Toxic Situation

...

A Bi-Coastal Perspective

Chelsea Thompson, Deirdre Clark, Enjoli Baez, Kaleah French, Kareem Todman, Saeeda Choudhury

Table of Contents

Motivation & Summary - Kareem

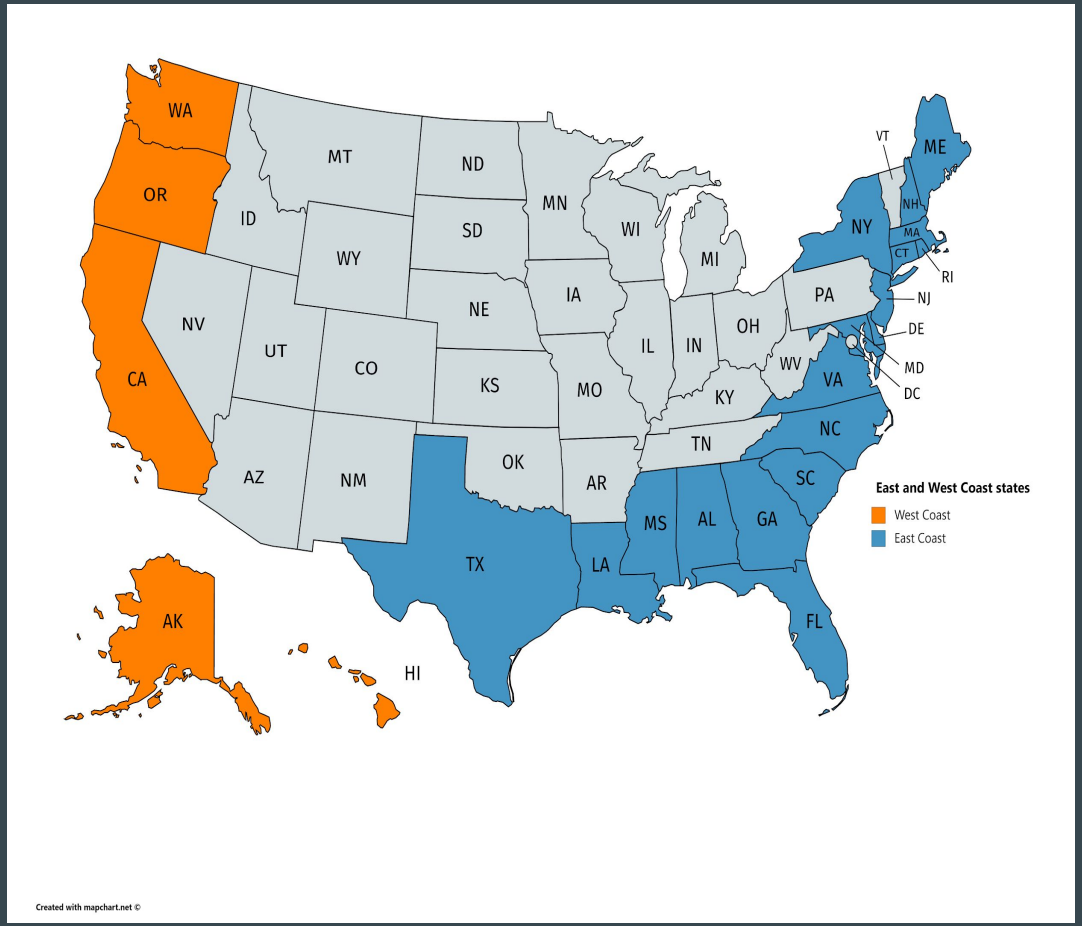
What is point vs nonpoint sources of pollutants?

Hypothesis

Using census data (from 2014-2018) to formulate our analyses.

Our purpose of analyzing this information is to develop more accurate conclusions.

In summary...



Questions & Data - Deirdre

1. Nonpoint pollution

Nonpoint - Ocean Conservancy

2. Population Density

U.S. Census Bureau

3. Total Estimated GDP (Gross Domestic Product)

U.S. Bureau of Economic Analysis

NOAA - National Oceanic & Atmospheric Administration

4. Laws & Regulations

U.S. EPA (Environmental Protection Agency)

Data Clean-Up and Exploration - Saeeda

- Laws and Regulations, State Pollution, State GDP
- State GDP as API
- Laws and Regulations & State Pollution as CSVs

```
# Delete unnecessary columns
EC_dropped = east_coast.drop(columns = ['Cleanup ID', 'Zone', 'Country', 'Cleanup Date',
                                        'Group Name', 'Adults', 'Children', 'People'])
EC_dropped.head()
```

```
# Rename long column names to simpler forms
EC_named = EC_dropped.rename(columns = {"Food Wrappers (candy, chips, etc.)" : "Food Wrappers", "Take Out/Away Containers (Pl
                                        "Take Out/Away Containers (Foam)" : "Food Containers (Foam)", "Fish
                                        "Fishing Net & Pieces" : "Fishing Nets", "Fishing Line (1 yard/mete
                                        "Rope (1 yard/meter = 1 piece)" : "Rope", "Fishing Gear (Clean Swel
                                        "Other Plastic Bottles (oil, bleach, etc.)" : "Other Plastic Bottle
                                        "Other Packaging (Clean Swell)" : "Other Packaging", "Other Trash (
                                        "Personal Hygiene (Clean Swell)" : "Hygiene"})
EC_named.head()
```

Data Clean-Up and Exploration continued

```
# Group plastic columns

EC_plastics = EC_named[["State", "Food Containers (Plastic)", "Bottle Caps (Plastic)", "Lids (Plastic)", "Straws, Stirrers",
                        "Beverage Bottles (Plastic)", "Grocery Bags (Plastic)", "Other Plastic Bags",
                        "Cups, Plates (Plastic)", "6-Pack Holders", "Other Plastic Bottles",
                        "Condoms", "Plastic Pieces"]]
EC_plastics.head()
```

	State	Food Containers (Plastic)	Bottle Caps (Plastic)	Lids (Plastic)	Straws, Stirrers	Beverage Bottles (Plastic)	Grocery Bags (Plastic)	Other Plastic Bags	Cups, Plates (Plastic)	6-Pack Holders
0	Maine, USA	0	0	0	0	0	0	0	0	0
1	Maine, USA	0	0	0	0	0	0	0	0	0
2	Maine, USA	11	4	3	6	17	7	0	7	0
3	Maine, USA	4	10	1	6	8	3	0	0	0
4	Maine, USA	0	0	6	14	10	19	0	0	0

H
START_DATE
11-APR-1201 00:00:00
11-APR-1201 00:00:00
3/28/2019 0:00
3/28/2019 0:00
3/14/2019 0:00

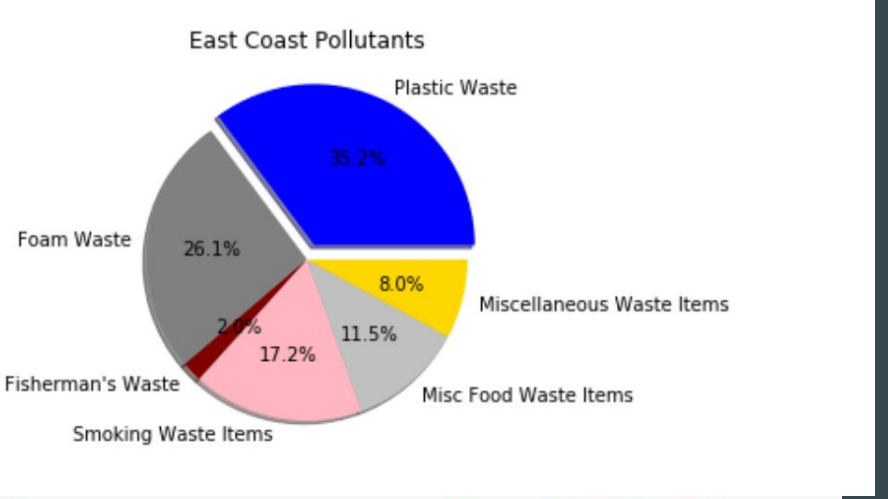
- Removing unnecessary columns
- Renaming long column names
- Sorting related information
- Formatting mismatching data

```
# Group Plastic pollution by state
EC_plastic_group = EC_plastics.groupby(["State"])
EC_plastic_count = EC_plastic_group.sum()
EC_plastic_count
```

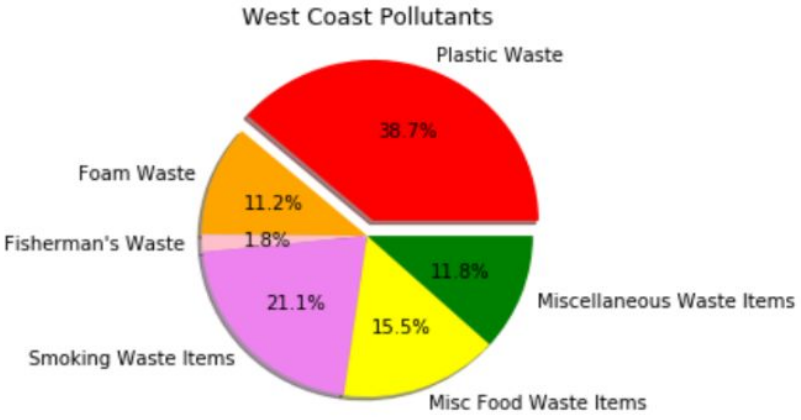
State	Food Containers (Plastic)	Bottle Caps (Plastic)	Lids (Plastic)	Straws, Stirrers	Beverage Bottles (Plastic)	Grocery Bags (Plastic)	Other Plastic Bags	Cups, Plates (Plastic)	6-Pack Holders	Other Plastic Bottles	Condoms	Plastic Pieces
Alabama, USA	1932	6682	2411	3206	9613	2385	1793	1326	333	1649	64	9337
Connecticut, USA	712	4830	1409	2242	5901	1956	2089	1178	73	1669	79	10014
Delaware, USA	939	7364	1179	2836	2547	922	1145	801	61	225	51	10098
Florida, USA	12455	72144	14585	26604	37356	23125	10607	11279	732	2080	615	186765
Georgia, USA	4757	8073	3971	4180	24848	9891	4290	3622	603	2070	529	7123
Louisiana, USA	572	5370	900	1625	4761	1032	683	835	77	200	58	8198
Maine, USA	443	1718	536	842	1566	540	884	375	41	308	26	7273
Maryland, USA	719	4908	805	4761	8863	1162	1212	515	69	233	30	10627
Massachusetts, USA	1036	8181	1404	3504	6902	1981	2145	1139	133	2233	125	19321
Mississippi, USA	3156	8776	3600	3862	11273	3453	3340	1535	526	821	396	22857
New Hampshire, USA	130	875	97	258	357	215	218	84	3	153	7	4677
New Jersey, USA	1357	23857	2434	15841	8605	8802	4047	1986	476	2207	286	24614
New York, USA	3258	23590	5149	13984	14904	6526	7616	4688	523	1343	362	54424
North Carolina, USA	883	4704	2849	3156	4443	1854	1461	939	54	663	87	12633
Rhode Island, USA	1215	10063	2130	5274	7118	2128	3764	1826	263	420	218	24085

Data Analysis - Chelsea

Non-point Pollution direct analysis

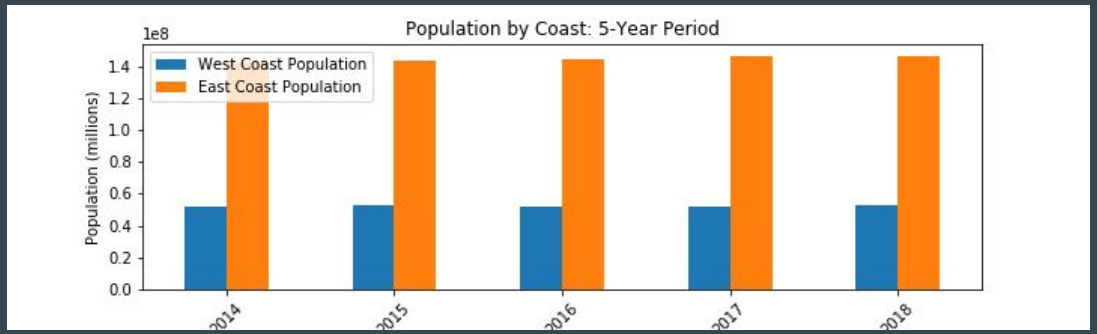


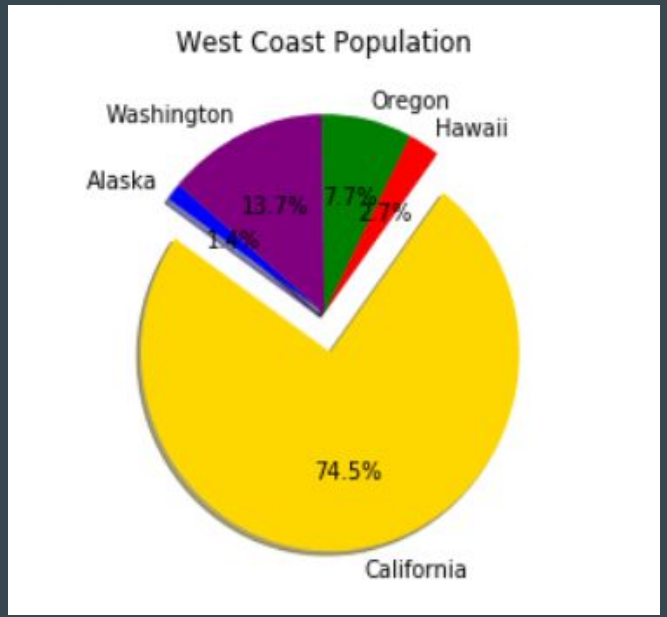

```
#East v West Coast Pollution, GDP
stats.ttest_ind(east_total_litter['Total Pollution'], west_total_litter['Total Pollution'])
# pvalue=0.8510624099227789 not different
```

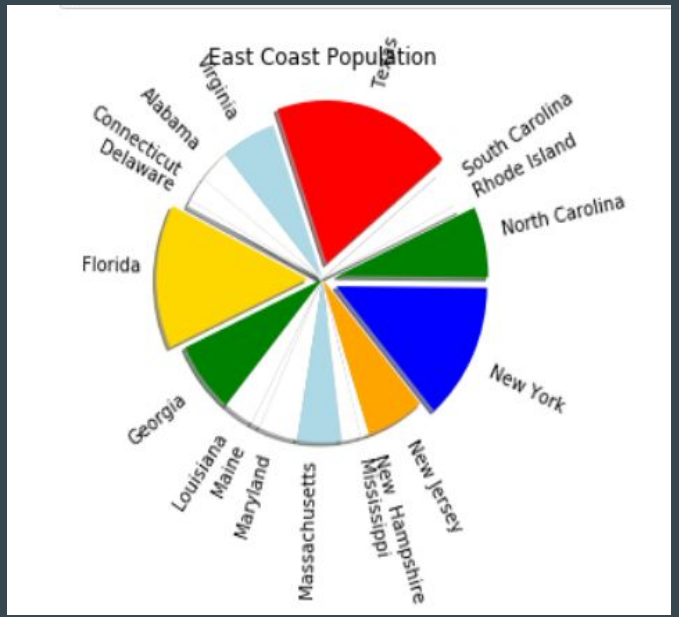


Data Analysis

Greater East Coast population means more pollution?







Data Analysis

High correlation between East Coast Pollution and Population; no correlation on West Coast (more in Data Discussion section!)

```
#Statistics
#West Coast Population vs West Coast Pollution
stats.linregress(west_all['Pollution'], new_total['West Coast Population'])
#rvalue=0.10125963115013017
```

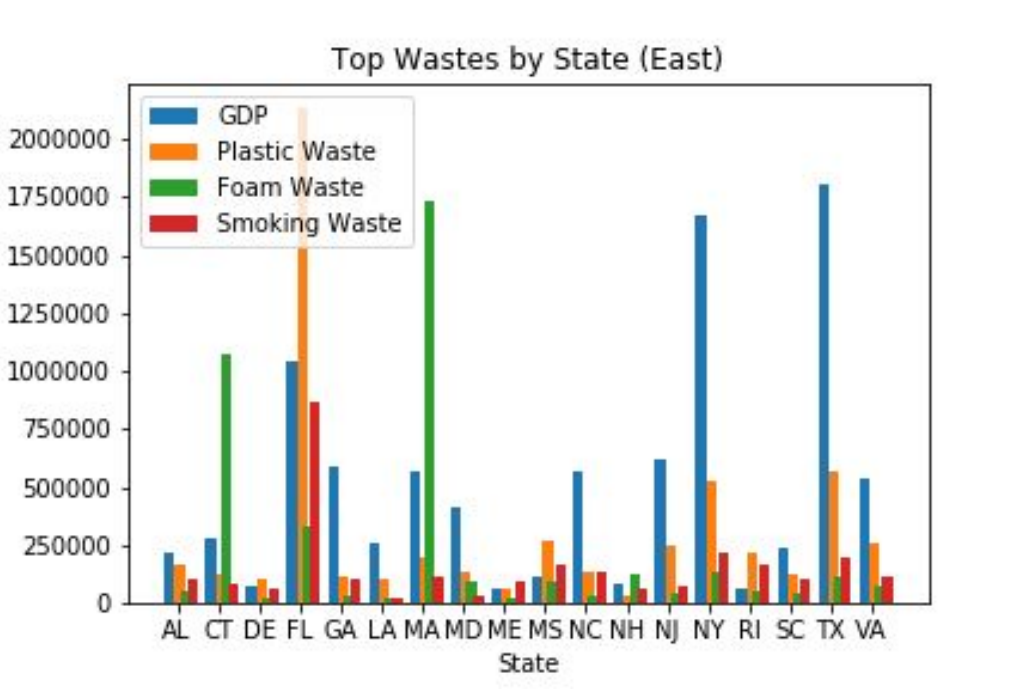
```
LinregressResult(slope=0.05305014099978954, intercept=52239536.920204565, rvalue=0.10125963115013017, pval
ue=0.8712929005369607, stderr=0.30092035021203767)
```

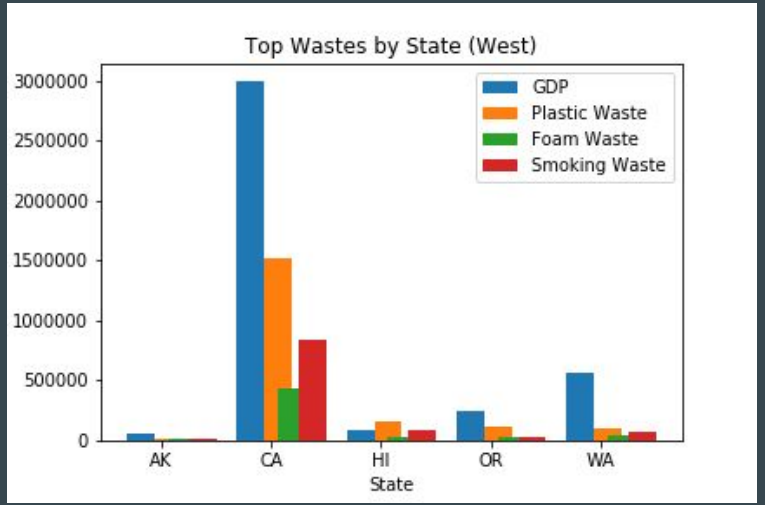
```
#East Coast Population vs East Coast Pollution
stats.linregress(east_all['Pollution'], new_total['East Coast Population'])
#rvalue=0.931577796825101
```

```
LinregressResult(slope=0.8385227414677713, intercept=142793577.41473663, rvalue=0.931577796825101, pvalue=
0.021262837187200612, stderr=0.18892512514674148)
```

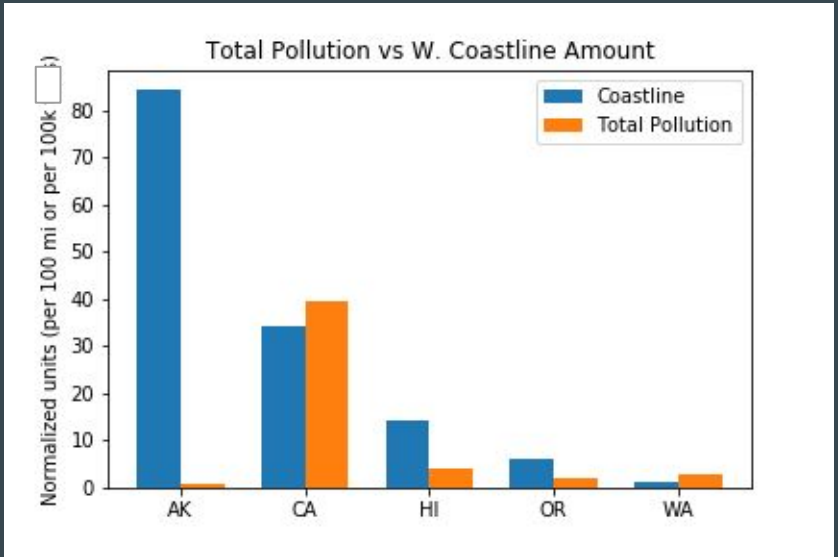
Data Analysis

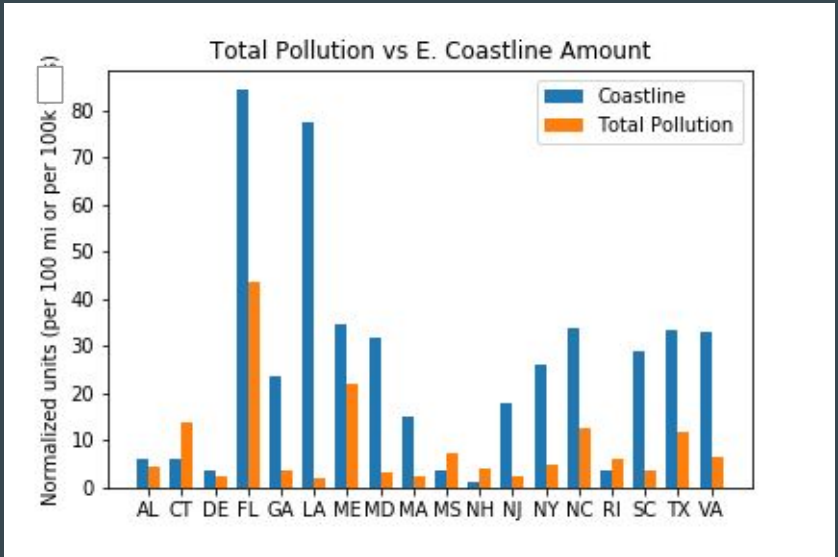
```
#GDP v Total Non-point Pollution
stats.linregress(west_df['2018'], west_total_litter['Total Pollution'])
# rvalue=0.987650265665757 very strong correlation
stats.linregress(east_df['2018'], east_total_litter['Total Pollution'])
# rvalue=0.44829529090626186
```





Data Analysis

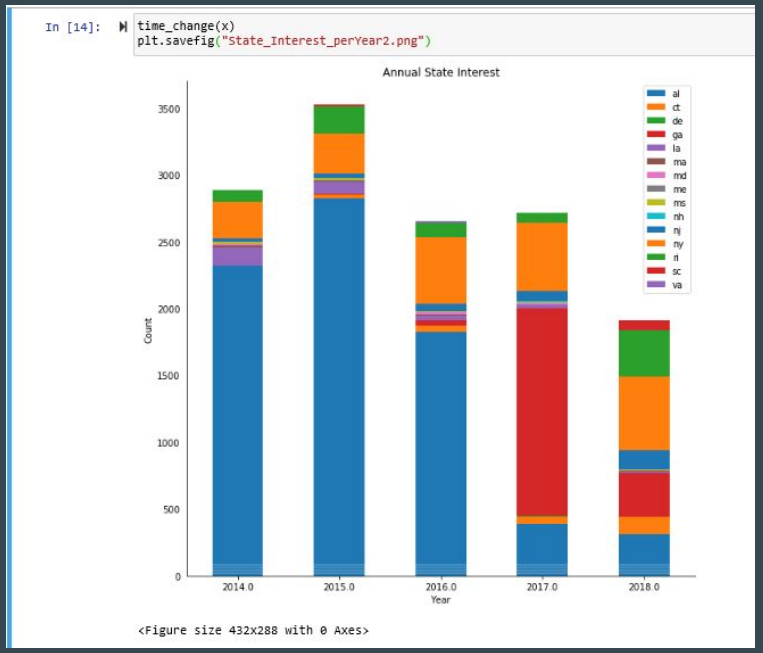


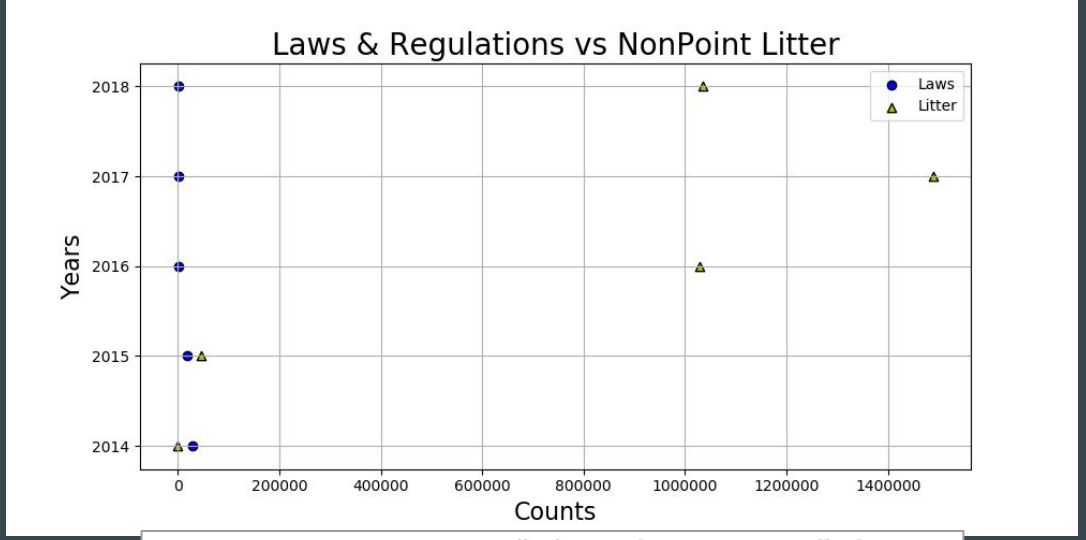


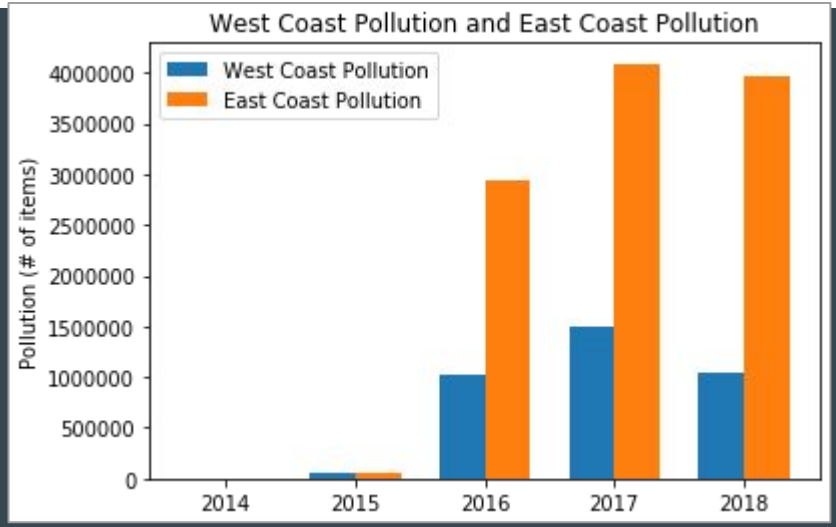

```
#Coastline v Total Non-point Pollution
stats.linregress(west_coast['Coastline'], west_total_litter['Total Pollution'])
# rvalue=-0.25912315944531666
stats.linregress(east_coast['Coastline'], east_total_litter['Total Pollution'])
# rvalue=0.5543644933615807
```

Data Analysis

Laws and Regulations had little to no bearing on pollution.







Discussion - Kaleah

-West versus East

-Correlation Coefficient (r)

-Additional factors

	Pollution vs. Population	GDP vs. Pollution	Pollution vs. Coastline
West	r= 0.10	r= 0.988	r= -0.259
East	r= 0.93	r= 0.45	r= 0.55

Lessons Learned and Future Research - Enjoli

Searching through EPA and BEA/GPI API's

Merging time frames and column names

Dealing with Big Data limit

```
$ git push origin ebaez
Enumerating objects: 122, done./24), 490 MB | 1.2 MB/s
Counting objects: 100% (120/120), done.
Delta compression using up to 4 threads
Compressing objects: 100% (104/104), done.
Writing objects: 100% (104/104), 34.96 MiB | 1.15 MiB/s, done.
Total 104 (delta 46), reused 0 (delta 0)
remote: Resolving deltas: 100% (46/46), completed with 5 local objects.
remote: warning: File Resources/epa_files_v2/TX_SUPP_INTEREST_FILE.CSV is 63.45
MB; this is larger than GitHub's recommended maximum file size of 50.00 MB
remote: warning: File Resources/epa_files_v2/NY_SUPP_INTEREST_FILE.CSV is 65.38
MB; this is larger than GitHub's recommended maximum file size of 50.00 MB
remote: error: GH001: Large files detected. You may want to try Git Large File S
torage - https://git-lfs.github.com.
remote: error: Trace: 87f5bec07eb0eb27d306010a1b900e4c
remote: error: See http://git.io/iEPt8g for more information.
remote: error: File Resources/epa_files_v2/NJ_SUPP_INTEREST_FILE.CSV is 109.27 M
B; this exceeds GitHub's file size limit of 100.00 MB
To https://github.com/deirdrebclark/Project1.git
 ! [remote rejected] ebaez -> ebaez (pre-receive hook declined)
error: failed to push some refs to 'https://github.com/deirdrebclark/Project1.gi
t'
```

Any Questions?