The Impacts of Geographic Proximity between European Countries on Carbon Emissions

Abstract

Problem

Does geographic proximity of high emission countries and low emission countries have an impact on carbon emissions? My hypothesis would be that the closer a low emission country is to a higher emission country, the more carbon emissions they would produce because they are being impacted by the high emission country.

Approach

I used countries (I removed the EU27 + UK column since it was just a sum of all other values), values (I added in values for 0's that did not seem like true 0's), sector, and year. I also imported the latitude and longitude points of each countries capital as well as the population size for each country. I grouped the data into high emission and low emission countries, depending on their total emissions for the 4 years. I also calculated the average distance to all high emission countries from each low emission countries.

Results

When comparing the average distance (between low and high emission countries) and the total residential emissions for low emission countries, I found a significant correlation. This correlation is found when looking at both capita and non capita data. When looking at total emissions, there was no significant correlation found.

Conclusion

My main findings were that there is a significant correlation between average distance and total residential emissions of low emission countries. However, different confounding variables could be impacting this, such as GDP or temperature. More models would need to be made in order to conclude that this correlation is because of the average distance, and not other variables.

Background

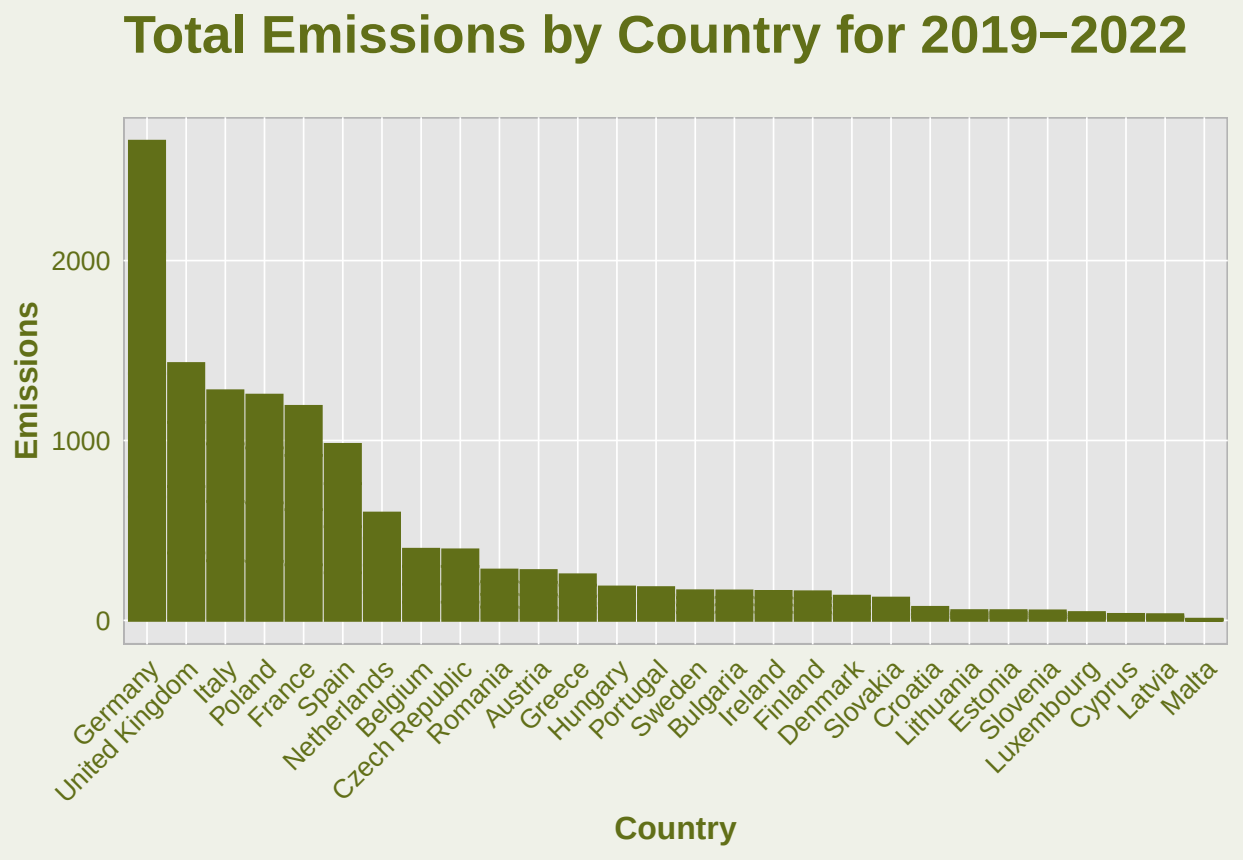
The dataset has carbon emissions data for every country in the European Union and the United Kingdom, with daily carbon emission values for 6 different sectors of emissions for the years 2019 to . The 6 different sectors were Power, Industry, International Aviation, Domestic Aviation, Residential, and Ground Transport. It came from the Carbon Monitor, a reputable source that has studied carbon emissions for years.

My research question was "Does geographic proximity of high emission countries and low emission countries have an impact on carbon emissions?". My hypothesis would be that the closer a low emission country is to a higher emission country, the more carbon emissions they would produce because they are being impacted by the high emission country.

Results

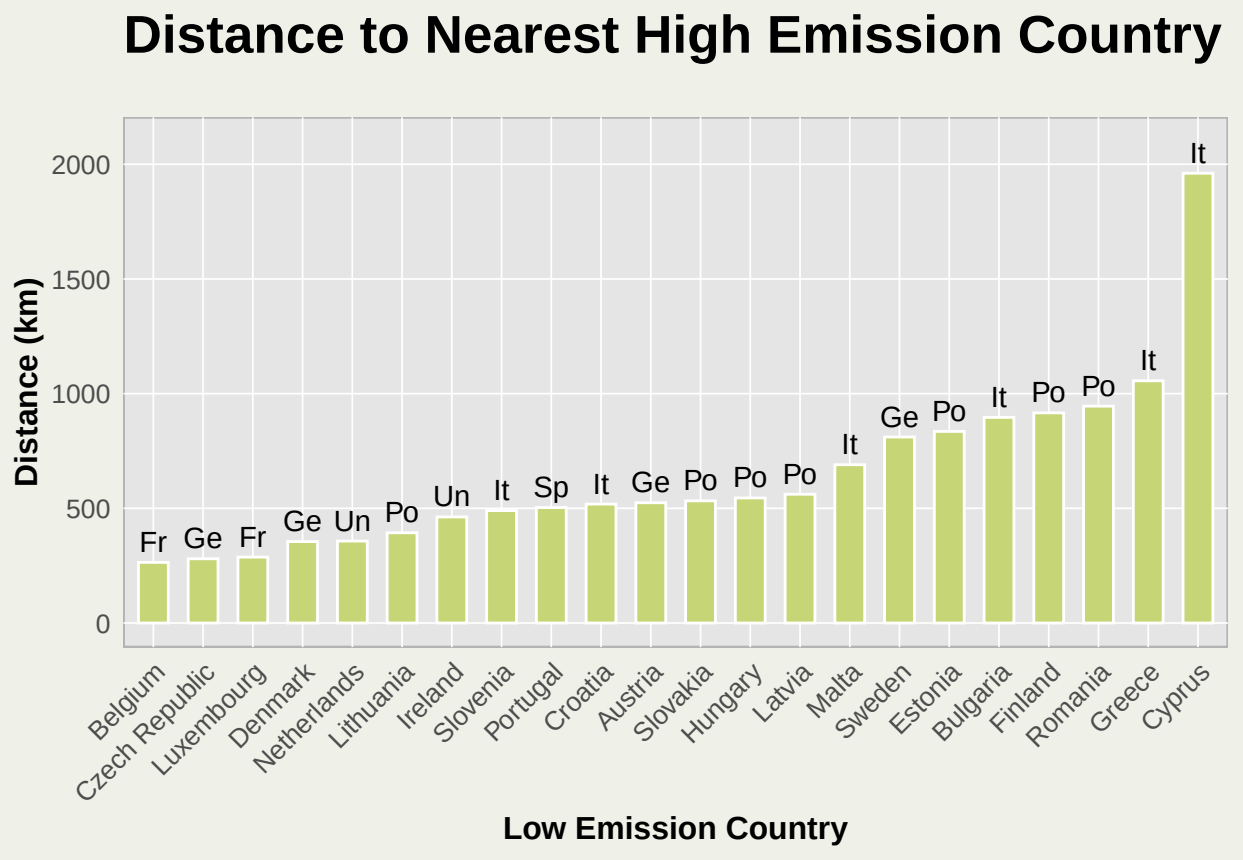
#1: Total Emissions by Country for 2019-2022

This graph shows a visual of all emissions by country for the years 2019-2022. It was used to then later visually see which countries belonged to each group, being either a high emission or low emission country. The top 6 countries here were selected as high emission countries, and the rest were selected as low emission countries.



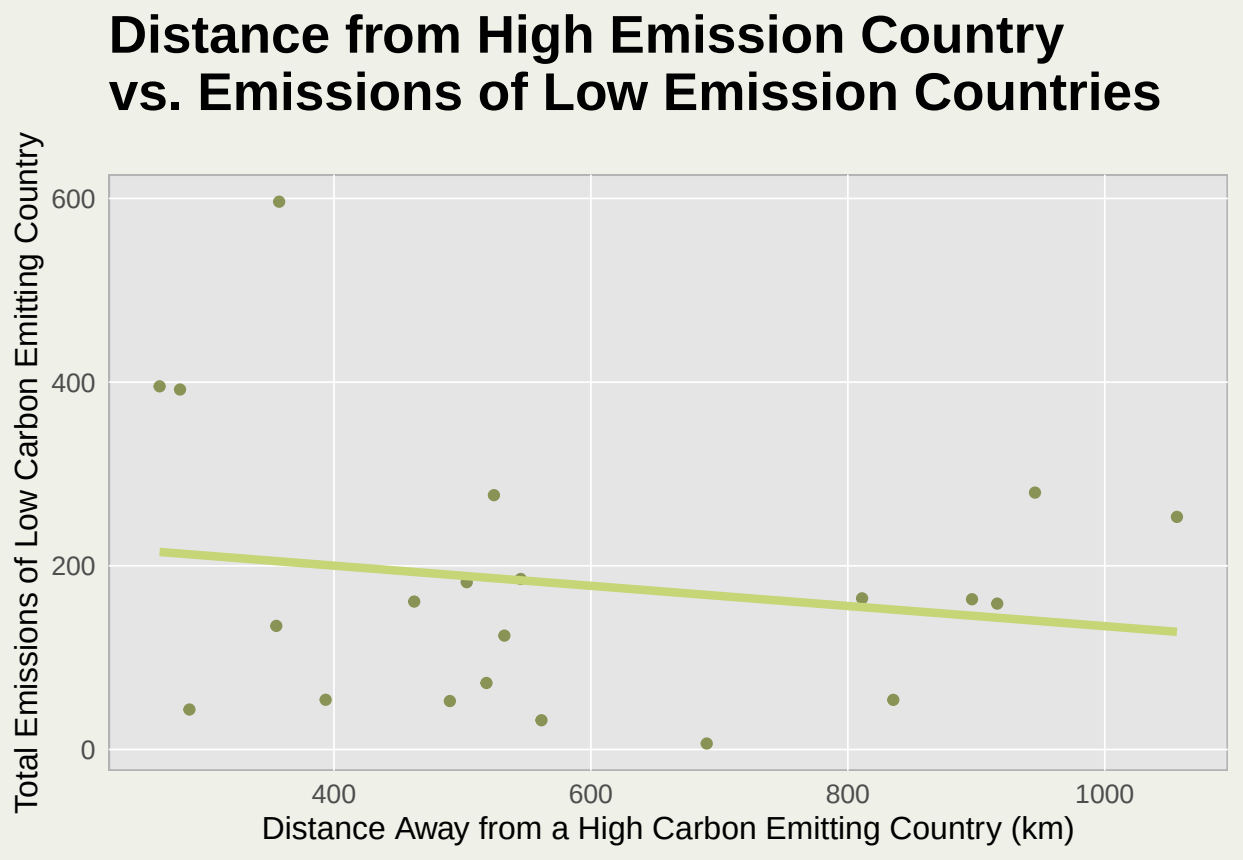
#2: Distance to Nearest High Emission Country for each Low Emission Country

This plot shows the nearest high emission country by distance to each low emission country. The closest high emission country is shown by it's first two letters above each countries bar.



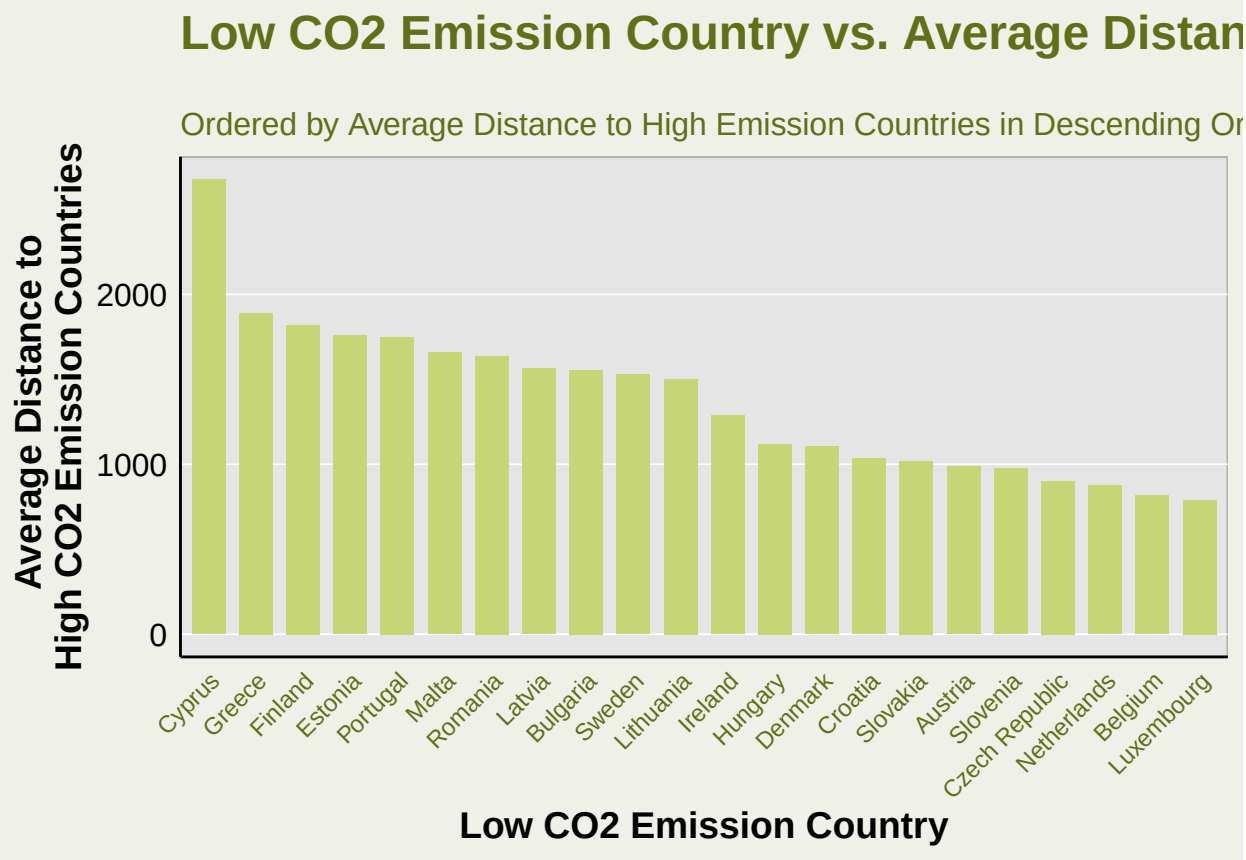
#3: Distance to Nearest High Emission Country vs. Emissions of Low Emission Country

This shows the correlation between the distance between each low emission country and it's closest high emission country and then the total emissions of the low emission country. There is a very weak correlation here, and I tried to approach different angles after seeing poor results from this correlation.



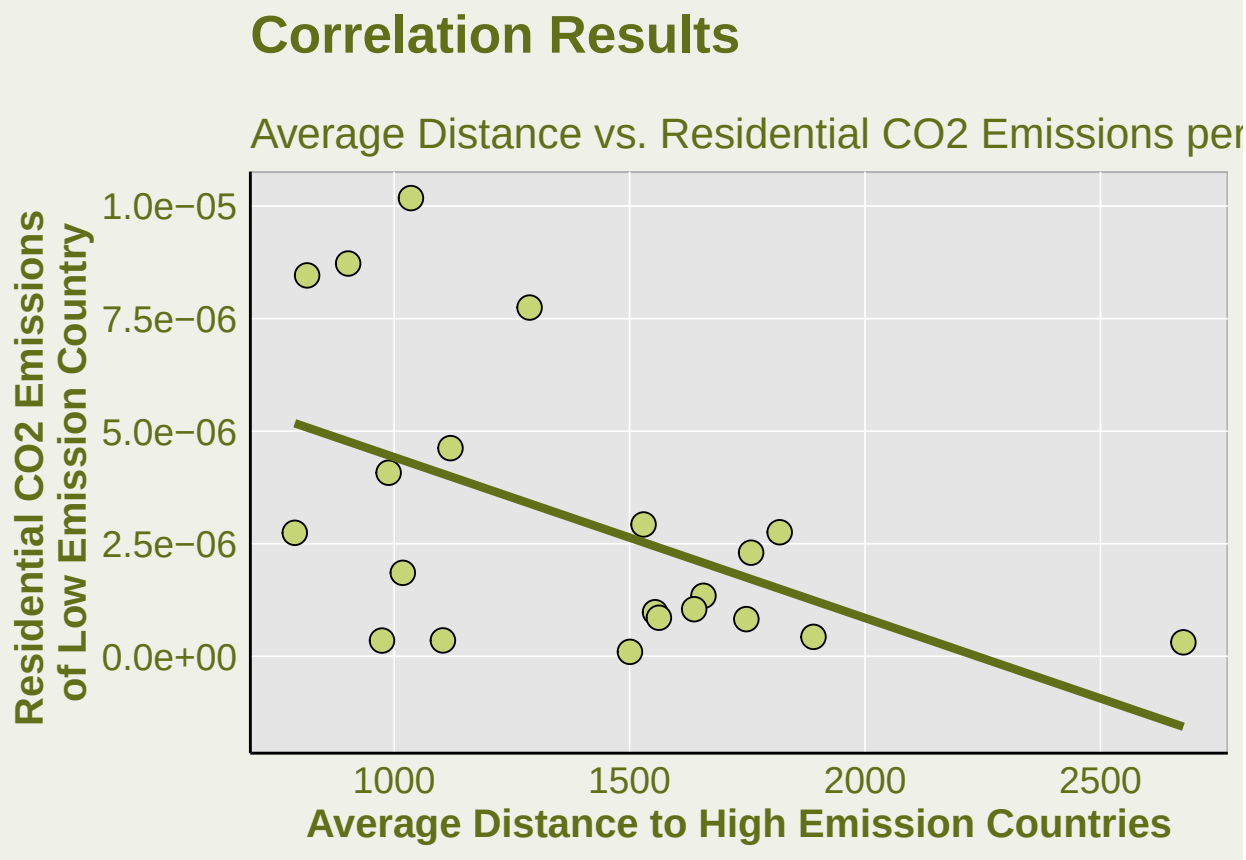
#4: Low CO2 Emission Country vs. Average Distance

This shows the average distance to all 6 high emission countries for each low emission country.



#5: Average Distance vs. Residential Emissions of Low Emission Countries per capita

This showed that there is a statistically significant correlation between the average distance from low emission countries to all of the high emission countries and the residential emissions of low emission countries per capita. P-value of 0.01488803.

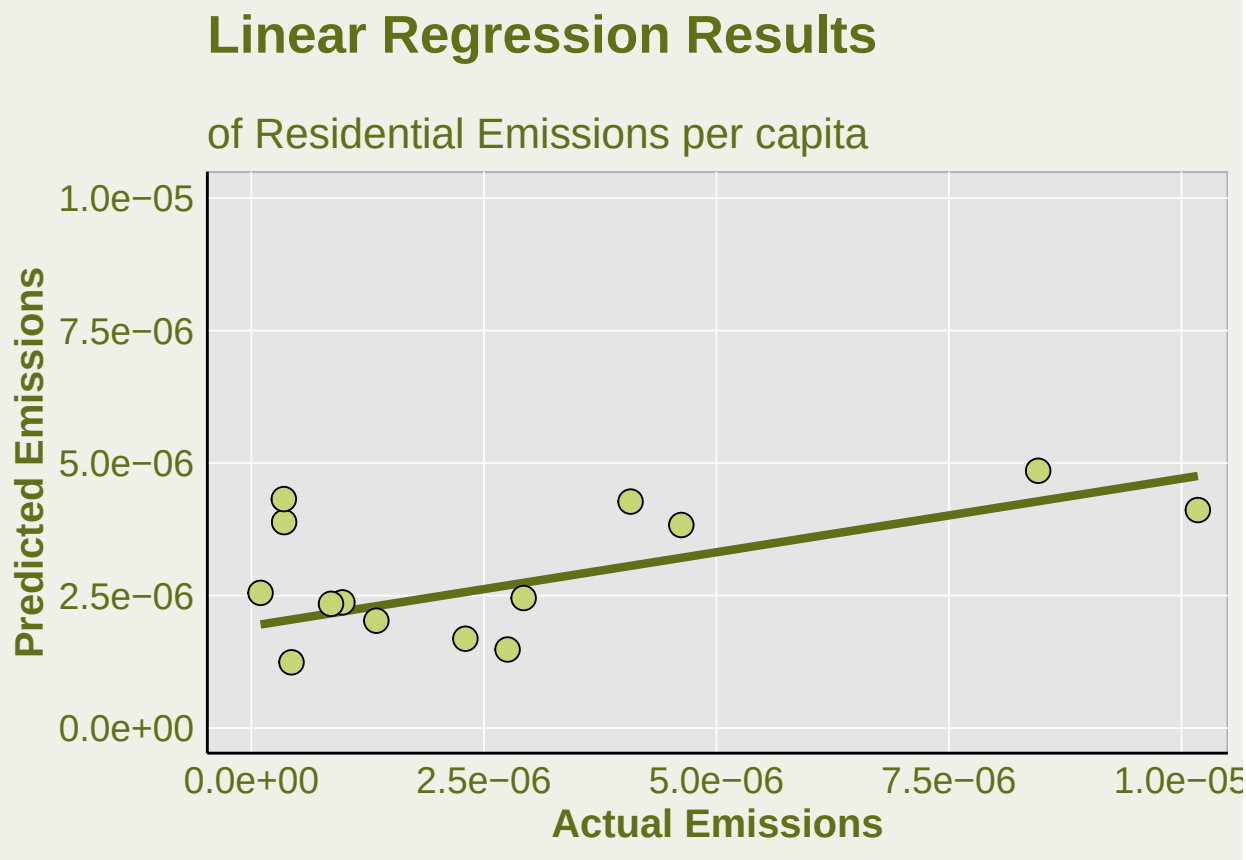



```
## [1] 0.01488803
```

#6: Linear Regression of Residential Emissions per capita given Average Distance

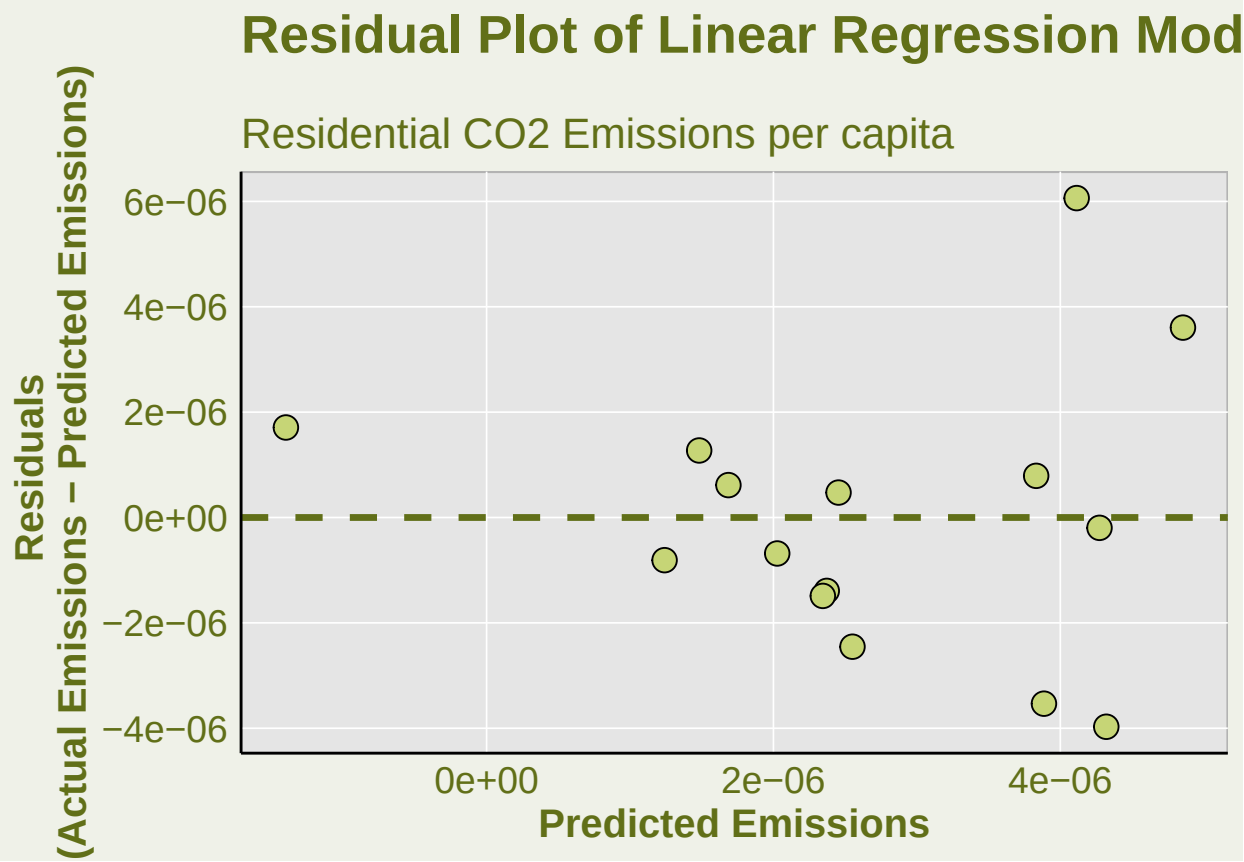
I used a linear regression model to predict the residential emissions per capita by just giving it the average distance. This shows the actual emissions vs. the predicted emissions by the model. It has a p-value of 0.04339, which makes the results significant.

```
##
## Call:
## lm(formula = capita ~ avgDistance, data = reg_train_capita)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -3.972e-06 -1.441e-06 -1.961e-07  1.032e-06  6.063e-06
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.594e-06  2.309e-06   3.289  0.00587 **
## avgDistance -3.361e-09  1.502e-09  -2.238  0.04339 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.705e-06 on 13 degrees of freedom
## Multiple R-squared:  0.2781, Adjusted R-squared:  0.2225
## F-statistic: 5.007 on 1 and 13 DF,  p-value: 0.04339
```



#7: Residual Plot of Linear Regression

This shows the residual plot of the linear regression plot, noted as #6. There is no noticeable pattern within this residual plot, so it can be said that the model is working as intended and not predicting the emissions incorrectly as a pattern.



Discussion

With my results, I can conclude that as the average distance to high emission countries increases, the residential emissions of low emission countries decreases. It is not currently known whether average distance is the true driving factor for this, as there could be outside confounding variables causing this that have not been accounted for. This could include the GDP of countries, their average temperature, and/or different political agreements between countries. In the future, I will attempt to measure the impact of each of these variables, and then remove that impact from the data to then see if average distance still shows a strong correlation with emissions.

Code and Data Availability

Here is a link to my Github repository for this project: https://github.com/the-codingschool/DSRP-2023-Batzer/tree/main/Caitlyn

Link to carbon emissions dataset: https://carbonmonitor.org/

Link to latitude and longitude of capitals: https://gist.github.com/ofou/df09a6834a8421b4f376c875194915c9

Link to population for each country for each year: https://github.com/the-codingschool/DSRP-2023-Batzer/blob/main/Data/worldbank_population_bycountry%20-%20data.csv ** Some of the populations here are swapped with each other, I had to manually fix them

Acknowledgements

I would like to acknowledge my mentor for my research group, Evan Batzer, The Coding School for giving me this opportunity to explore data science and research, and Alex Dong, the TA for my group.