# Predicting a Country's Greenhouse Gas Emission

Jennifer Sun

Metis, Fall 2020

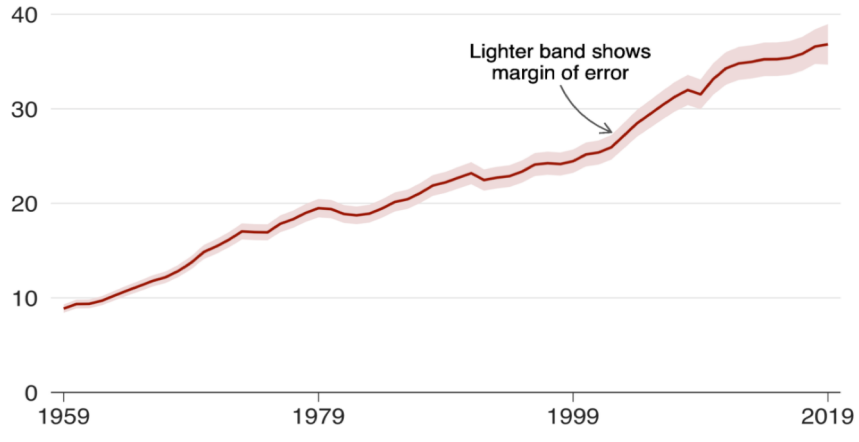
# Is there an easier way to predict a country's greenhouse gas emission?

1. Carbon emissions raising globally
2. Effects of greenhouse gas emission
3. Current greenhouse gas accounting: Complicated!

1

**Global CO2 emissions continue to rise**

Emissions in gigatonnes of carbon dioxide




Source: Global Carbon Project/CICERO BBC

2

3

Some emissions producing activities and methods used to estimate emissions

| Activity | GHG | IPCC tier | Method used to estimate emissions |
|---|---|---|---|
| Public electricity and heat production | $CO_2$ | 2 | An emissions factor is applied to fuel consumption data from DUKES.[1] Some data are also collected from individual point sources at generation facilities. The emissions factors are UK specific factors obtained by sampling average UK carbon content of fuels. |
| Road transportation | $CO_2$, $CH_4$, $N_2O$ | 3 | Emissions from road transport are estimated from a combination of total fuel consumption data taken from the Digest of UK Energy Statistics and fuel properties, and from a combination of drive related emission factors and road traffic data on fuel use, car type, miles driven, road types, and fuel type from DfT. |
| Domestic aviation | $CO_2$, $CH_4$, $N_2O$ | 3 | Data from DfT and CAA on aircraft movements is broken down by aircraft type at each UK airport. The model takes into account the lengths of time spent on different parts of an aircraft's take off and landing cycle and different types of aircraft used in the UK. |

# Web Scraping + Data Preparation

Source: https://ourworldindata.org/countries


Observations: 228 countries and major territories

Target Variable: Annual Greenhouse Gas Emissions 2018

Selected Features – 16, Continuous

- Population density
- Population (Gapminder & UN (1800 to 2019))
- GDP per capita
- Unemployment rate
- Daily supply of calories
- Share of the population living in urban areas
- Share of the population using the Interne
- Share of migrants among the population
- International tourism: Number of arrival
- Beef and buffalo meat consumption per person
- Beef production
- Share of land area used for agriculture
- Forest area (% of land area)
- Share of important terrerstrial biodiversity sites that are protected
- Fossil-fuel subsidies per capita
- Death rates from air pollution

# Designing Regression Model

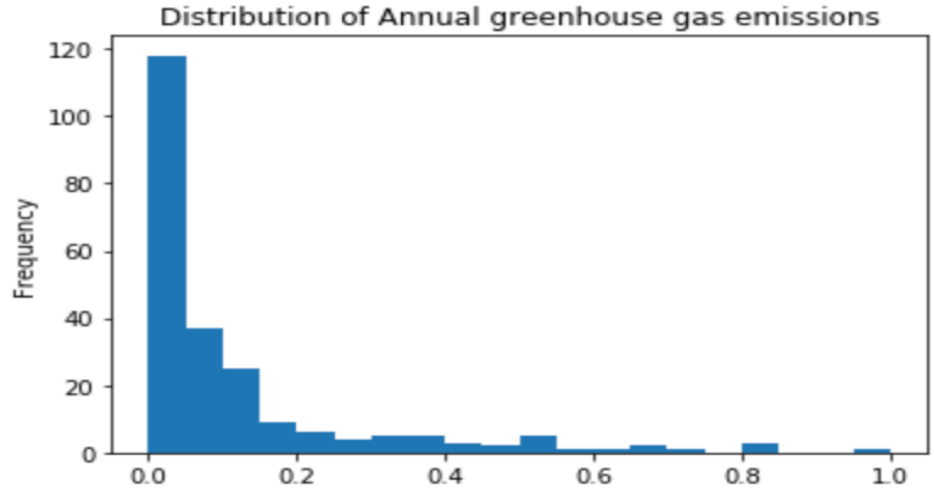
Overfit

**Base Model**
R-squared: 0.228

**Log Target Variable**

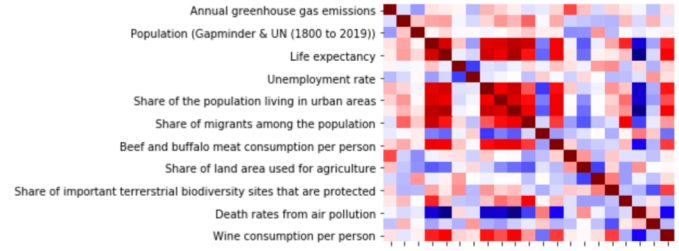
**Single Feature - Beef Production**
R-squared: 0.130

**Feature Selection – 7**
R-squared: .220




| | |
|---|---|
| Annual greenhouse gas emissions | 1.000000 |
| Beef production | 0.366607 |
| Daily supply of calories | 0.143005 |
| Share of land area used for agriculture | 0.123104 |
| Share of the population living in urban areas | 0.091321 |
| Share of important terrerstrial biodiversity sites that are protected | 0.089620 |
| Beef and buffalo meat consumption per person | 0.080377 |
| GDP per capita | 0.062474 |
| Wine consumption per person | 0.055572 |
| Fossil-fuel subsidies per capita | 0.039434 |
| Life expectancy | 0.037771 |
| Share of the population using the Internet | 0.035676 |
| Female employment-to-population ratio | 0.018635 |
| Death rates from air pollution | -0.022853 |
| International tourism: Number of arrivals | -0.046365 |
| Share of migrants among the population | -0.054854 |
| Population density | -0.069963 |
| Forest area (% of land area) | -0.071714 |
| Direct disaster economic loss | -0.124489 |
| Unemployment rate | -0.135822 |
| Population (Gapminder & UN (1800 to 2019)) | -0.200831 |

# Features Selected

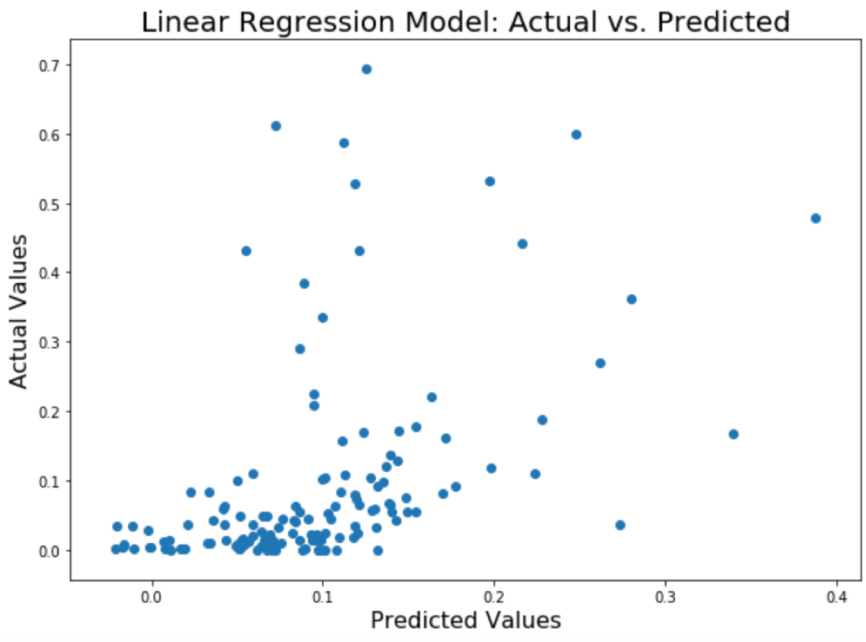


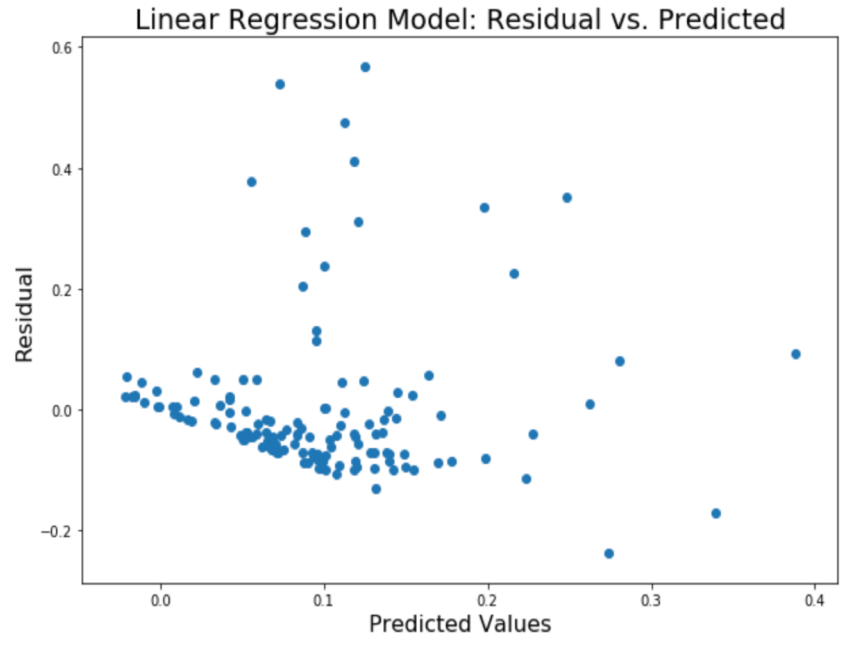

| | Positive Correlation | Negative Correlation |
|---|---|---|
| Strong Correlation | 1. Beef production<br>2. Daily supply of calories<br>3. Share of land area used for agriculture<br>4. Share of the population living in urban areas | 1. Population (Gapminder & UN (1800 to 2019))<br>2. Unemployment rate<br>3. Direct disaster economic loss |
| Weak Correlation | 1. Share of important terrerstrial biodiversity sites that are protected<br>2. Fossil-fuel subsidies per capita<br>3. Beef and buffalo meat consumption per person<br>4. GDP per capita | 1. Death rates from air pollution<br>2. International tourism: Number of arrivals<br>3. Share of migrants among the population<br>4. Population density |

# Evaluating the model

Train Data
R-squared: 0.220

Test Data
R-squared: 0.210



# Key Insights

- Food systems and habits related indicators have a stronger positive correlation with emissions output
- Population and wealth related indicators have a stronger negative correlation with emissions output

| | Positive Correlation | Negative Correlation |
|---|---|---|
| Strong Correlation | 1. Beef production<br>2. Daily supply of calories<br>3. Share of land area used for agriculture<br>4. Share of the population living in urban areas | 1. Population (Gapminder & UN (1800 to 2019))<br>2. Unemployment rate<br>3. Direct disaster economic loss |

# Challenges & Next Steps

- Inconsistent data:
    1. Countries had between 200 – 800 indicators
    2. Smaller, less developed countries tend to have missing data
    3. Many indicators were redundant = needed more domain knowledge
        a) (Meat consumption vs. GDP per capita, Meat production, Meat production by livestock type)
- Expand analysis across historical data instead of just most recent (2018)

# Questions?

Additional Resources:

extinction rebellion