CARBON FOOTPRINTS

Using data to predict
carbon emissions by country

Problem

CARBON DIOXIDE AND METHANE ACCOUNT FOR 92% OF GREENHOUSE GAS EMISSIONS

CARBON DIOXIDE ALONE IS 82% OF CARBON EMISSIONS

INCREASED CARBON EMISSIONS ARE CONTRIBUTING TO AN INCREASE IN GLOBAL TEMPERATURES

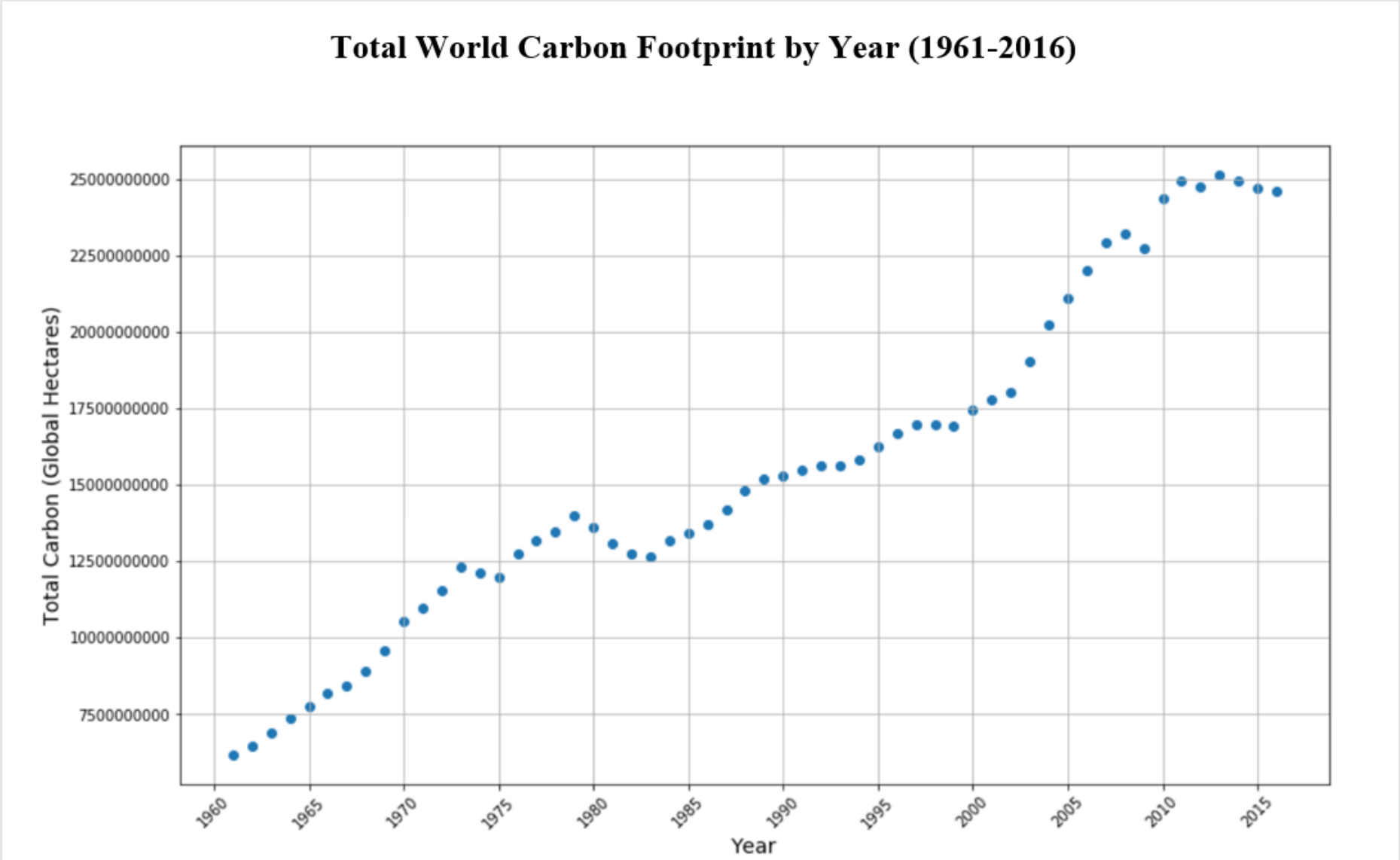
Hypothesis:

- A country's carbon footprint can be predicted by its other footprints (cropland footprint, grazing footprint, forest footprint, built up land footprint, fishing footprint)

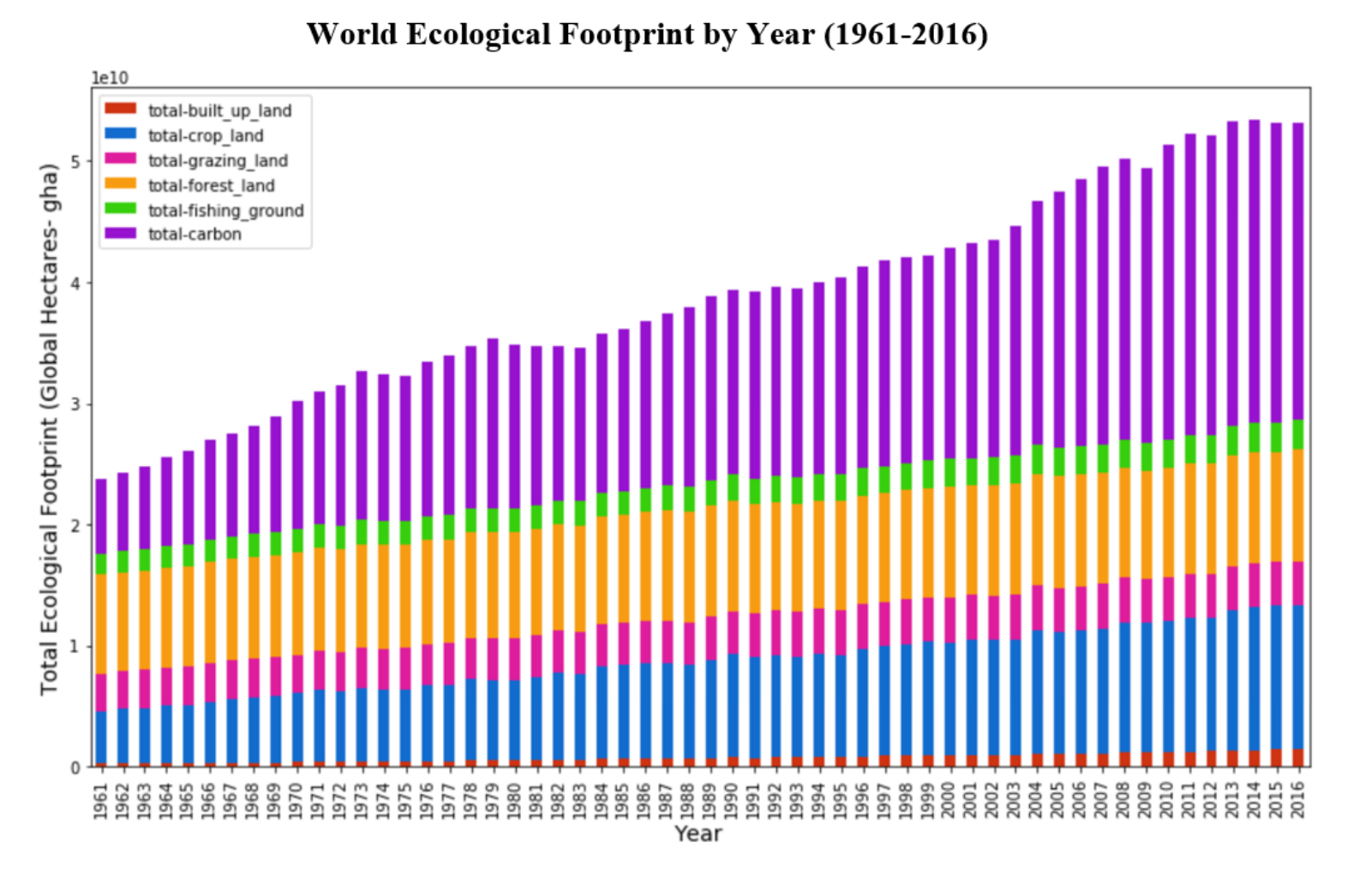
DATA ANALYSIS

What the data tells us

DATA ANALYSIS: WORLD DATA



Total carbon emissions have rapidly increased since 1961

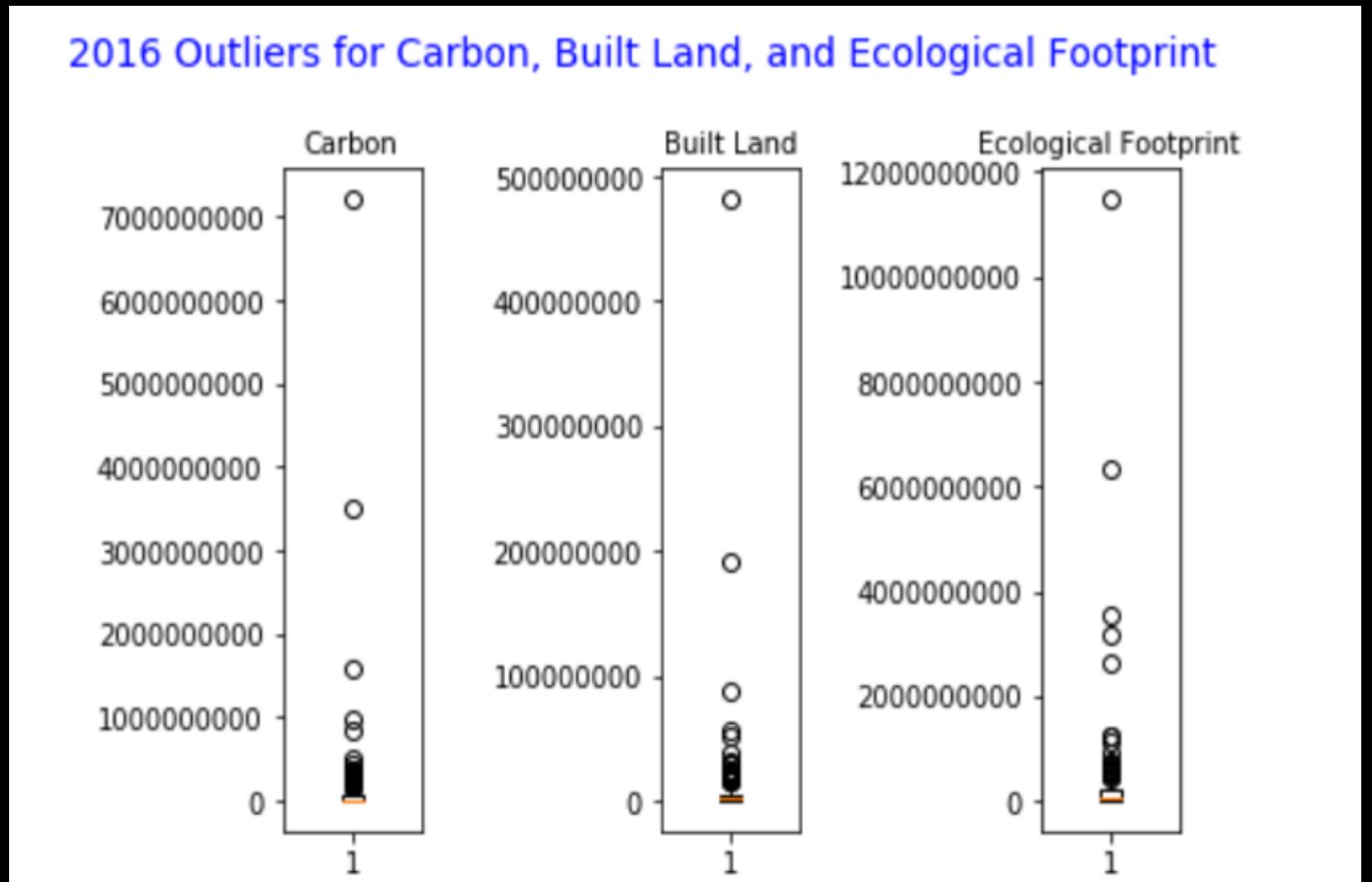


World Ecological Footprint Composition by Year

Note:

- Increase in Carbon
- Increase in Crop Land
- Increase in built up land

DATA ANALYSIS: COUNTRIES

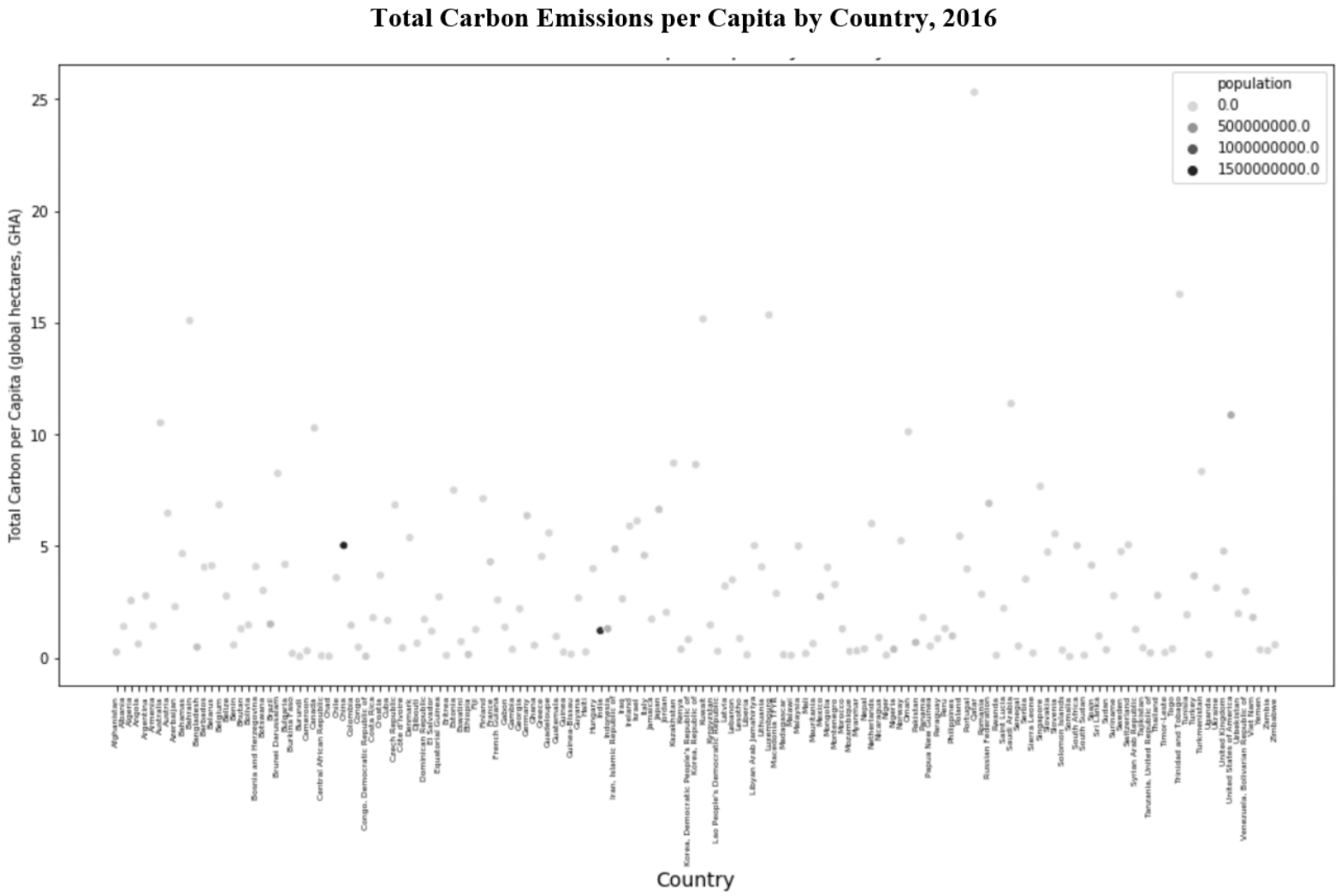


Top country outliers for Carbon, Built up Land, and Ecological Footprints

- Median values are indicated by red line
- Top 3 outliers have extreme footprints

2016 Top 10 Outliers for Carbon Emissions

	country	year	country_code	built_up_land-AreaPerCap	built_up_land-AreaTotHA	built_up_land-BiocapPerCap	built_up_land-BiocapTotGHA	built_up_land-EFConsPerCap	built_up_land-EFConsTotGHA	built_up_land-EFProdPerCap	...
1783	China	2016	351	0.0221	31691300.7800	0.1119	160538830.6000	0.1119	160538830.6000	0.1119	...
8578	United States of America	2016	231	0.0272	8761080.0780	0.0907	29215883.2200	0.0907	29215883.2200	0.0907	...
3955	India	2016	100	0.0187	24802199.2200	0.0479	63470660.5800	0.0479	63470660.5800	0.0479	...
6804	Russian Federation	2016	185	0.0334	4810729.9800	0.0418	6014960.1840	0.0418	6014960.1840	0.0418	...
4376	Japan	2016	110	0.0193	2470149.9020	0.0542	6930153.6600	0.0542	6930153.6600	0.0542	...
3327	Germany	2016	79	0.0369	3020124.8740	0.1328	10878955.9000	0.1328	10878955.9000	0.1328	...
4625	Korea, Republic of	2016	117	0.0148	749236.0229	0.0588	2987266.0920	0.0588	2987266.0920	0.0588	...
4067	Iran, Islamic Republic of	2016	102	0.0415	3334580.0780	0.0700	5617958.8500	0.0700	5617958.8500	0.0700	...
1538	Canada	2016	33	0.0292	1061449.9510	0.0695	2523558.4810	0.0695	2523558.4810	0.0695	...
7120	Saudi Arabia	2016	194	0.0548	1767329.9560	0.0348	1124531.4590	0.0348	1124531.4590	0.0348	...



Darker spots are larger populations, and we find that China is NOT the top outlier when the data is analyzed using per capita calculations

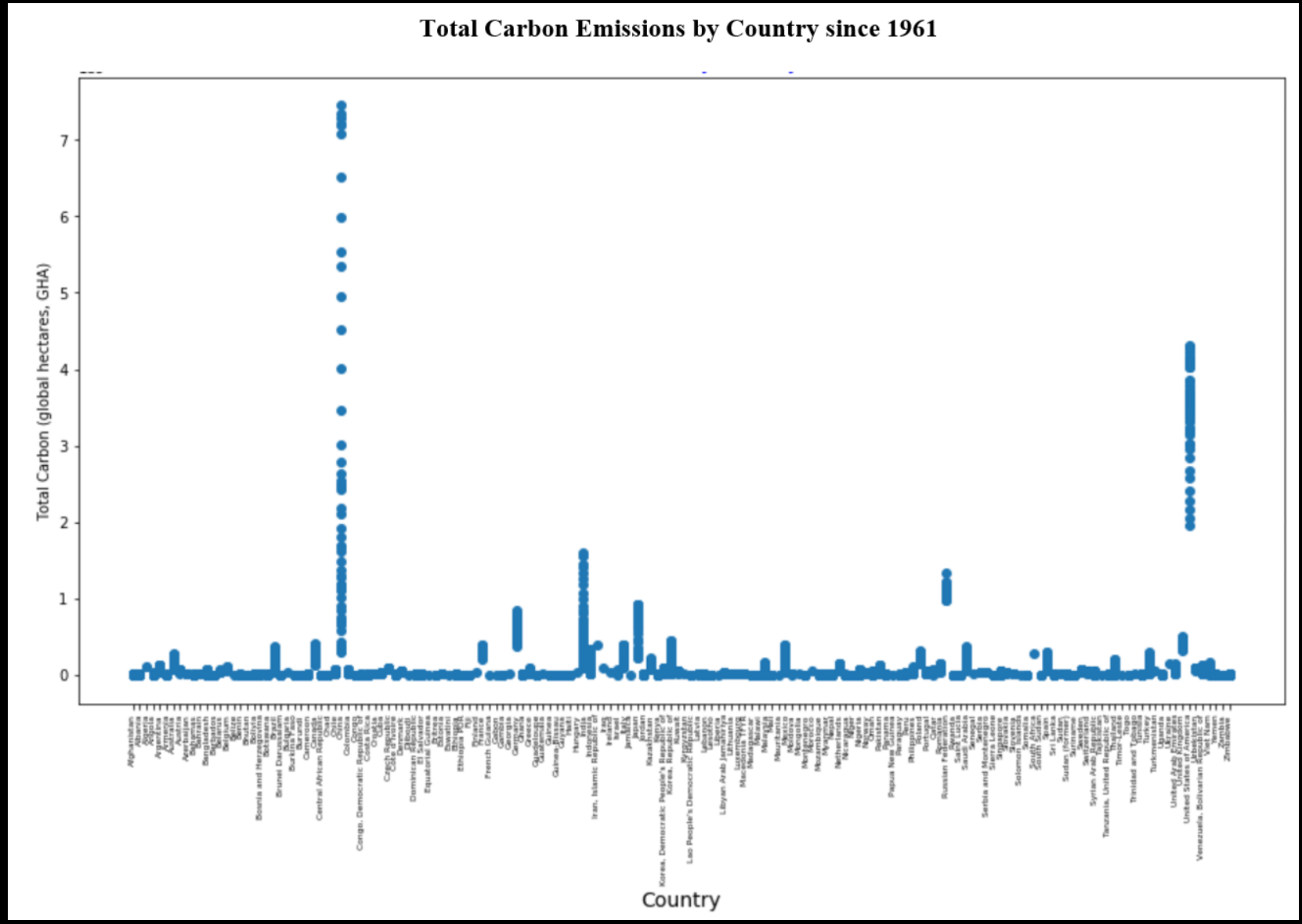
2016 Top 10 Outliers for Per Capita Carbon Emissions

	country	year	country_code	built_up_land-AreaPerCap	built_up_land-AreaTotHA	built_up_land-BiocapPerCap	built_up_land-BiocapTotGHA	built_up_land-EFConsPerCap	built_up_land-EFConsTotGHA	built_up_land-EFProdPerCap	...
6723	Qatar	2016	179	0.0422	108472.9996	0.0461	118584.2115	0.0461	118584.2115	0.0461	...
8180	Trinidad and Tobago	2016	220	0.0009	1291.7800	0.0012	1668.7126	0.0012	1668.7126	0.0012	...
5029	Luxembourg	2016	256	0.0518	29838.9108	0.0817	47037.3901	0.0817	47037.3901	0.0817	...
4643	Kuwait	2016	118	0.0283	114616.9968	0.2501	1013683.8160	0.2501	1013683.8160	0.2501	...
588	Bahrain	2016	13	0.0296	42243.9003	0.0785	111923.9712	0.0785	111923.9712	0.0785	...
7120	Saudi Arabia	2016	194	0.0548	1767329.9560	0.0348	1124531.4590	0.0348	1124531.4590	0.0348	...
8578	United States of America	2016	231	0.0272	8761080.0780	0.0907	29215883.2200	0.0907	29215883.2200	0.0907	...
428	Australia	2016	10	0.0266	642822.0215	0.0589	1422135.1640	0.0589	1422135.1640	0.0589	...
1538	Canada	2016	33	0.0292	1061449.9510	0.0695	2523558.4810	0.0695	2523558.4810	0.0695	...
6238	Oman	2016	221	0.0423	187292.0074	0.1872	828209.6882	0.1872	828209.6882	0.1872	...

10 rows × 74 columns

- Most outliers are very wealthy countries with small to moderately sized populations

- Gaps indicate substantial growth in carbon output
- Countries that had industrial revolutions prior to 1961 started above the zero mark



PREDICTIVE MODELING

Models used

Multiple Linear Regression

Ridge Regression

Random Forest Regression

Multiple Linear Regression

Metric	Value
R^2 Value	1.00
Mean Absolute Error	0.02088
Mean Squared Error	0.00696
Root Mean Squared Error	0.08342

- High R^2 value
- Low errors
- *But* is there a problem?

Variance Inflation Factor

	VIF Factor	features
24	3002399751580330.5000	fishing_ground-AreaTotHA
33	2251799813685248.0000	forest_land-BiocapPerCap
65	643371375338642.2500	total-forest_land_perCap
47	360287970189639.6875	total-AreaPerCap
23	321685687669321.1250	fishing_ground-AreaPerCap
3	300239975158033.0000	built_up_land-AreaTotHA
61	130539119633927.4062	total-carbon_perCap
29	120095990063213.2031	fishing_ground-EFProdPerCap
62	101204486008325.7969	total-EF_perCap
32	94812623734115.7031	forest_land-AreaTotHA
40	84179432287299.0000	grazing_land-AreaTotHA
16	73829502088040.9062	crop_land-AreaTotHA
36	72057594037927.9062	forest_land-EFConsTotGHA
13	70368744177664.0000	carbon-EFConsPerCap
28	66229406284860.2031	fishing_ground-EFConsTotGHA
31	63430980667190.1016	forest_land-AreaPerCap
181	3399992622.2000	country_Papua New Guinea
202	3305598016.6000	country_South Africa
83	3053500075.3000	country_Belize
142	2924628042.7000	country_Jamaica
167	1834672746.1000	country_Mongolia
200	1761577122.2000	country_Solomon Islands
137	1482597234.1000	country_Iran, Islamic Republic of
169	1159226957.3000	country_Morocco
217	950903978.0000	country_Trinidad and Tobago
134	538490760.7000	country_Hungary
121	391036805.3000	country_French Guiana
119	248243195.4000	country_Finland
52	5674.0000	total-EFConsTotGHA
54	4947.2000	total-EFProdTotGHA
55	2304.6000	population
0	3.8000	year

- All VIF values are extremely high, indicating multicollinearity across all features

Ridge Regression

Metric	Value
R^2 Value	0.99
Mean Absolute Error	1.86×10^7
Mean Squared Error	3.55×10^{15}
Root Mean Squared Error	5.96×10^7

- High R^2 value
- High error values
- Not a good model for prediction on the data

Random Forest Regression

Metric	Value
R^2 Value	1.00
Mean Absolute Error	5.43×10^6
Mean Squared Error	1.14×10^{15}
Root Mean Squared Error	3.38×10^7

- High R^2 value
- High error values
- Not a good model for prediction on this data

Conclusion

The features in this data set suffer from a very high degree of multicollinearity despite the use of scaling, PCA, and feature removal

The amount of total carbon produced by a country could not be realistically predicted using this dataset

Recommendations

- Improve prediction by:
  - Including additional outside data
  - Analyzing highest carbon emission countries separately to limit the influence of extreme outliers