Climate Change: A Global Analysis

Asifur Rahman
Kevin Gates
Manuel Moina
Sara Zhou

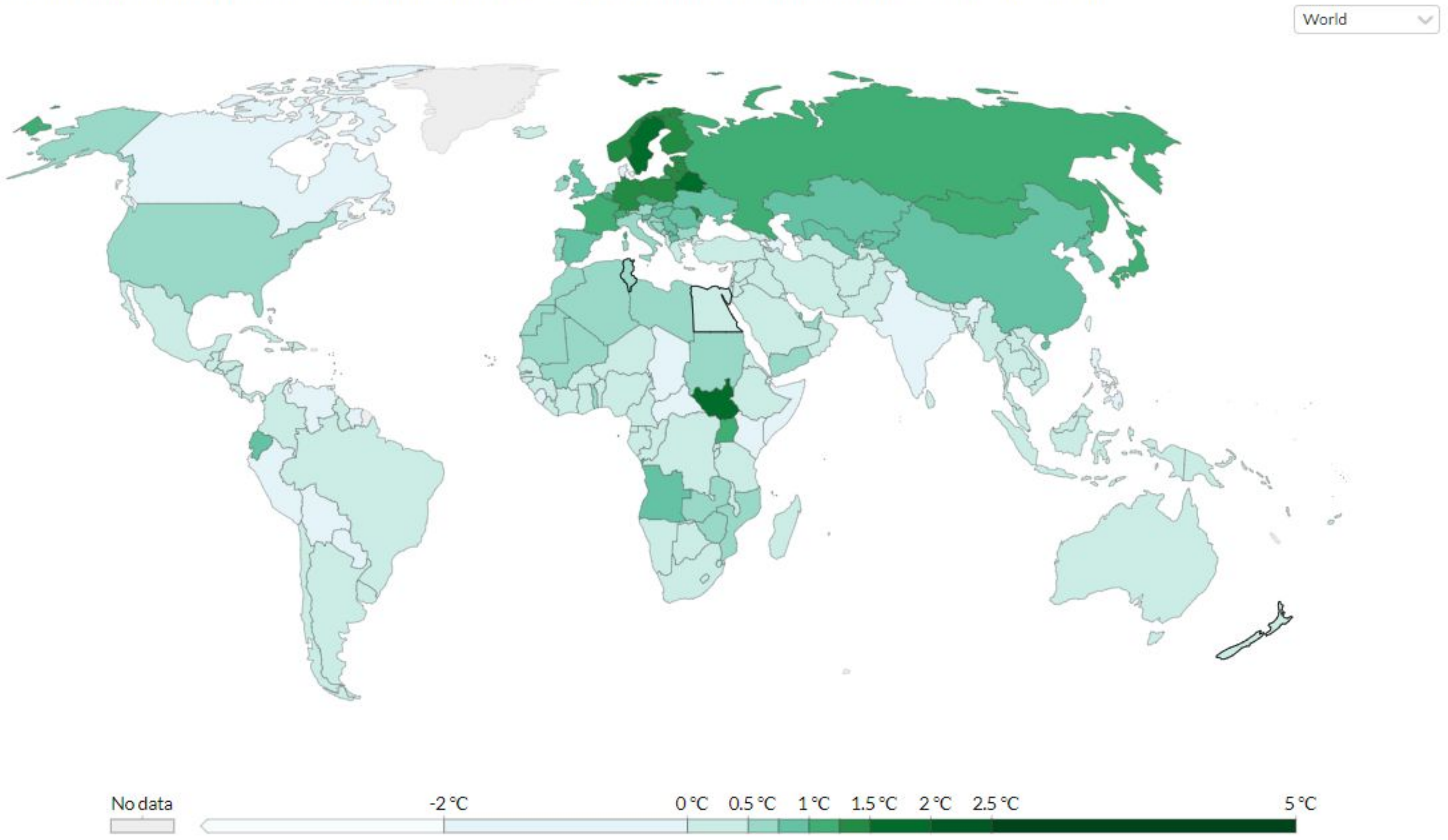
CONTENT

01

Introduction

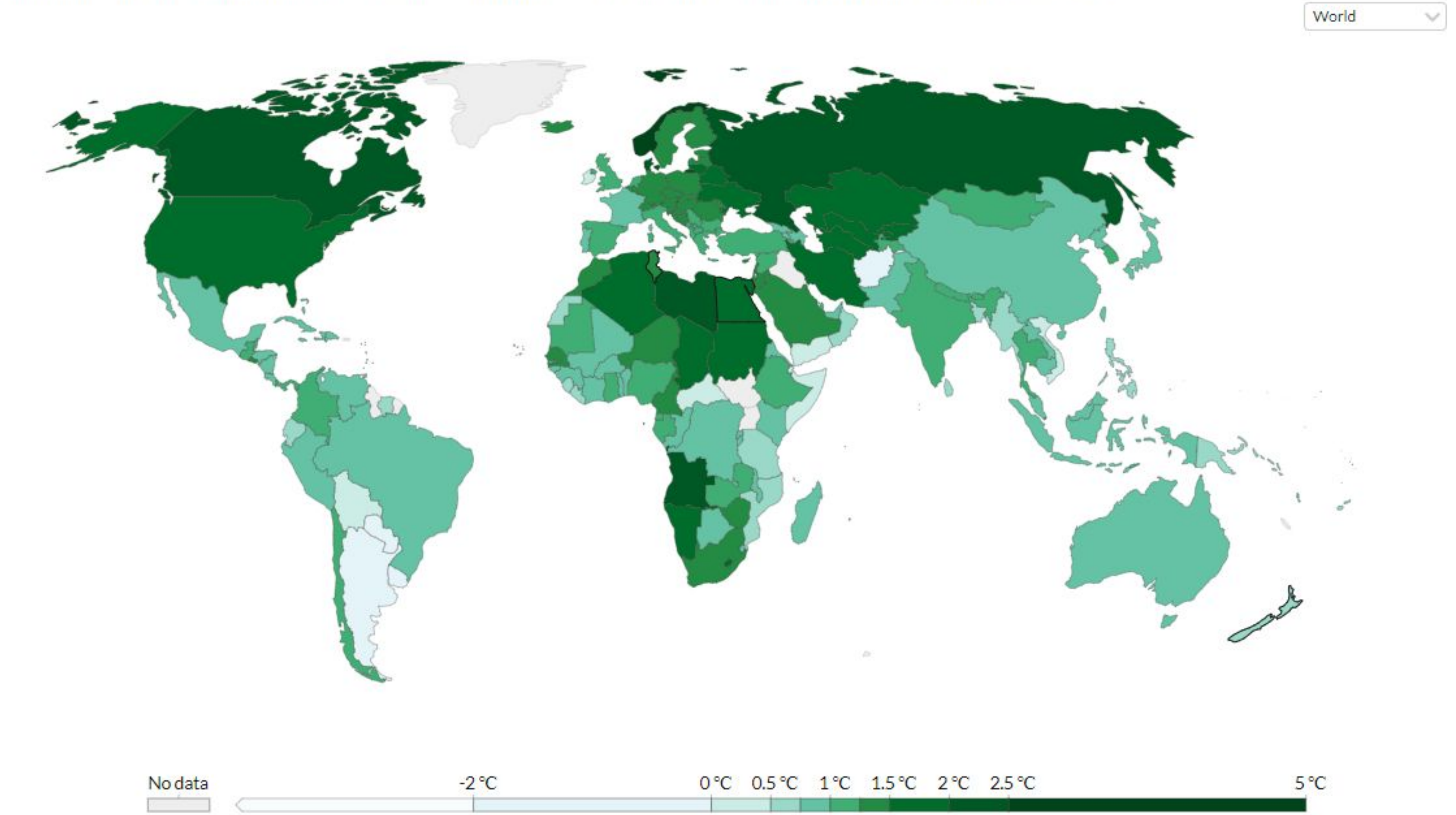
HadCRUT Surface Temperature Anomaly, 1990

Surface temperature anomaly, measured in degrees celcius The temperature anomaly is relative to the 1951-1980 global average temperature. Data is based on the HadCRUT analysis from the Climatic Research Unit (University of East Anglia) in conjunction with the Hadley Centre (UK Met Office).



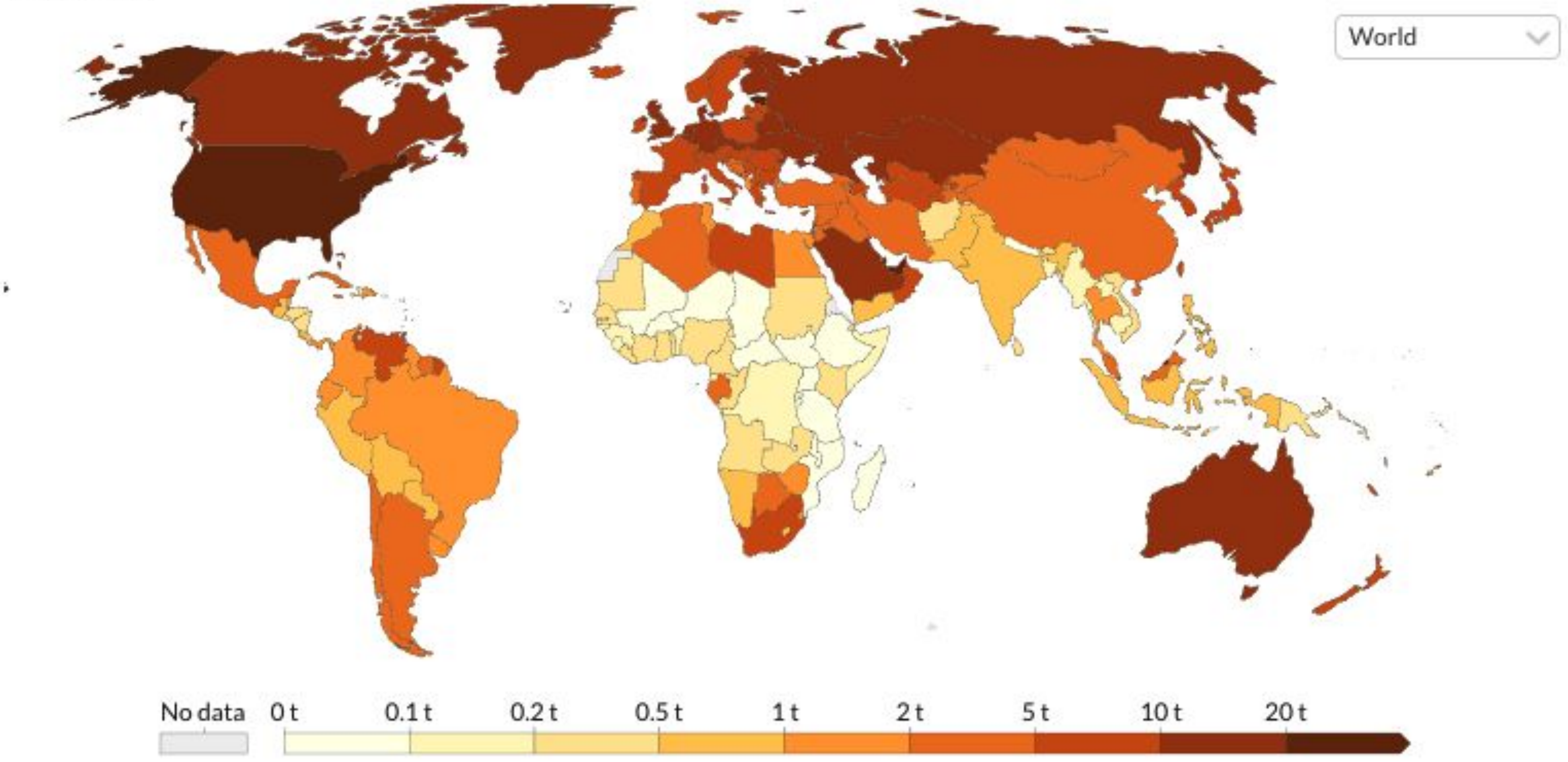
HadCRUT Surface Temperature Anomaly, 2016

Surface temperature anomaly, measured in degrees celcius The temperature anomaly is relative to the 1951-1980 global average temperature. Data is based on the HadCRUT analysis from the Climatic Research Unit (University of East Anglia) in conjunction with the Hadley Centre (UK Met Office).



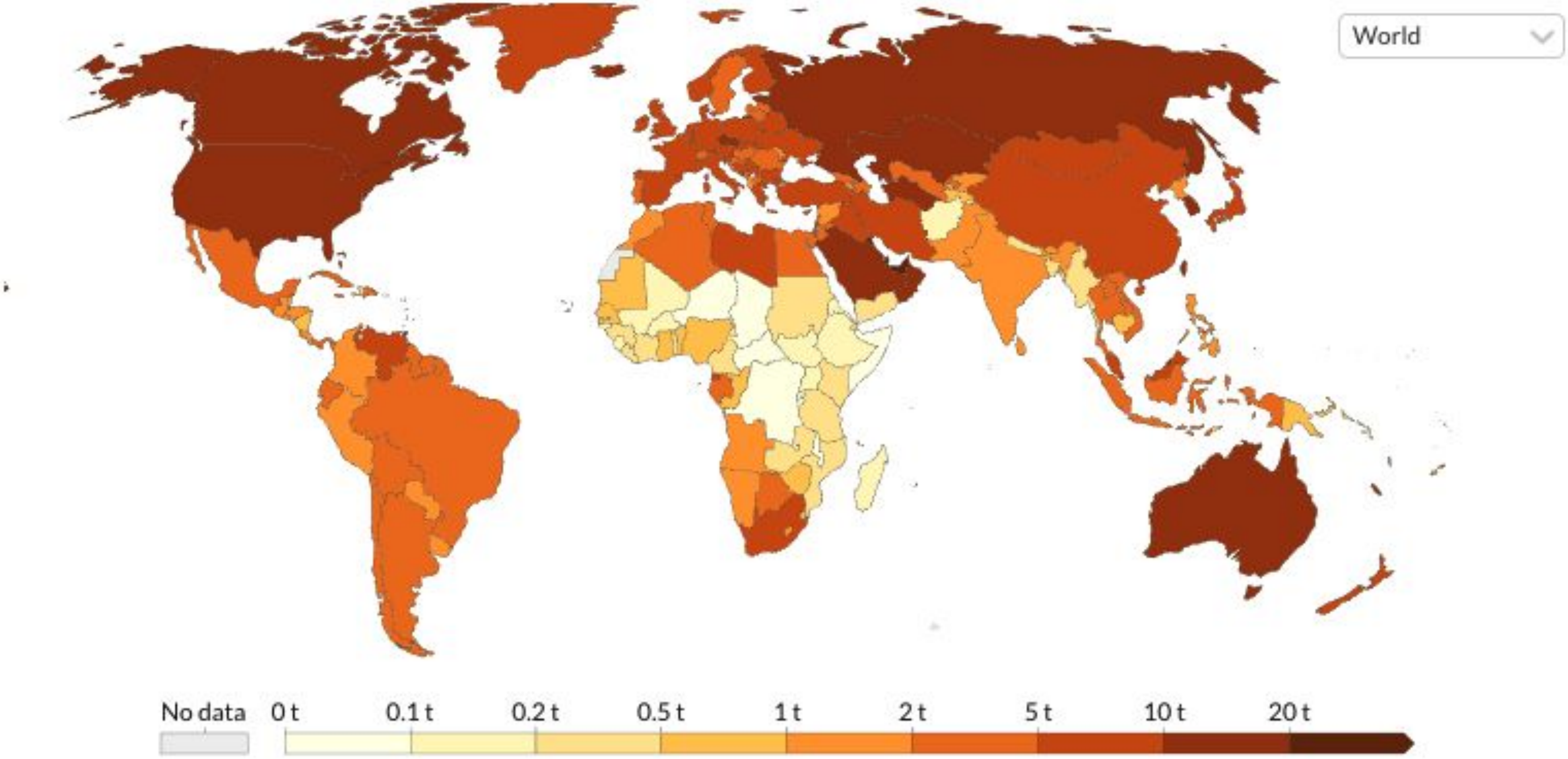
Per capita CO2 emissions, 1990

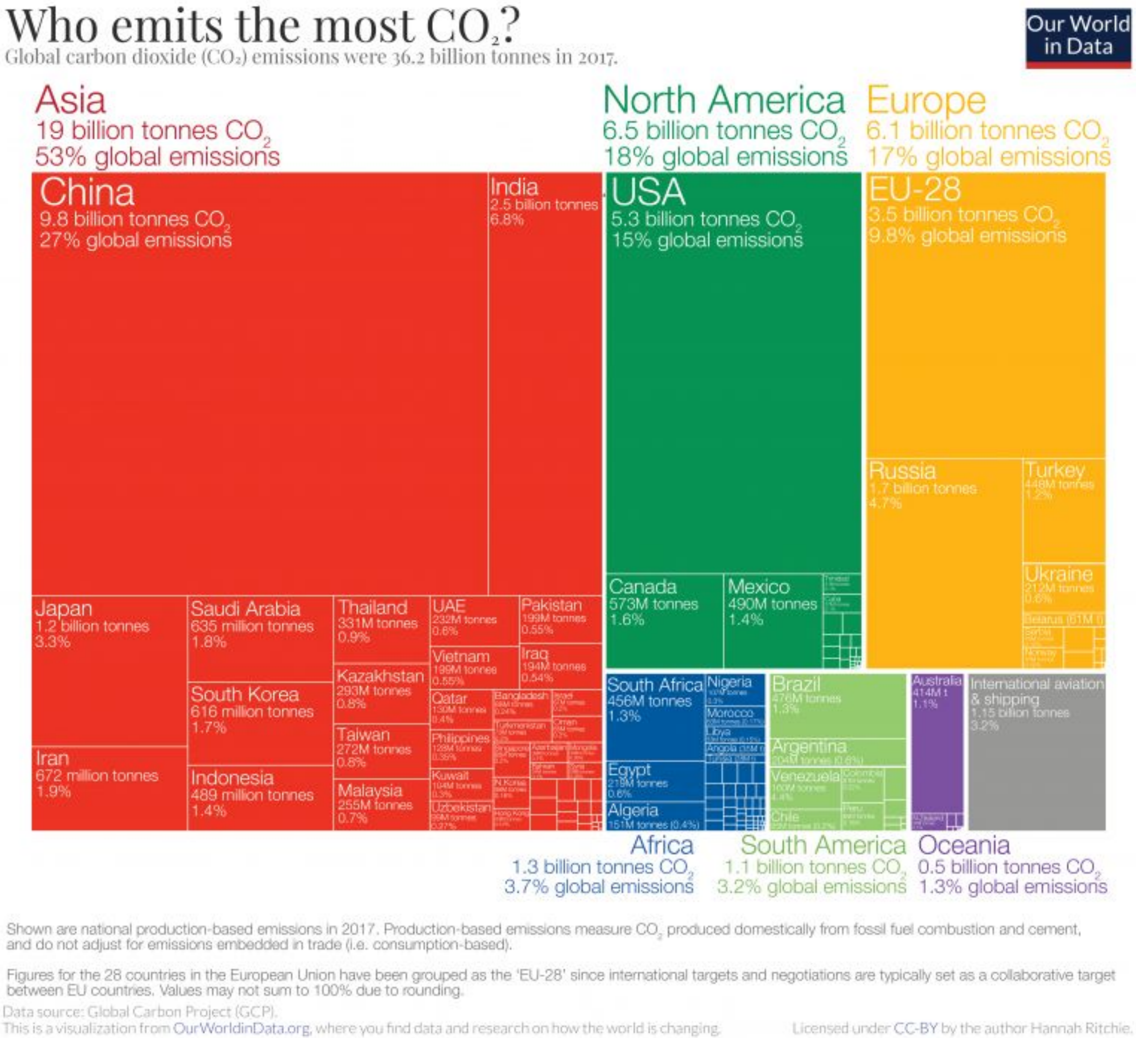
Carbon dioxide (CO_2) emissions from the burning of fossil fuels for energy and cement production. Land use change is not included.



Per capita CO_2 emissions, 2016

Carbon dioxide (CO_2) emissions from the burning of fossil fuels for energy and cement production. Land use change is not included.




Who emits the most CO_2?
Global carbon dioxide (CO_2) emissions were 36.2 billion tonnes in 2017.
Our World in Data
Asia
19 billion tonnes CO_2
53% global emissions
China
9.8 billion tonnes CO_2
27% global emissions
India
2.5 billion tonnes
6.8%
Japan
1.2 billion tonnes
3.3%
Saudi Arabia
635 million tonnes
1.8%
Thailand
331M tonnes
0.9%
UAE
232M tonnes
0.6%
Pakistan
199M tonnes
0.55%
Vietnam
199M tonnes
0.55%
Iraq
194M tonnes
0.54%
Kazakhstan
293M tonnes
0.8%
South Korea
616 million tonnes
1.7%
Qatar
130M tonnes
0.4%
Bangladesh
Taiwan
272M tonnes
0.8%
Philippines
128M tonnes
0.35%
Iran
672 million tonnes
1.9%
Indonesia
489 million tonnes
1.4%
Malaysia
255M tonnes
0.7%
Kuwait
104M tonnes
0.3%
Uzbekistan
North America
6.5 billion tonnes CO_2
18% global emissions
USA
5.3 billion tonnes CO_2
15% global emissions
Canada
573M tonnes
1.6%
Mexico
490M tonnes
1.4%
Europe
6.1 billion tonnes CO_2
17% global emissions
EU-28
3.5 billion tonnes CO_2
9.8% global emissions
Russia
1.7 billion tonnes
4.7%
Turkey
448M tonnes
1.2%
Ukraine
212M tonnes
0.6%
South Africa
456M tonnes
1.3%
Nigeria
Morocco
Libya
Angola
Egypt
219M tonnes
0.6%
Algeria
151M tonnes (0.4%)
Brazil
476M tonnes
1.3%
Argentina
204M tonnes (0.6%)
Venezuela
Colombia
Chile
Peru
Australia
414M t
1.1%
International aviation & shipping
1.15 billion tonnes
3.2%
Africa
1.3 billion tonnes CO_2
3.7% global emissions
South America
1.1 billion tonnes CO_2
3.2% global emissions
Oceania
0.5 billion tonnes CO_2
1.3% global emissions
Shown are national production-based emissions in 2017. Production-based emissions measure CO_2 produced domestically from fossil fuel combustion and cement, and do not adjust for emissions embedded in trade (i.e. consumption-based).
Figures for the 28 countries in the European Union have been grouped as the 'EU-28' since international targets and negotiations are typically set as a collaborative target between EU countries. Values may not sum to 100% due to rounding.
Data source: Global Carbon Project (GCP).
This is a visualization from OurWorldinData.org, where you find data and research on how the world is changing.
Licensed under CC-BY by the author Hannah Ritchie.

Project Goals

Countries contribute to and are affected by climate change differently. In this project we seek to uncover specific global **relationships** between country inputs to climate change (i.e. CO2 emissions) and their effects on a country (i.e. death rates caused by pollution).

02

Data

DATA

FEATURES

Range: 1990-2014

26 Features

COUNTRIES

114 Countries

SOURCE

https://ourworldindata.org/

Merged on Date & Country

FEATURES

CO2, GHG, CO2 Per Cap, Ozone Consumption, Industry Emissions, Forest area

Climate Metrics

Urbanization (% of Pop), Population Density, GDP Per Cap, Commitment to a Net Zero Future

Demographic Metrics

Air Pollution Death Rates, Death Rates By Age, Child Mortality

Outcome Metrics

03: Methods

Our Data Science Process

CLUSTER

With our wide array of interrelated features and our goal of investigating relationships, we knew clustering would be the best technique.

ANALYSIS

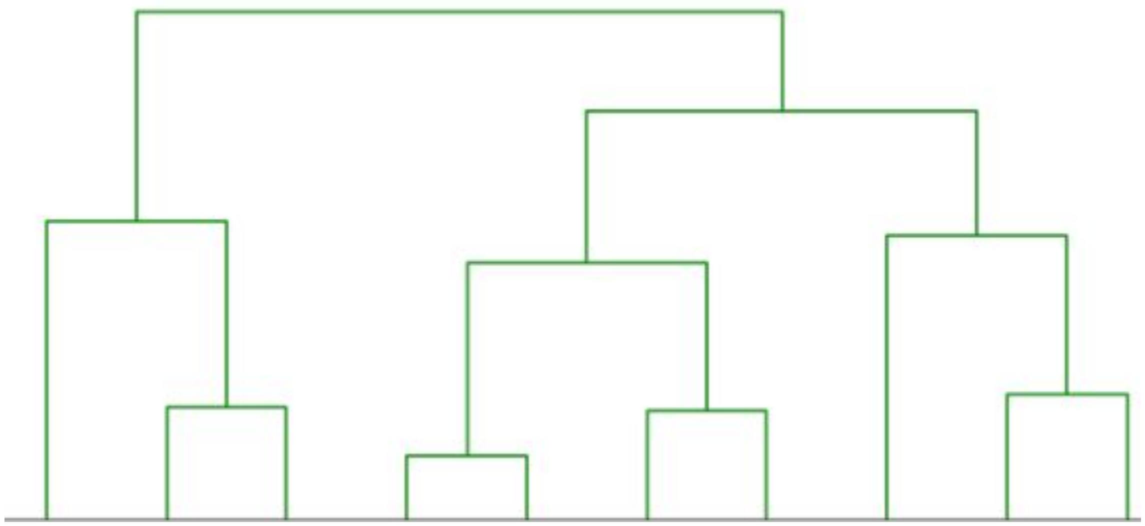
With a clustering problem in hand, our inference can be derived from deep diving into our features and model outputs with EDA.

PREDICT

After thoroughly analyzing our cluster models we utilized predictive modeling techniques to test the validity of relationships we identified.

MODEL SELECTION

1. Hierarchical Clustering
   a. Agglomerative

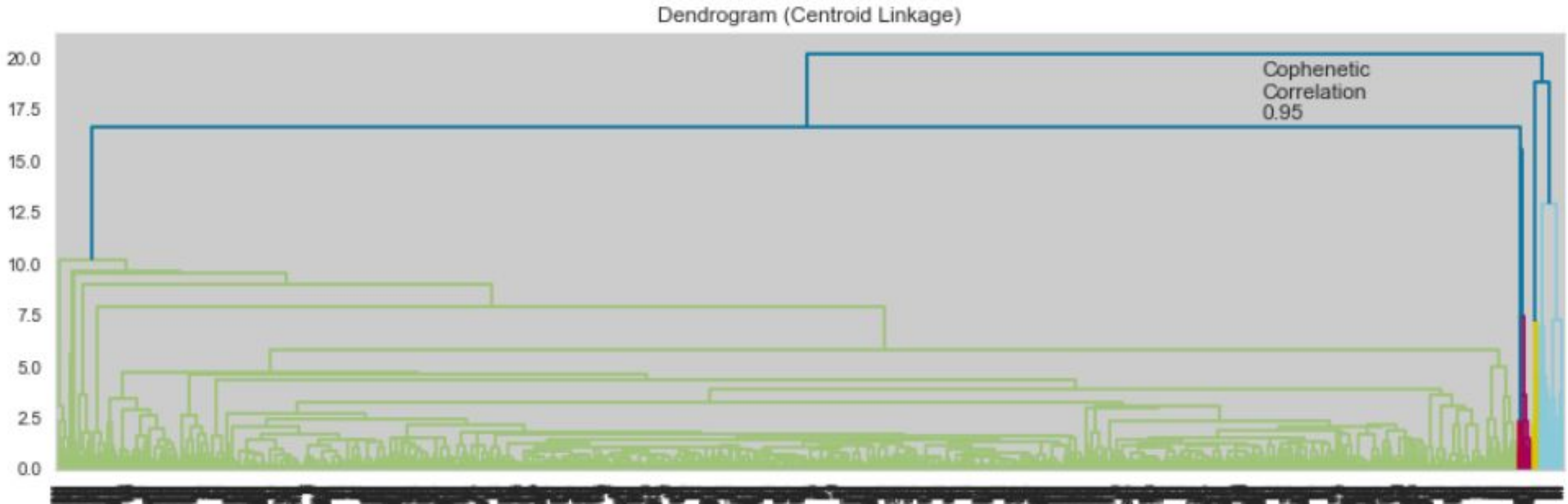


2. DBSCAN
3. KMeans

Hierarchical Clustering

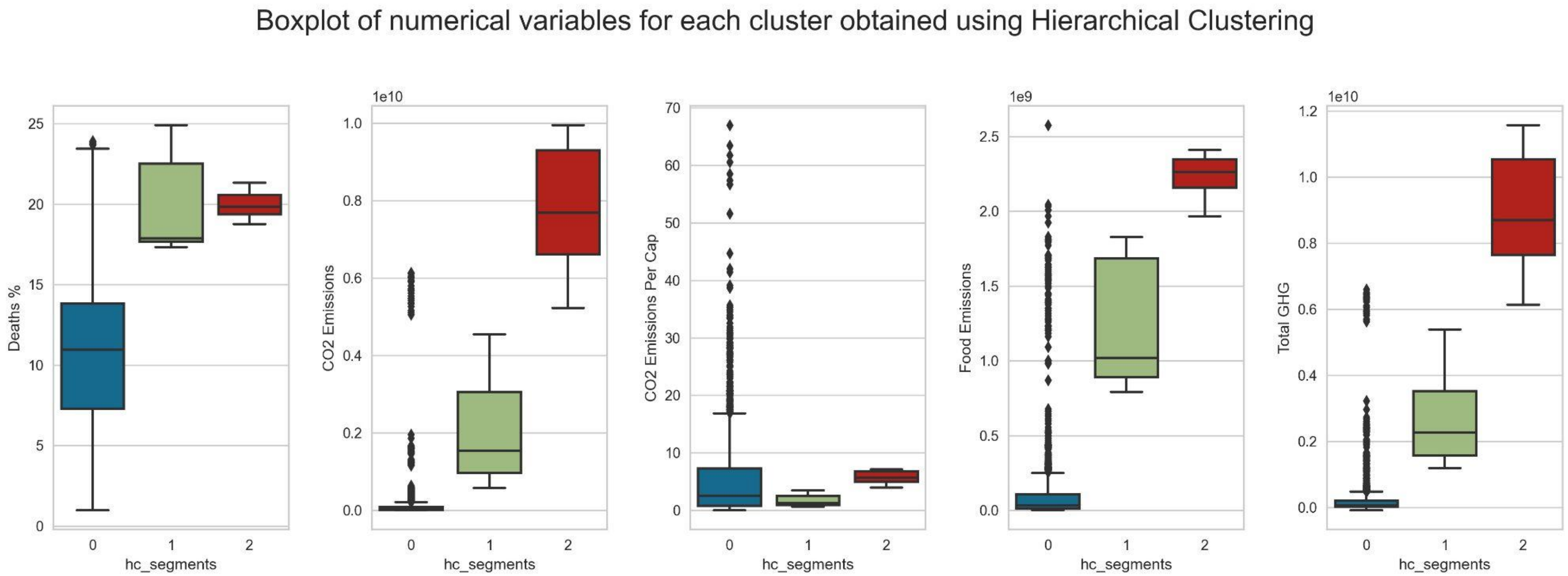
- For hierarchical clustering, the different linkages methods tested are single, complete, weighted, centroid, average, and Ward.
- Eucledian distance with average linkage gave separate and distinct clusters, and also had the highest cophenetic
  correlation (~0.95).

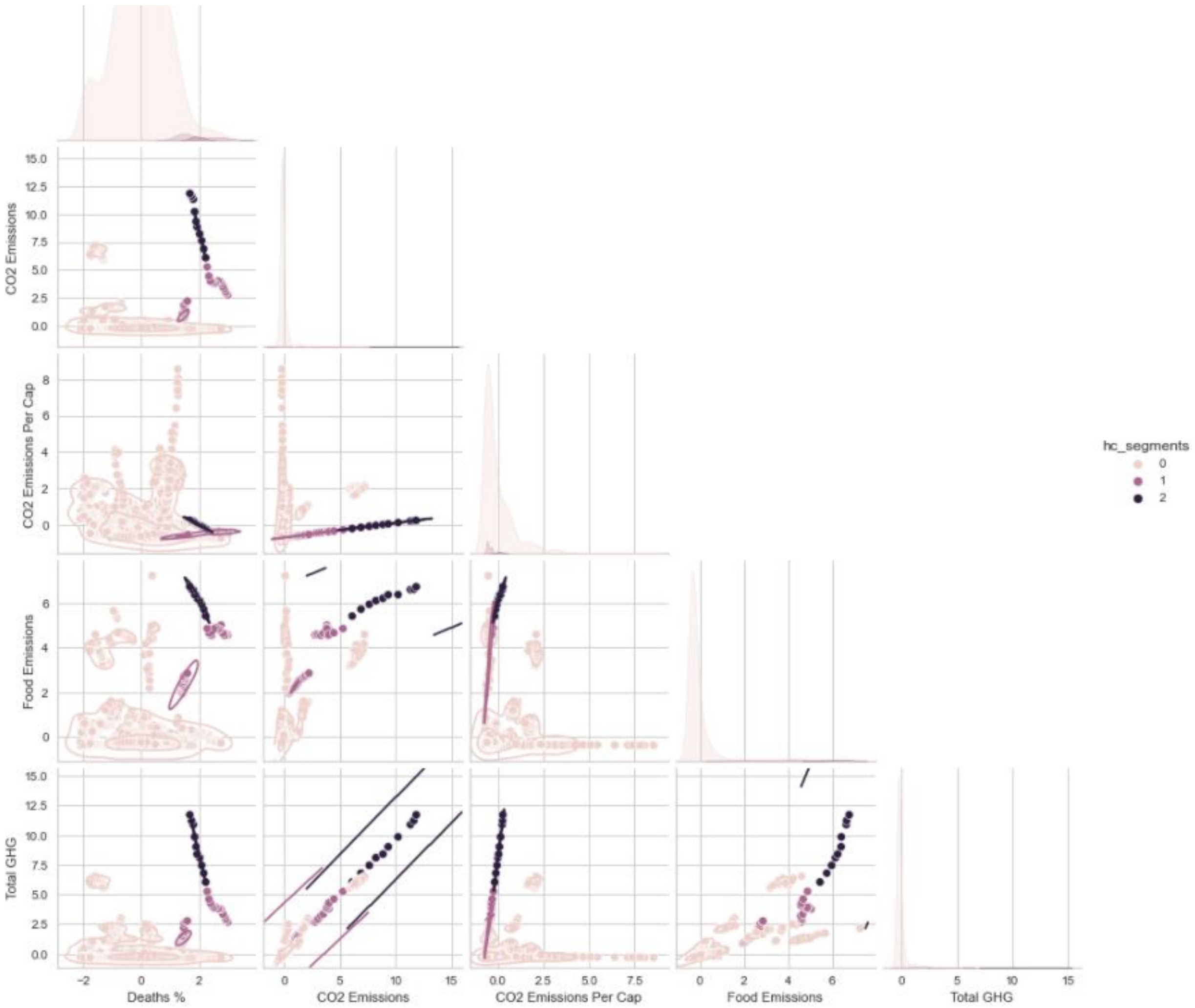
- The dendrogram for average linkage is shown below.



Hierarchical Clustering

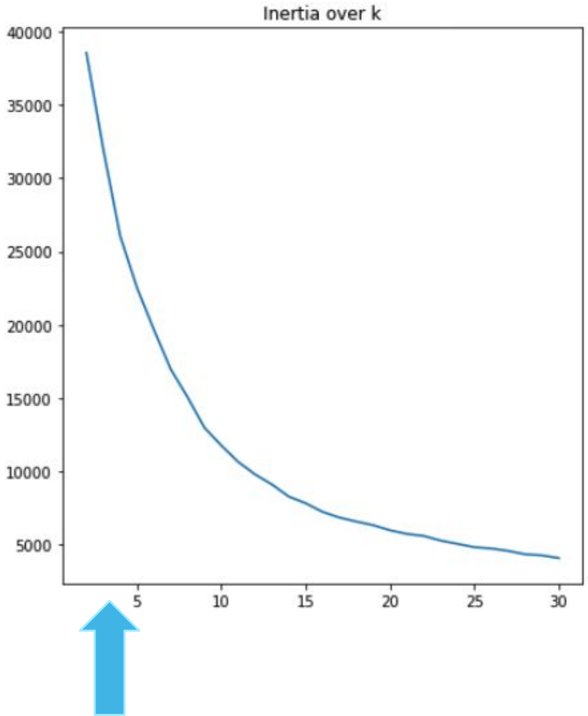
Distribution of the different variables for each cluster.

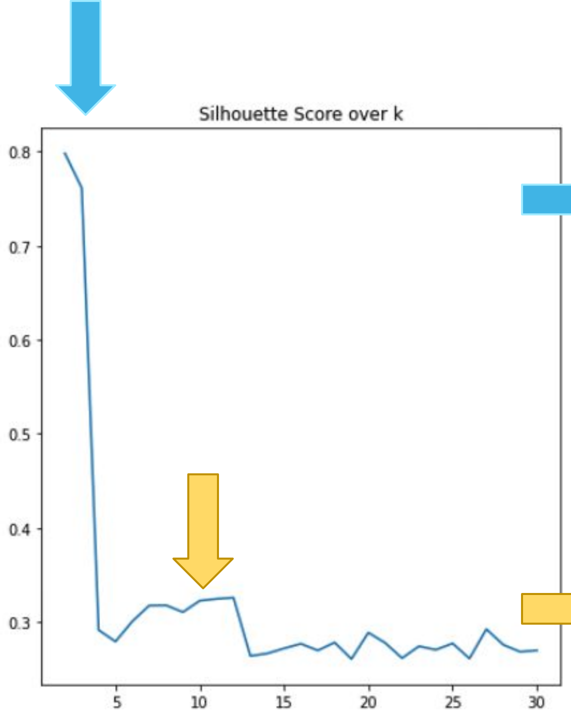




CO2 Emissions
CO2 Emissions Per Cap
Food Emissions
Total GHG
Deaths %
CO2 Emissions
CO2 Emissions Per Cap
Food Emissions
Total GHG
hc_segments
0
1
2

KMEANS MODEL





k	inertia	silhouette
2	38563.314	0.798
3	32001.744	0.761
4	25712.987	0.277
5	22522.354	0.279
6	19631.994	0.301
7	16958.828	0.316
8	14944.614	0.325
9	13832.446	0.326
10	11874.338	0.312
11	10624.942	0.324
12	9809.654	0.271

KMEANS EDA (3 clusters)

cluster	0	1	2
CO2 Emissions	102877714.826	3149207442.771	5664404257.417
CO2 Emissions Per Cap	5.169	2.464	19.947
Total GHG	195452293.539	3915458541.667	6159869583.333
Consumption of Ozone	1187.150	35517.384	47467.957
Urban%	58.067	33.545	78.875
Population	180.856	249.008	31.124
GDP	15806.941	3911.556	49302.893
Forest area	30985970.730	124395229.562	305191875.000
Deaths %	10.697	19.831	3.912
Child Mortality	4.473	5.999	0.858
Death_rate_all_causes	55.409	80.021	20.981
Death_under5	1316.642	60912.312	964.144
Death_5-14	34.251	1765.129	7.433
Death_70+	4986.685	411923.571	61425.546
net_zero_bin	0.344	1.000	1.000

What's Inside?

- Cluster 2: United States
- Cluster 1: China & India
- Cluster 0: ROW

CLUSTERS

Cluster 0

Rest of the world (111)

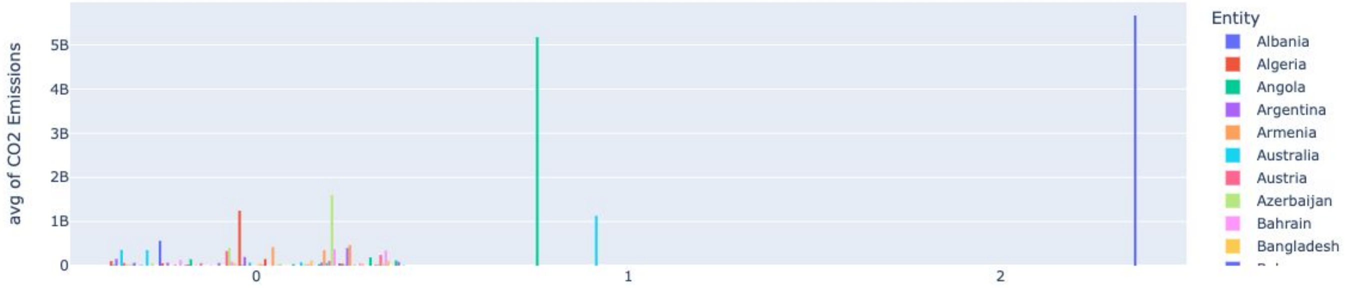
Cluster 1

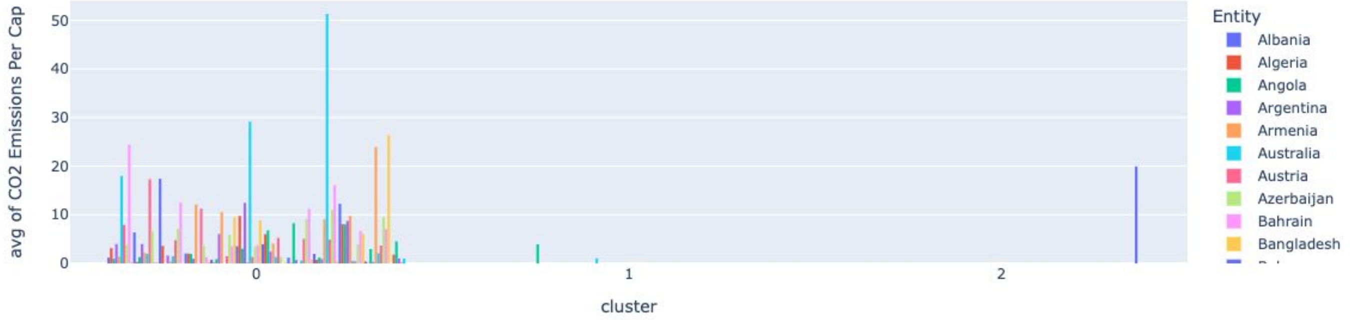
China & India

Cluster 2

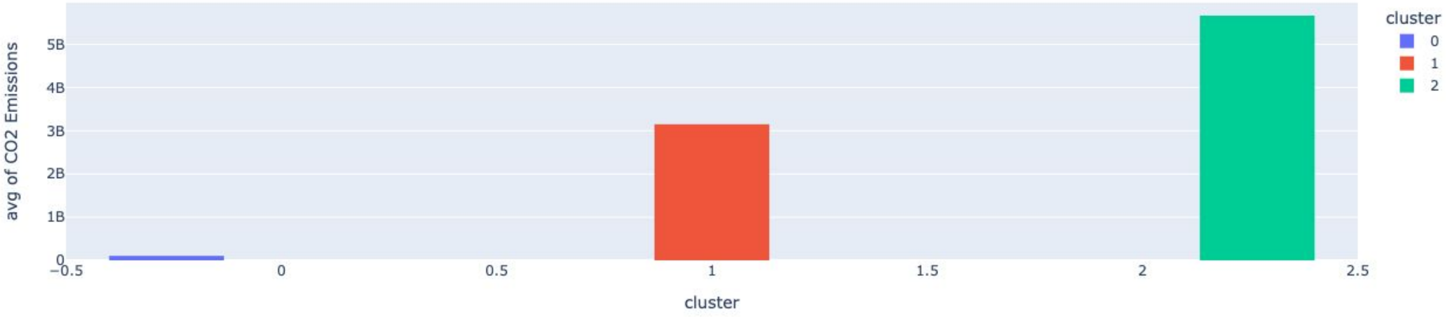
US

KMEANS EDA (3 clusters)

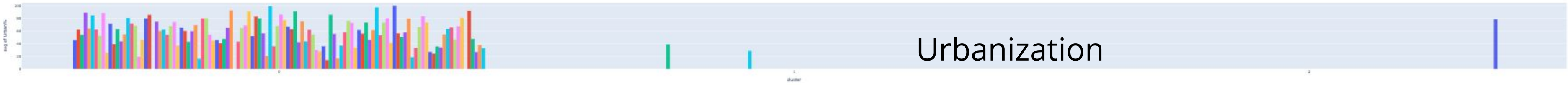


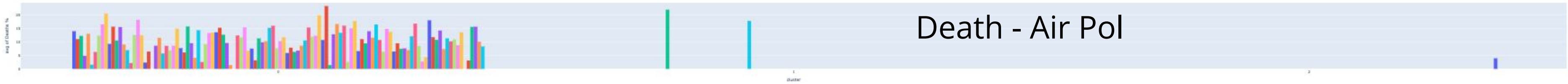


KMEANS EDA (3 clusters)

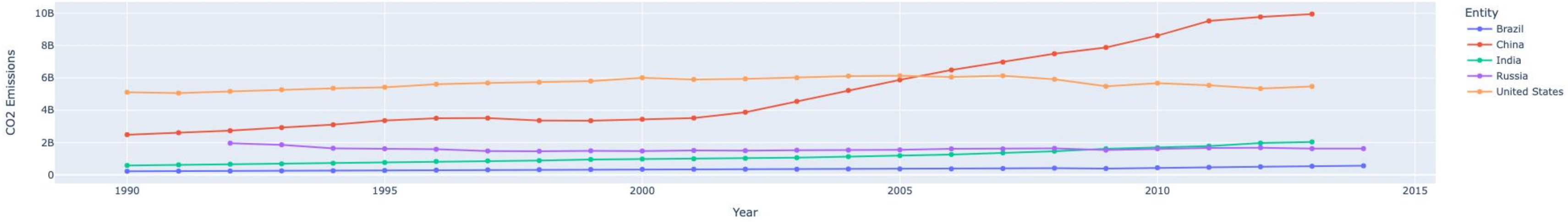


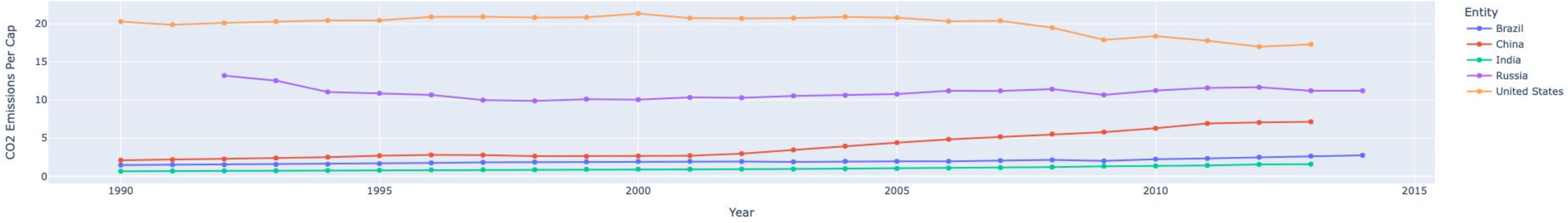
KMEANS EDA (3 clusters)



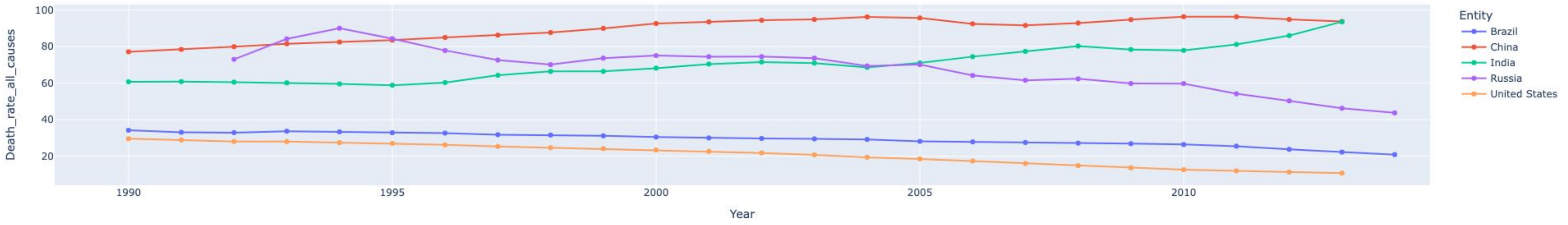


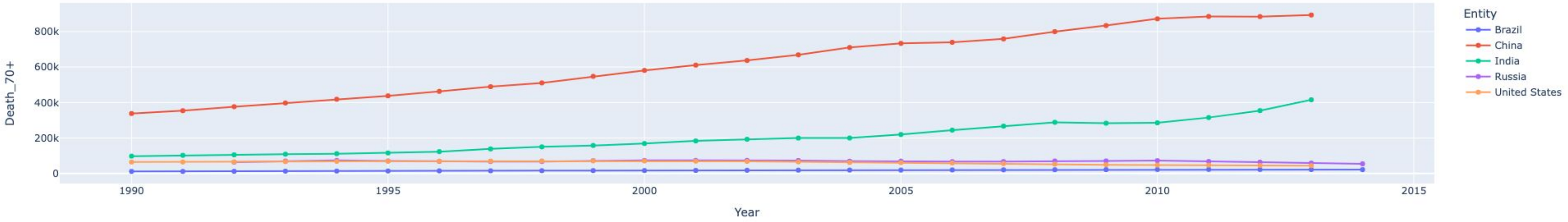
KMEANS EDA (3 clusters)





KMEANS EDA (3 clusters)





Predictions (linear regression)

Predictor	Train	Test
Total GHG	0.994219	0.995571
CO2 Emissions	0.996917	0.994699
Shared CO2 Emissions	0.989121	0.990754
Transport	0.991486	0.98432
Death_actual_household	0.969355	0.974721
clusterx3	0.97311	0.972311
Food Emissions	0.941125	0.934144
Death_rate_household_air_pol	0.926519	0.926603
Deaths %	0.852396	0.852154
Child Mortality	0.827584	0.836838
GDP	0.789815	0.799169
CO2 Emissions Per Cap	0.760061	0.746024
Death_rate_all_causes	0.7395	0.72434
Urban%	0.727791	0.715992
Forest area	0.686289	0.653941
Death_rate_ambient_ozone_pol	0.584691	0.613677
clusterx10	0.590787	0.560587
Consumption of Ozone	0.620467	0.527336
Population	0.222576	0.219751

Classification of Countries

Model	Train	Test
logreg	0.976744	0.970149
rfc	1.000000	0.998342

LIMITATIONS & RECOMMENDATIONS

TIMEFRAME

Data time frame limitation

DATA

Variety & scale mismatch. Limitation of features & countries

MODEL

Inclusion of death rates bake in causation/weigh death; population and country size

SPECIFICITY

Doing an investigation for specific countries or years would tailor highlight particular effects

NEXT STEP

Exclusion of US, China, & India to change the behavior of the clusters; inclusion of recent years

MORE INPUTS

Focus on different streams of inputs (uv rays, water quality, energy consumption, weather anomalies, etc.)

Conclusion

...

Large countries (population & size) contribute most to climate change while not uniformly receiving the worst effects of climate change.

The cluster model groups heavily based on factors interrelated by emissions and population.

Cluster 2 (China and India) and Cluster 3 (United States) are some of the highest emitters on a total CO2 basis. Yet the ROW separates itself from cluster 1 and 2 by having higher emissions on a per capita basis.

Cluster 1 separates itself further also having the highest death rates attributable to an air pollution risk factor, followed by the ROW, while the US, the leading emitter over most of time period seems immune to the consequences looking at death rates.

Check-out Our Streamlit to see for yourself !!

- https://share.streamlit.io/sara-zhou/project-5/main/code/sz/streamlit.py