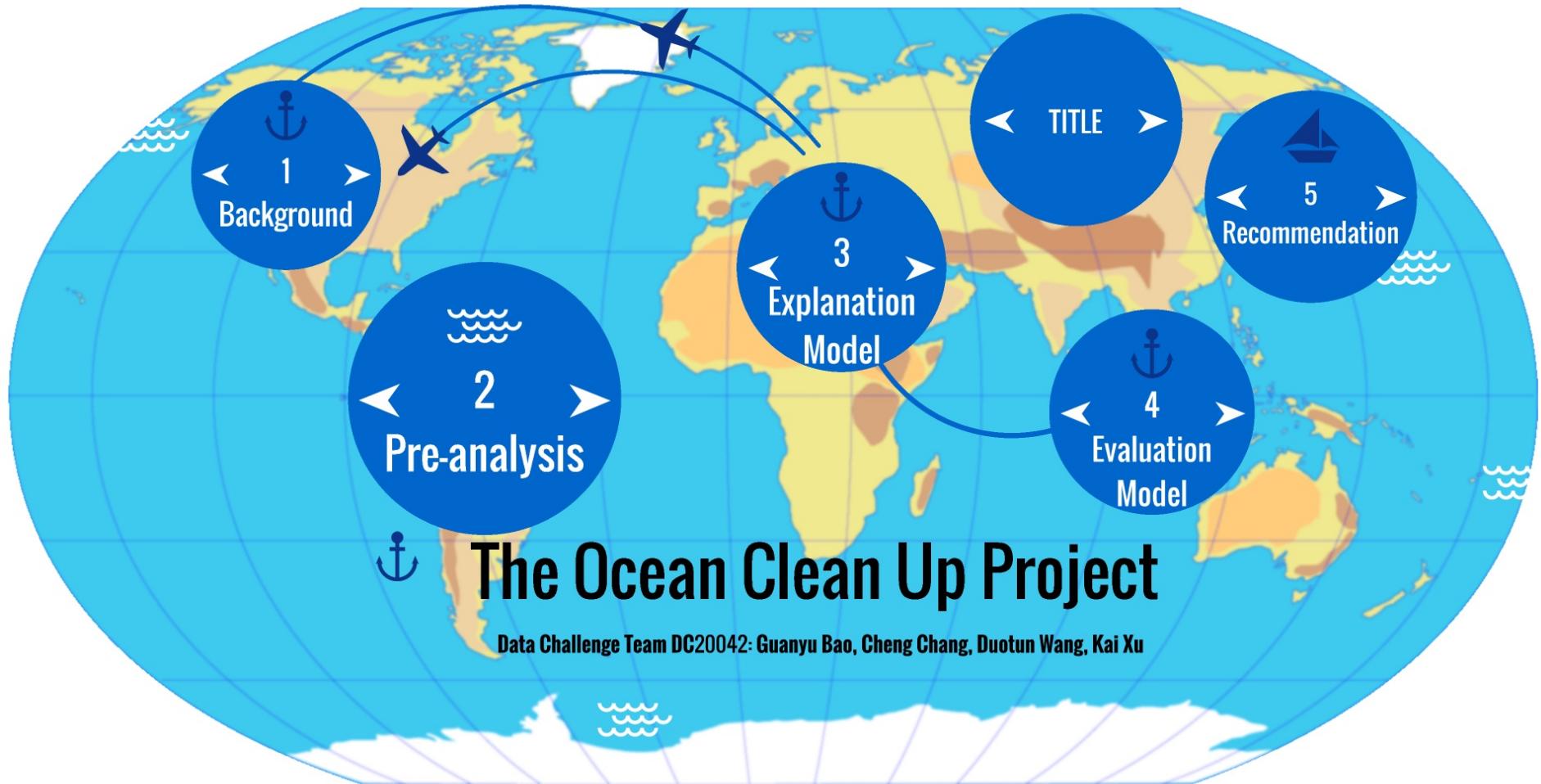


• Plastic and Recycle Trash
always go together

within 0.05 miles



Plastic and Recycle Trash
always go together





TOPIC 3

What factors affect people's participation in cleaning up the Ocean trash?

EPA

Data

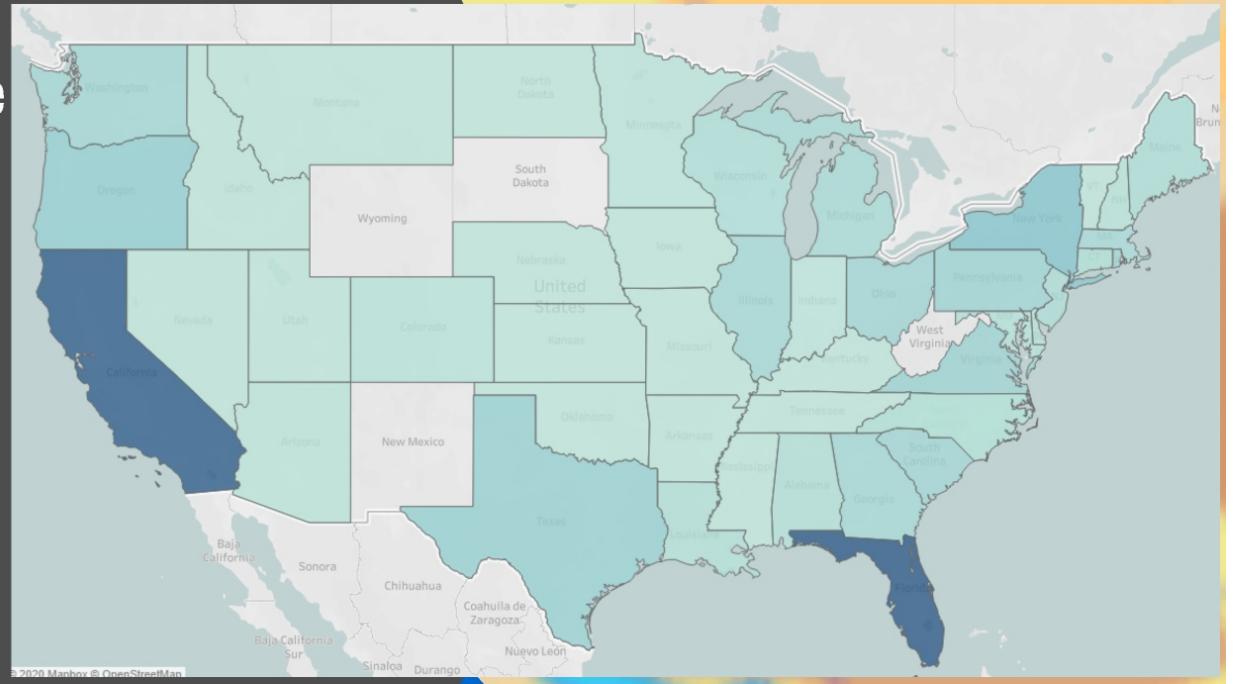
Model

Predict

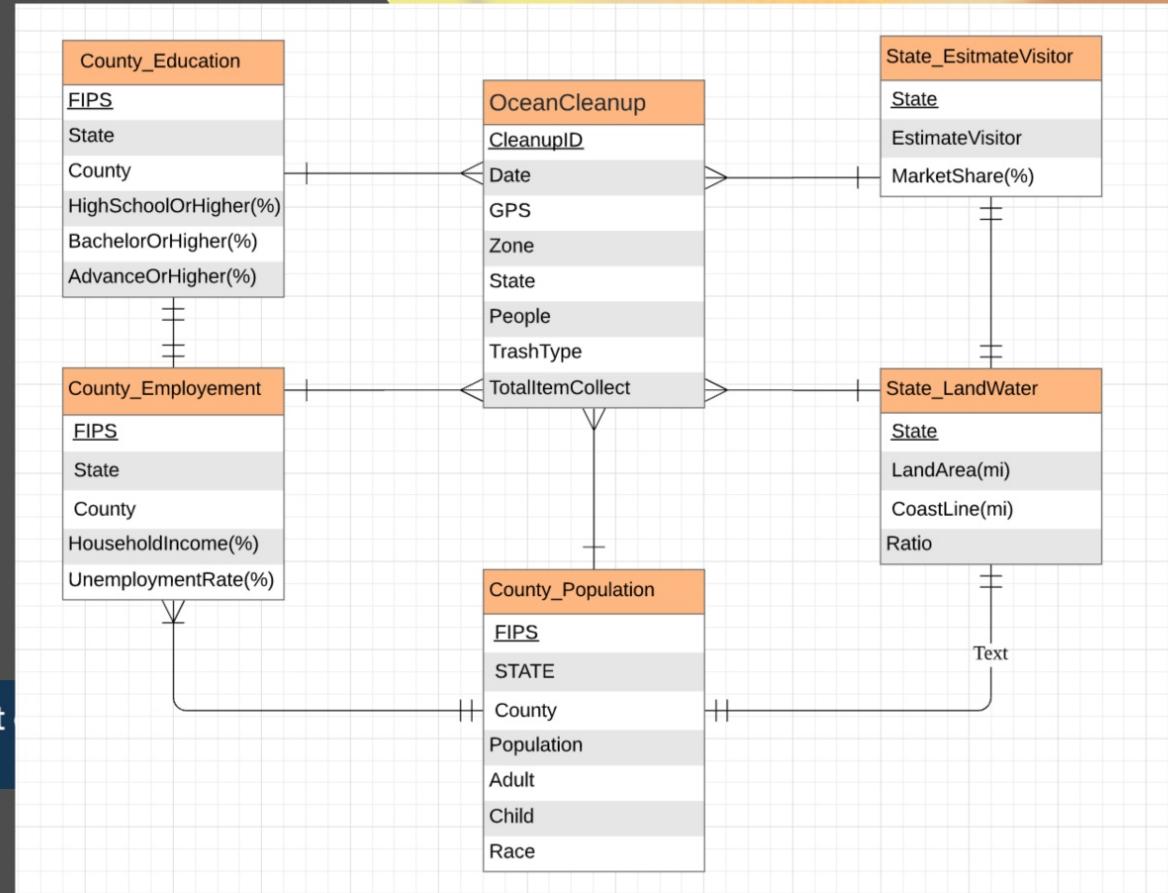
Exploratory Data Analysis

There are significant difference between states. But why?

- Tourism?
- Coastline?
- Education level?



External Data resource:



Model

Independent variable:

- Population
- Income
- Education
- Tourism
- Coastline

Most significant variable:

- Coastline
- Population
- Education

OLS Regression Results								
Dep. Variable:	People	R-squared:	0.424					
Model:	OLS	Adj. R-squared:	0.417					
Method:	Least Squares	F-statistic:	56.74					
Date:	Fri, 28 Feb 2020	Prob (F-statistic):	1.51e-85					
Time:	15:09:40	Log-Likelihood:	-7088.1					
No. Observations:	781	AIC:	1.420e+04					
Df Residuals:	770	BIC:	1.425e+04					
Df Model:	10							
Covariance Type:	nonrobust							
	coef	std err	t	P> t	[0.025]			
Intercept	513.7678	646.263	0.795	0.427	-754.878			
Sea[T.True]	-822.2510	723.515	-1.136	0.256	-2242.546			
Population	0.0010	0.000	3.700	0.000	0.000			
Population:Sea[T.True]	0.0020	0.000	6.378	0.000	0.001			
medHousInc	-0.0136	0.016	-0.849	0.396	-0.045			
medHousInc:Sea[T.True]	0.0233	0.018	1.305	0.192	-0.012			
Bachelor	13.7636	22.125	0.622	0.534	-29.670			
Bachelor:Sea[T.True]	-28.6540	25.526	-1.123	0.262	-78.763			
MarketShare	-1320.6466	3510.895	-0.376	0.707	-8212.708			
MarketShare:Sea[T.True]	827.7041	3793.261	0.218	0.827	-6618.656			
Coastline	0.6603	0.177	3.723	0.000	0.312			
Omnibus:	710.708	Durbin-Watson:	2.017					
Prob(Omnibus):	0.000	Jarque-Bera (JB):	38047.037					
Skew:	3.880	Prob(JB):	0.00					
Kurtosis:	36.301	Cond. No.	6.08e+07					

Prediction

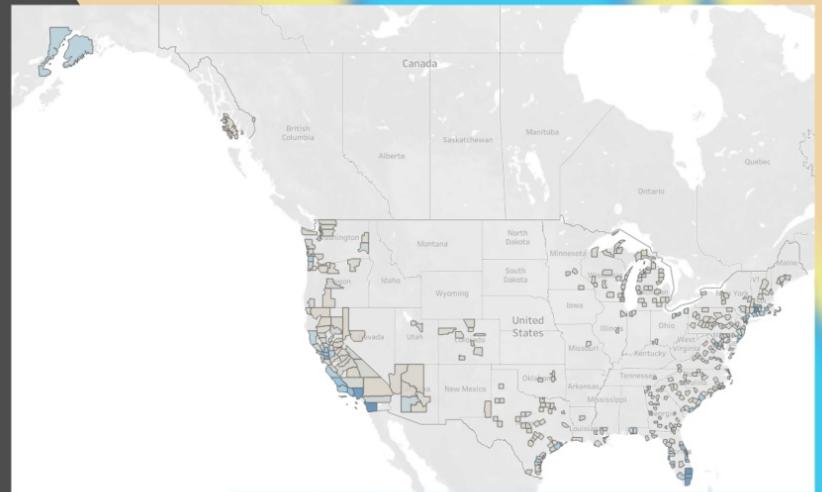
Base on our model, people in these area participate less than expectation.

So why?

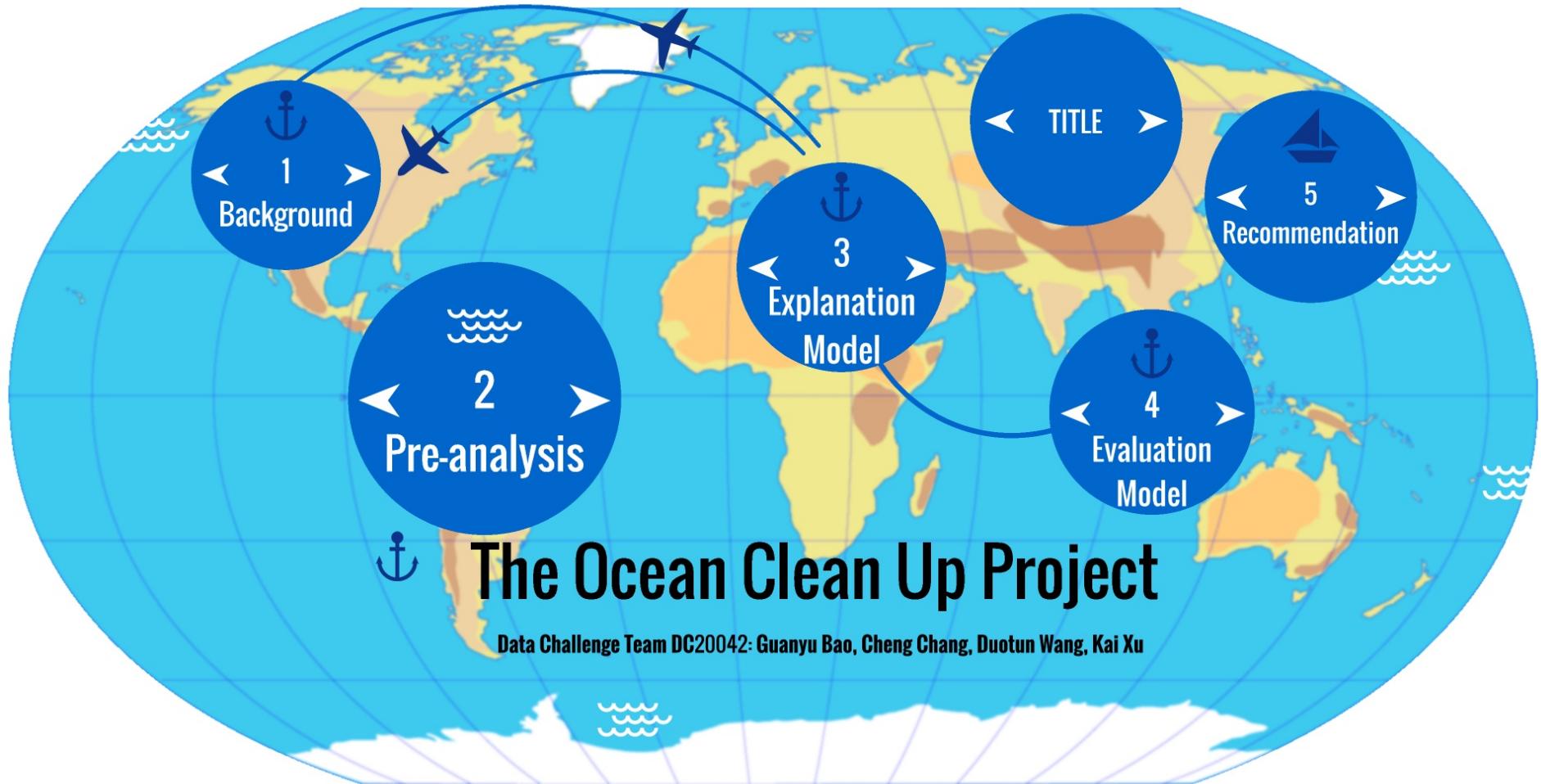
- Lack of awareness
- Lack of local volunteering events
- Not many trash (Ocean Current)

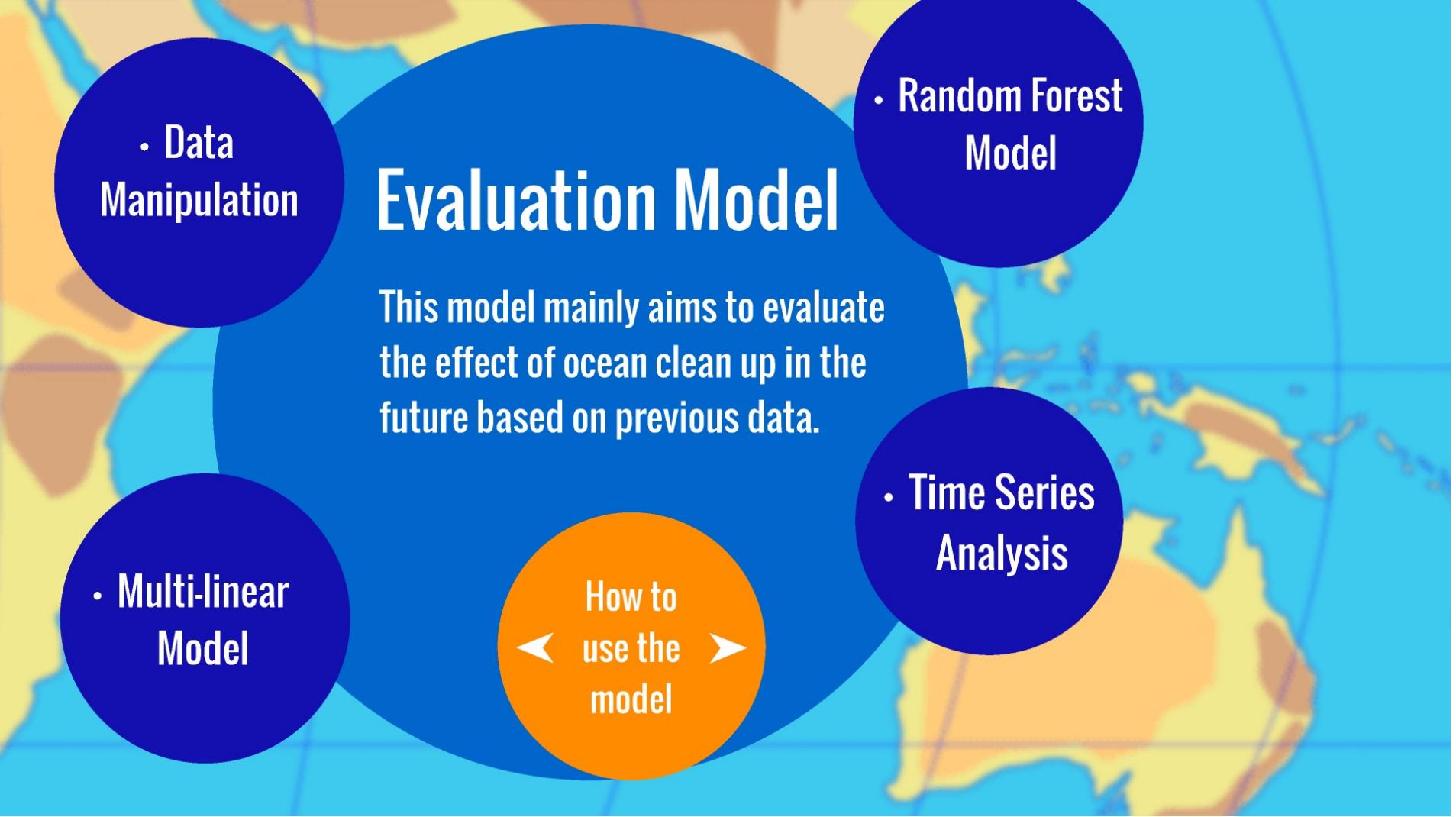
What's next?

- Promote at local neighbor
- Cooperate with local organization
- More Data about marine debris



State	County	Predict	Actual
TX	Bexar County	5437	310
TX	Dallas County	7169	3018
NV	Clark County	5533	1963
AK	Prince of Wales-Hyder Censu	4624	1060
CA	San Bernardino County	6380	3487
MI	Oakland County	2892	27
OH	Franklin County	2945	208
MN	Hennepin County	2822	404
TX	Collin County	2925	518
TX	EL Paso County	2383	0



- 
- Data Manipulation

Evaluation Model

- Random Forest Model

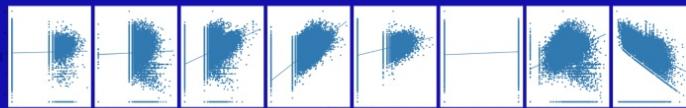
- Multi-linear Model

How to use the model

- Time Series Analysis

- 
- **Data Manipulation**
 1. Correction of input mistake
 2. Clean time in the future
 3. Factors created
 4. Log transformation

Relationship between garbage density and independent variables



- Density=1.0309-0.028 Children-0.27 People+0.11 recycle+ 0.81 plastic+0.01 toxic+ 0.027 east+ 0.31 Pounds-1.04 Miles
- Using CV: R square score out-of-sample is **0.8051155645275859**
- Lasso Selection: Eliminating 'East'
- 10.47-9.14** is the benchmark range which defines the effect of ocean clean up activity



OLS Regression Results					
Dep. Variable:	density	R-squared:	0.817		
Model:	OLS	Adj. R-squared:	0.817		
Method:	Least Squares	F-statistic:	9012.		
Date:	Fri, 28 Feb 2020	Prob (F-statistic):	0.00		
Time:	16:23:16	Log-Likelihood:	-46130.		
No. Observations:	24571	AIC:	9.228e+04		
Df Residuals:	24562	BIC:	9.235e+04		
Df Model:	8				
Covariance Type:	HC1				
coef	std err	z	P> z	[0.025	0.975]
Intercept	1.0309	0.050	19.283	0.000	0.926 1.136
Children	-0.0280	0.004	-7.174	0.000	-0.036 -0.020
People	-0.2697	0.012	-22.579	0.000	-0.293 -0.246
recycle	0.1059	0.006	19.208	0.000	0.095 0.117
plastic	0.8142	0.009	89.250	0.000	0.796 0.832
toxic	0.0094	0.005	1.916	0.055	-0.000 0.019
east	0.0267	0.024	1.192	0.233	-0.017 0.071
Pounds	0.3093	0.010	31.685	0.000	0.290 0.328
Miles	-1.0357	0.006	-172.932	0.000	-1.047 -1.024
Omnibus:	3529.604	Durbin-Watson:	1.669		
Prob(Omnibus):	0.000	Jarque-Bera (JB):	47621.403		
Skew:	-0.199	Prob(JB):	0.00		
Kurtosis:	9.809	Cond. No.	31.4		

Random Forest

- Max Depth=9
- test R square 0.932040271677322
- -19.35~20.1 is the benchmark range which defines the effect of ocean clean up activity.

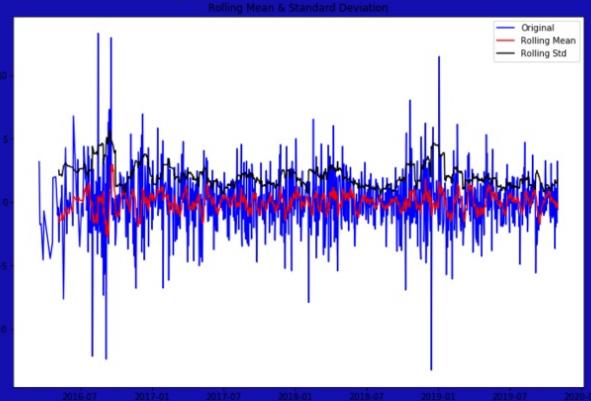
		importance
0	east	0.003378
1	Children	0.004184
2	toxic	0.004401
3	People	0.011428
4	recycle	0.024647
5	Pounds	0.081157
6	Miles	0.196827
7	plastic	0.673978



Time Series Analysis across the country

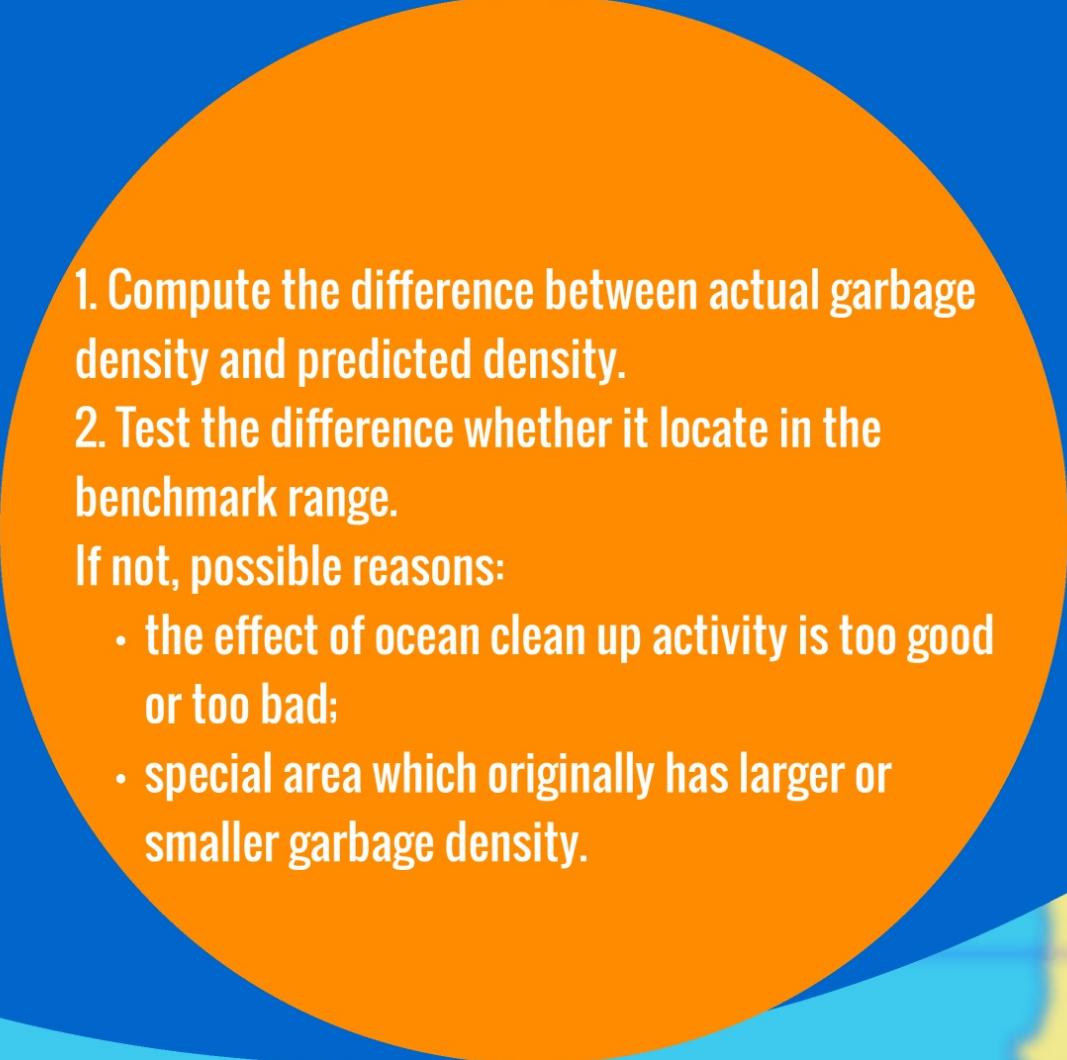
Model Results

• Stationary Time Series



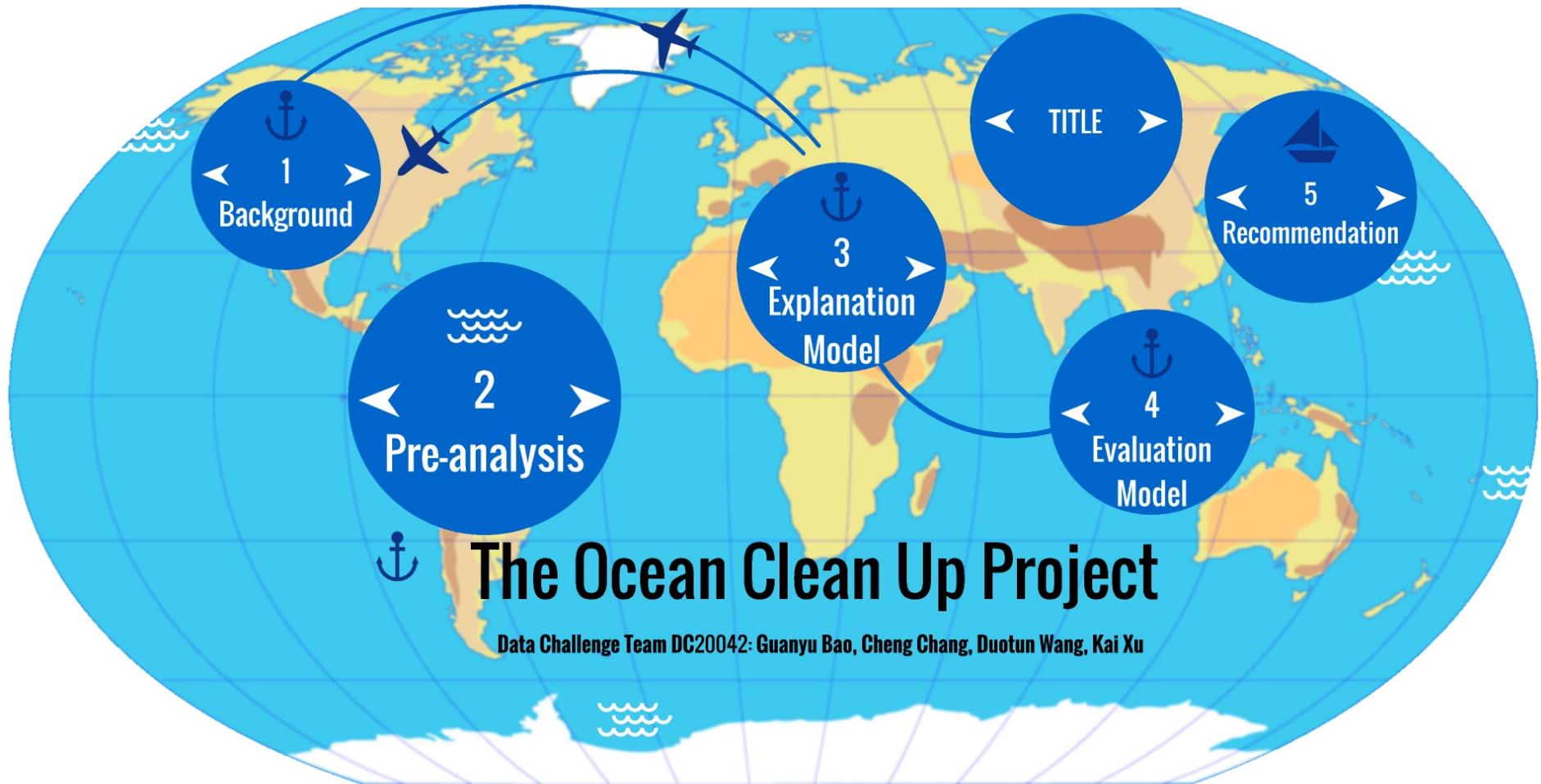
Model Prediction

	lower density	upper density
2019-10-30	2.797056	11.234539
2019-10-31	1.711105	10.148589
2019-11-01	4.993564	13.431047
2019-12-01	2.668208	11.105691
2020-01-01	1.183581	13.115984
2020-02-01	-0.322921	14.291228
2020-03-01	-1.535304	15.339662
2020-04-01	-2.682693	16.183911
2020-05-01	-3.577857	17.089339
2020-06-01	-4.246993	18.076027
2020-07-01	-5.116718	18.747513
2020-08-01	-5.864457	19.447314
2020-09-01	-6.358134	20.322758
2020-10-01	-7.326875	20.656231
2020-11-01	-7.659055	21.568303

- 
1. Compute the difference between actual garbage density and predicted density.
 2. Test the difference whether it locate in the benchmark range.

If not, possible reasons:

- the effect of ocean clean up activity is too good or too bad;
- special area which originally has larger or smaller garbage density.



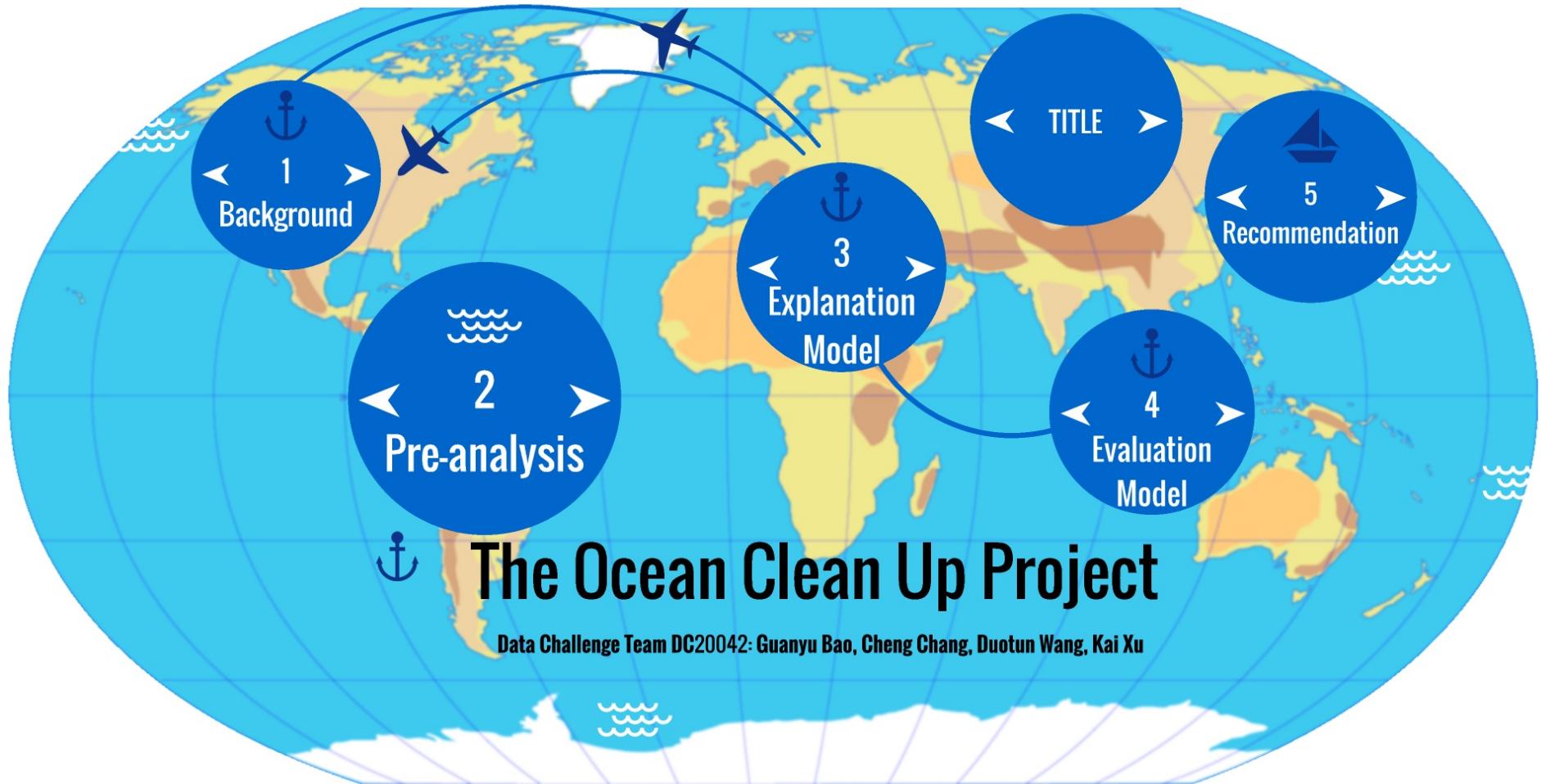


Recommendation

- Use other datasets to combine and improve our model result.

For
CleanSwell







Click to edit text

