

Fueling the Footprint:

Exploring How Parent Entity and Commodity
Types Impact Carbon Emissions

Cailey Plainte

Introduction

- As climate change continues to worsen, it's more important than ever to understand the factors that drive carbon emissions.
 - A lot of attention is focused on what companies or sectors produce the most carbon emissions
 - Not as much focus is put on exploring the patterns behind these large emission producers
- This project aims to use historical carbon emissions data from fossil fuel producers to explore whether a company's parent entity type and the type of commodity it produces can be used to predict whether its annual carbon emissions will fall above or below the industry average for that year.

The Data

- Data was generated using data from the group Climate Majors.
 - This group compiled a dataset of historical carbon emissions data spanning from 1854 to 2022 and including 180 of the world's largest fossil fuel producers.
- Provides information on year, parent entity, parent entity type, commodity produced, quantity of production, the units that production is measured in for the specified commodity, and the total amount of carbon emissions resulting from that commodity production, measured in metric tons.
- While this data does not cover all emissions globally or all sectors, Climate Majors claims that it covers 72% of all carbon emissions since the start of the Industrial Revolution, making it a relatively comprehensive dataset.

The Data

	year	parent_entity	parent_type	commodity	production_value	production_unit	total_emissions_MtCO2e
11851	1854	Westmoreland Mining	Investor-owned Company	Lignite Coal	0.011223970	Million tonnes/yr	0.01503125
11852	1854	Westmoreland Mining	Investor-owned Company	Sub-Bituminous Coal	0.034249083	Million tonnes/yr	0.06915200
11853	1854	Westmoreland Mining	Investor-owned Company	Bituminous Coal	0.005532118	Million tonnes/yr	0.01501494
11854	1855	Westmoreland Mining	Investor-owned Company	Lignite Coal	0.014595473	Million tonnes/yr	0.01954641
11855	1855	Westmoreland Mining	Investor-owned Company	Sub-Bituminous Coal	0.044536965	Million tonnes/yr	0.08992416
11856	1855	Westmoreland Mining	Investor-owned Company	Bituminous Coal	0.007193879	Million tonnes/yr	0.01952519
11857	1856	Westmoreland Mining	Investor-owned Company	Lignite Coal	0.017966976	Million tonnes/yr	0.02406156
11858	1856	Westmoreland Mining	Investor-owned Company	Sub-Bituminous Coal	0.054824848	Million tonnes/yr	0.11069633
11859	1856	Westmoreland Mining	Investor-owned Company	Bituminous Coal	0.008855640	Million tonnes/yr	0.02403544
11860	1857	Westmoreland Mining	Investor-owned Company	Lignite Coal	0.020884691	Million tonnes/yr	0.02796899
11861	1857	Westmoreland Mining	Investor-owned Company	Sub-Bituminous Coal	0.063728029	Million tonnes/yr	0.12867265
11862	1857	Westmoreland Mining	Investor-owned Company	Bituminous Coal	0.010293735	Million tonnes/yr	0.02793863
11863	1858	Westmoreland Mining	Investor-owned Company	Lignite Coal	0.023802405	Million tonnes/yr	0.03187642
11864	1858	Westmoreland Mining	Investor-owned Company	Sub-Bituminous Coal	0.072631209	Million tonnes/yr	0.14664898

Data Wrangling

- Narrow down data to relevant variables
- Find average emissions from production each year
- Utilize if else statements to determine if an observation is above or below the average for that year

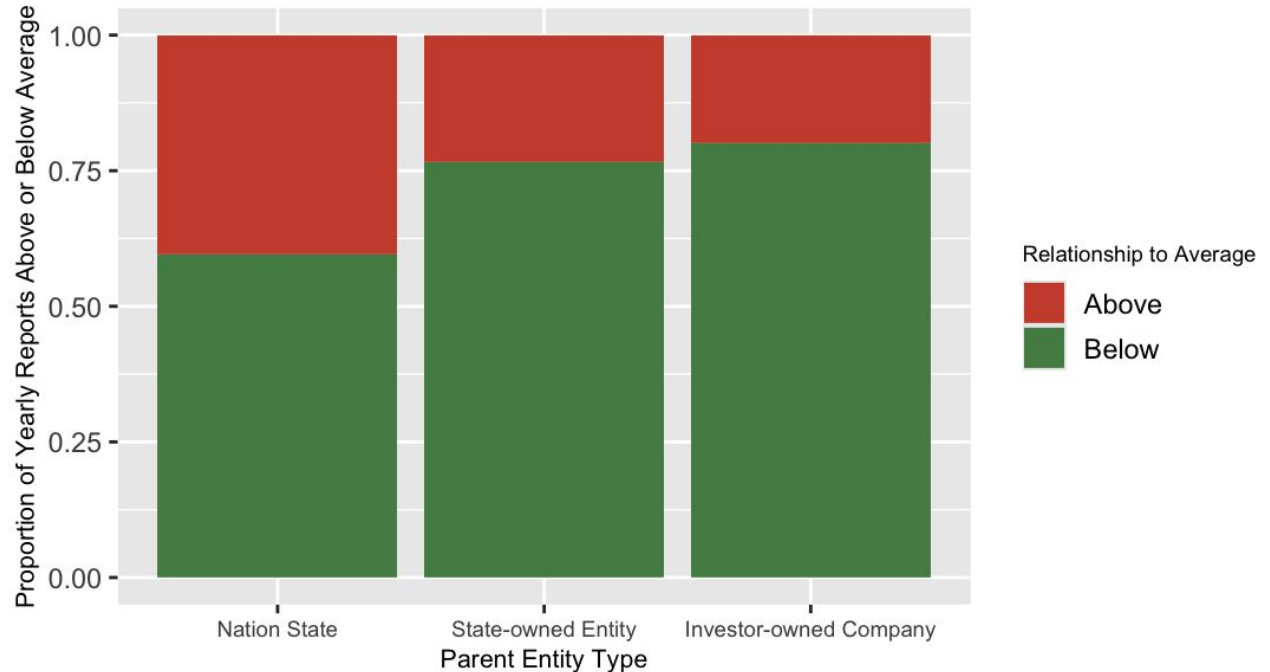
```
```{r message=FALSE, warning=FALSE, include=FALSE}
emissions <- emissions |>
 select(year, parent_type, total_emissions_MtCO2e, commodity) |>
 group_by(year) |>
 mutate(
 year_avg = mean(total_emissions_MtCO2e, na.rm = TRUE),
 above_average = ifelse(total_emissions_MtCO2e > year_avg, "Above",
"Below")) |>
 ungroup()
```
```

The Data

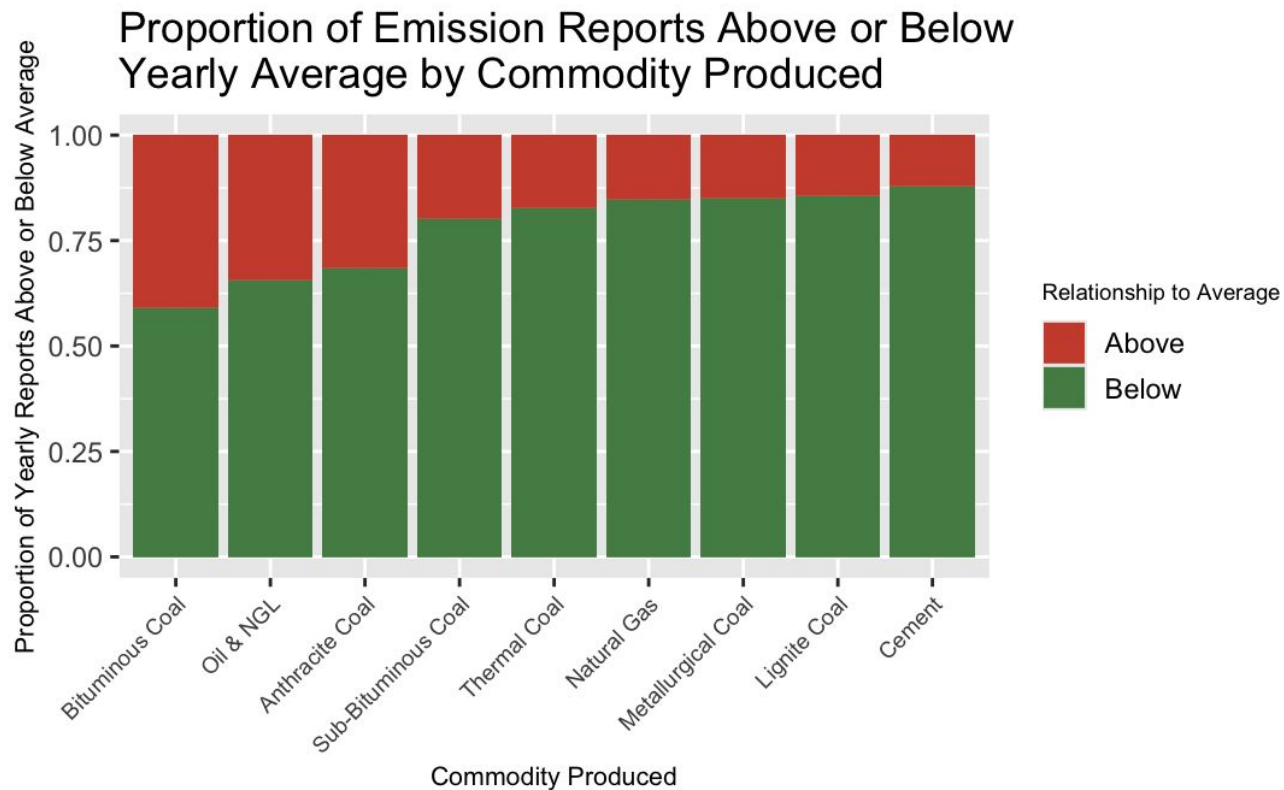
| year | parent_type | total_emissions_MtCO2e | commodity | year_avg | above_average |
|------|------------------------|------------------------|---------------------|------------|---------------|
| 1854 | Investor-owned Company | 0.01503125 | Lignite Coal | 0.03306606 | Below |
| 1854 | Investor-owned Company | 0.06915200 | Sub-Bituminous Coal | 0.03306606 | Above |
| 1854 | Investor-owned Company | 0.01501494 | Bituminous Coal | 0.03306606 | Below |
| 1855 | Investor-owned Company | 0.01954641 | Lignite Coal | 0.04299858 | Below |
| 1855 | Investor-owned Company | 0.08992416 | Sub-Bituminous Coal | 0.04299858 | Above |
| 1855 | Investor-owned Company | 0.01952519 | Bituminous Coal | 0.04299858 | Below |
| 1856 | Investor-owned Company | 0.02406156 | Lignite Coal | 0.05293111 | Below |
| 1856 | Investor-owned Company | 0.11069633 | Sub-Bituminous Coal | 0.05293111 | Above |
| 1856 | Investor-owned Company | 0.02403544 | Bituminous Coal | 0.05293111 | Below |
| 1857 | Investor-owned Company | 0.02796899 | Lignite Coal | 0.06152676 | Below |
| 1857 | Investor-owned Company | 0.12867265 | Sub-Bituminous Coal | 0.06152676 | Above |
| 1857 | Investor-owned Company | 0.02793863 | Bituminous Coal | 0.06152676 | Below |
| 1858 | Investor-owned Company | 0.03187642 | Lignite Coal | 0.07012241 | Below |
| 1858 | Investor-owned Company | 0.14664898 | Sub-Bituminous Coal | 0.07012241 | Above |
| 1858 | Investor-owned Company | 0.03184182 | Bituminous Coal | 0.07012241 | Below |
| 1859 | Investor-owned Company | 0.03578385 | Lignite Coal | 0.07871806 | Below |
| 1859 | Investor-owned Company | 0.16462531 | Sub-Bituminous Coal | 0.07871806 | Above |
| 1859 | Investor-owned Company | 0.03574501 | Bituminous Coal | 0.07871806 | Below |

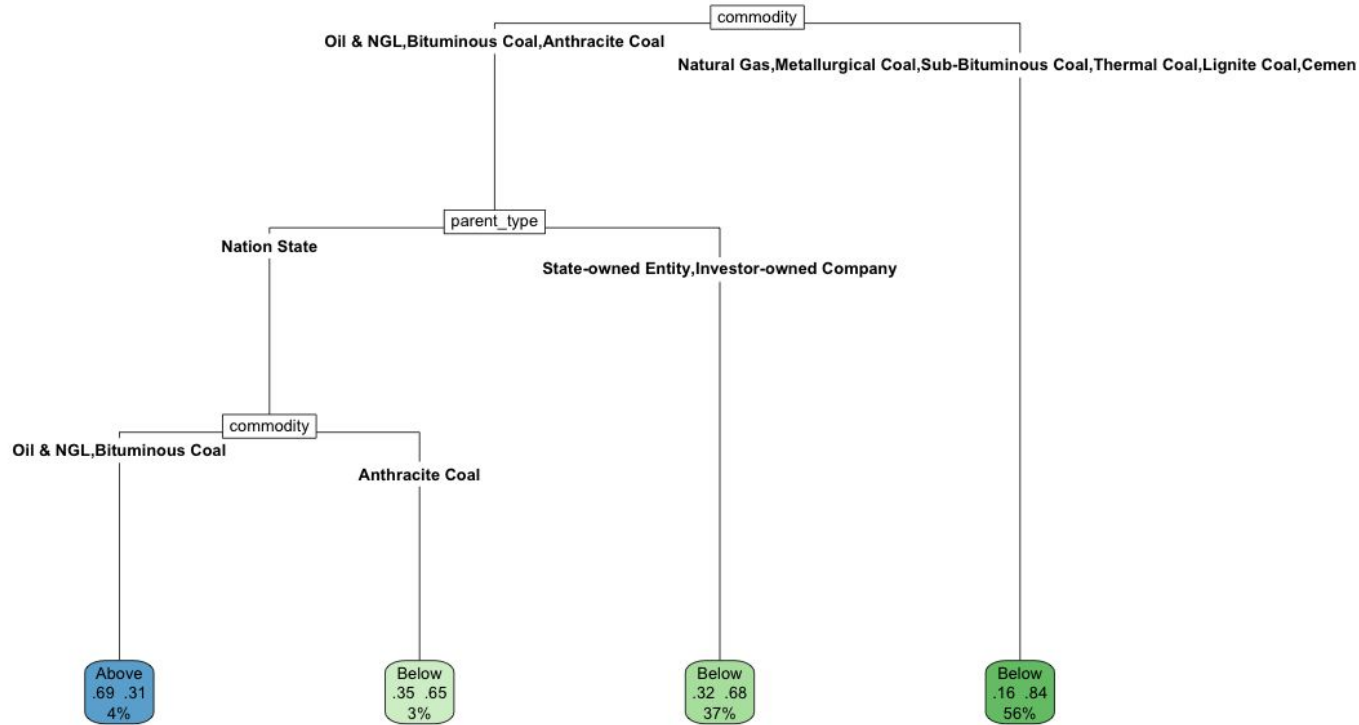
Proportion of Emission Reports Above or Below Yearly Average by Parent Entity Type

- Shows that fossil fuel producers owned or controlled by national governments have a significantly higher proportion of emissions reports that are above the yearly average.
- Companies owned by state governments or private investors have much lower proportions of emissions reports exceeding the average.



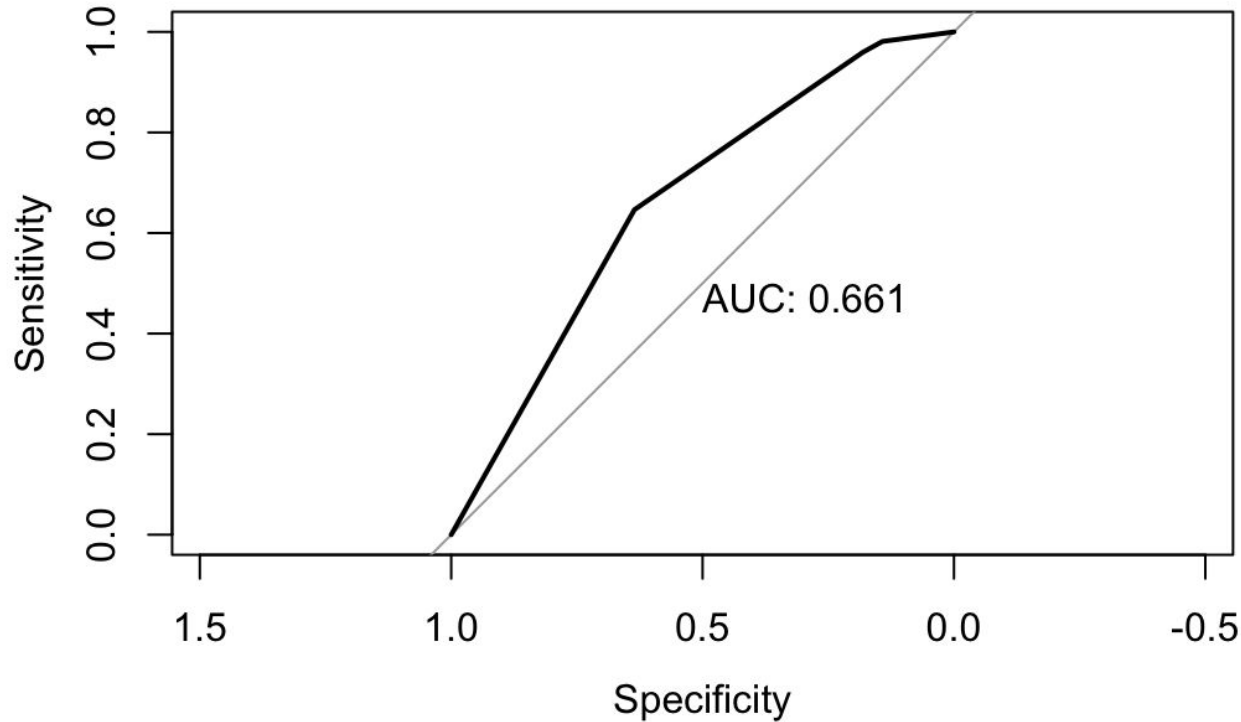
- Shows that the production of commodities such as bituminous coal, oil and NGL, and anthracite coal is associated with higher proportions of emissions reports that are above the yearly average than other commodities reported in this dataset.





- I used a classification tree to predict if a point was above or below average, rather than predicting the exact emission amount.
- The results of this tree reflect what we saw with the earlier visualizations
- Only 69% of the data classified as above average was correctly predicted, making it important to evaluate the strength of the model

ROC Curve for Decision Tree



- The model performs better than random guessing.
- While this performance is acceptable, there is substantial room for improvement
- Adding more predictive variables, such as the quantity of production, could increase the ability of this model to accurately classify observations.

Conclusion

- This project aimed to explore whether certain factors, specifically parent entity type and commodities produced, can help predict whether a fossil fuel producer's annual carbon emissions will fall above or below the industry average for that year.
 - The visualizations produced from this wrangled data revealed strong associations between above average emissions and certain company traits.
- The decision tree model developed to predict above or below average emissions supported these trends.
 - However, its performance, while better than random, has room for improvement.
 - One variable that could be added to the model is production quantity, which was not included in the initial model.

Sources

- <https://github.com/rfordatascience/tidytuesday/blob/main/data/2024/2024-05-21/readme.md>
- <https://carbonmajors.org/>