Final project

Introduction:

Question: How does GDP impact CO2 emissions?

Reasons for analysis: CO2 emissions are impacting climate change, and contribute to many current health impacts, so we are interested in seeing how the a country's GDP impacts the levels of CO2 that they emit. CO2 traps heat which can lead to climate change, and respiratory issues due to the smog and air pollution

Thesis: CO2 emissions are causing a wide variety of environmental and health issues throughout the world. We believe that people in countries with a higher GDP per capita are larger contributors to the rise in CO2 emissions than countries with a smaller GDP per capita.

Background:

Data: Our first set of data was found on Kaggle and was collected by OurWorldInData. It contains the CO2 emissions for countries in the years 1750 through 2017. The variables in the data are the countries, their country abbreviations, the year, and the CO2 emissions from that year [1]. Our second dataset was also found on Kaggle and was originally from Worldometers. This dataset contains the country, the year from 1994-2017, and the GDP per capita in units of U.S. dollars [2].

Note: For this project we will only be using the data for the years 2012-2017 from both of our datasets.

Variables:

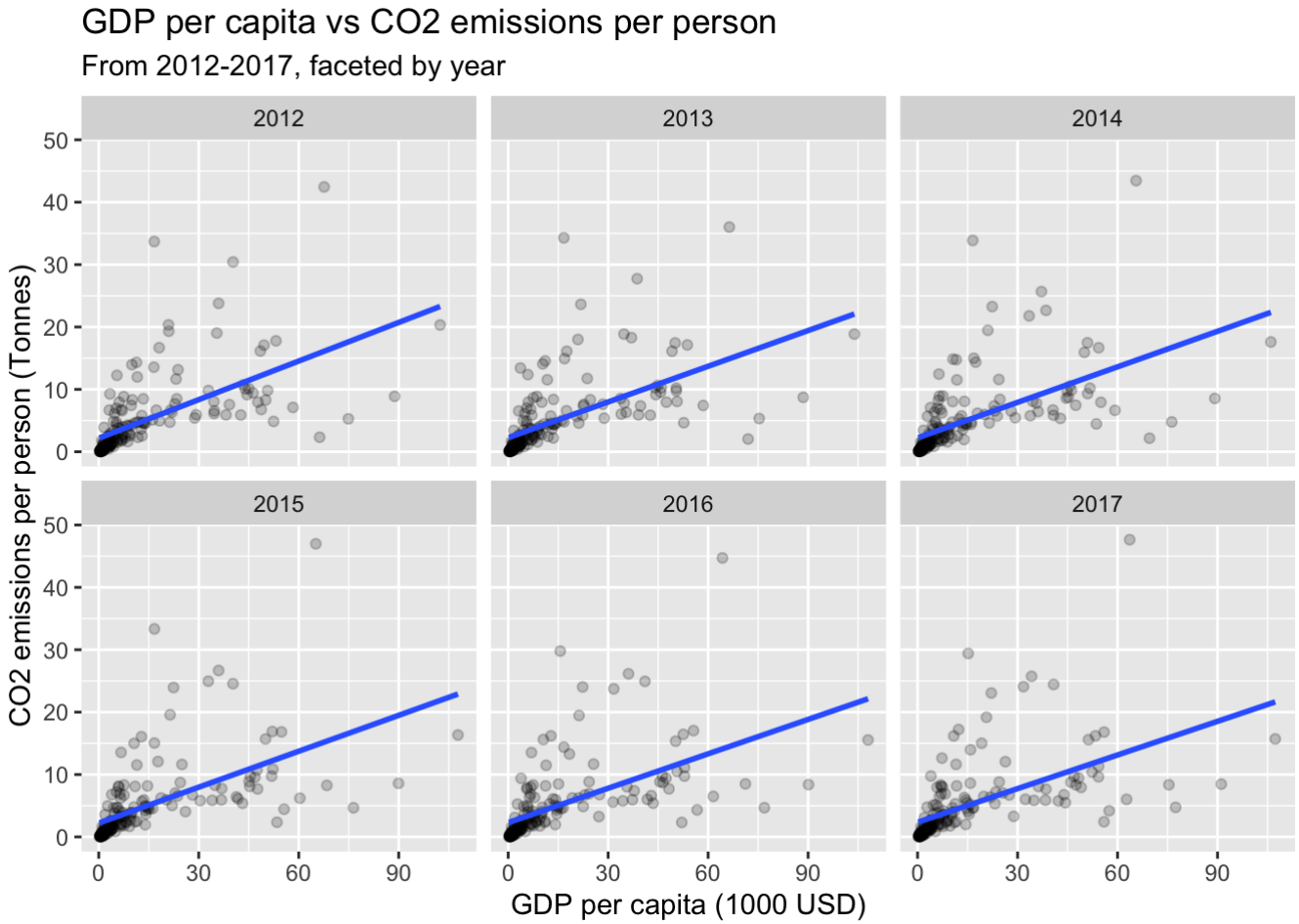
- CO2 emissions: CO2 emitted by a country in units of tonnes.
- GDP: standard measure of the value added created through the production of goods and services in a country during a certain period.
- GDP per capita: GDP of a country divided by its population, in units of 1000 USD.

Unusual Factors: Although a country's GDP may be one factor in determining their CO2 emissions, it also could be influenced by a variety of other factors such as people's behavior and a country's policies. Therefore, despite any correlation, we cannot make any definite conclusions on the impact of GDP on CO2 emissions as we are not taking into account the other factors.

Goals: We want to look at how the GDP per capita of the countries correlates with their CO2 emission levels per person, and compare the different nations.

Overview:

Graph faceted by year to see the general trend for different years:



This graph shows us that the trend between GDP per capita and CO2 emissions per person has not changed much over these six years. This means that the mean values of GDP per capita and CO2 emissions per person will be good approximations to use for our analysis.

Preview of the final dataset:

```
## # A tibble: 6 x 3
##   Country             avg_co2_emissions_per_person avg_gdp_per_capita
##   <chr>                                      <dbl>              <dbl>
## 1 Afghanistan                                0.324              0.577
## 2 Albania                                    1.97               4.5
## 3 Algeria                                    3.62               4.72
## 4 Andorra                                    6.19              41.1
## 5 Angola                                     1.25               3.68
## 6 Antigua and Barbuda                        5.85              13.5
```

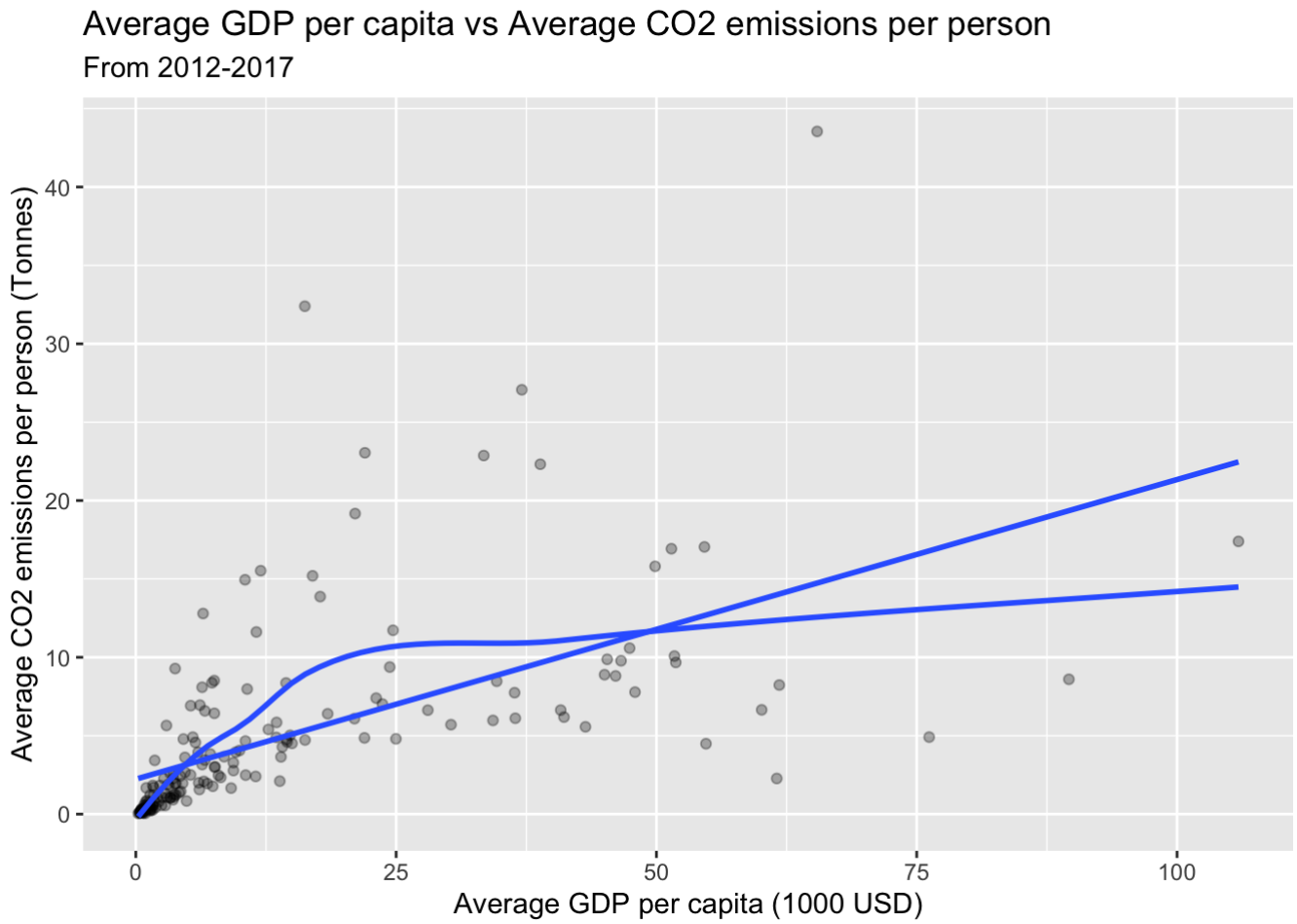
Variables:

- Country: name of country
- avg_co2_emissions_per_person: Average CO2 emissions per person in tonnes
- avg_gdp_per_capita: Average GDP per capita in 1000 USD

Analysis:

Method: Calculated the mean GDP per capita and mean CO2 emissions per person from 2012-2017 for each country, then graphed these variables using a scatterplot. Also added both a smooth curve and a regression line using the lm method to view their relationship.

Graph:



Results:

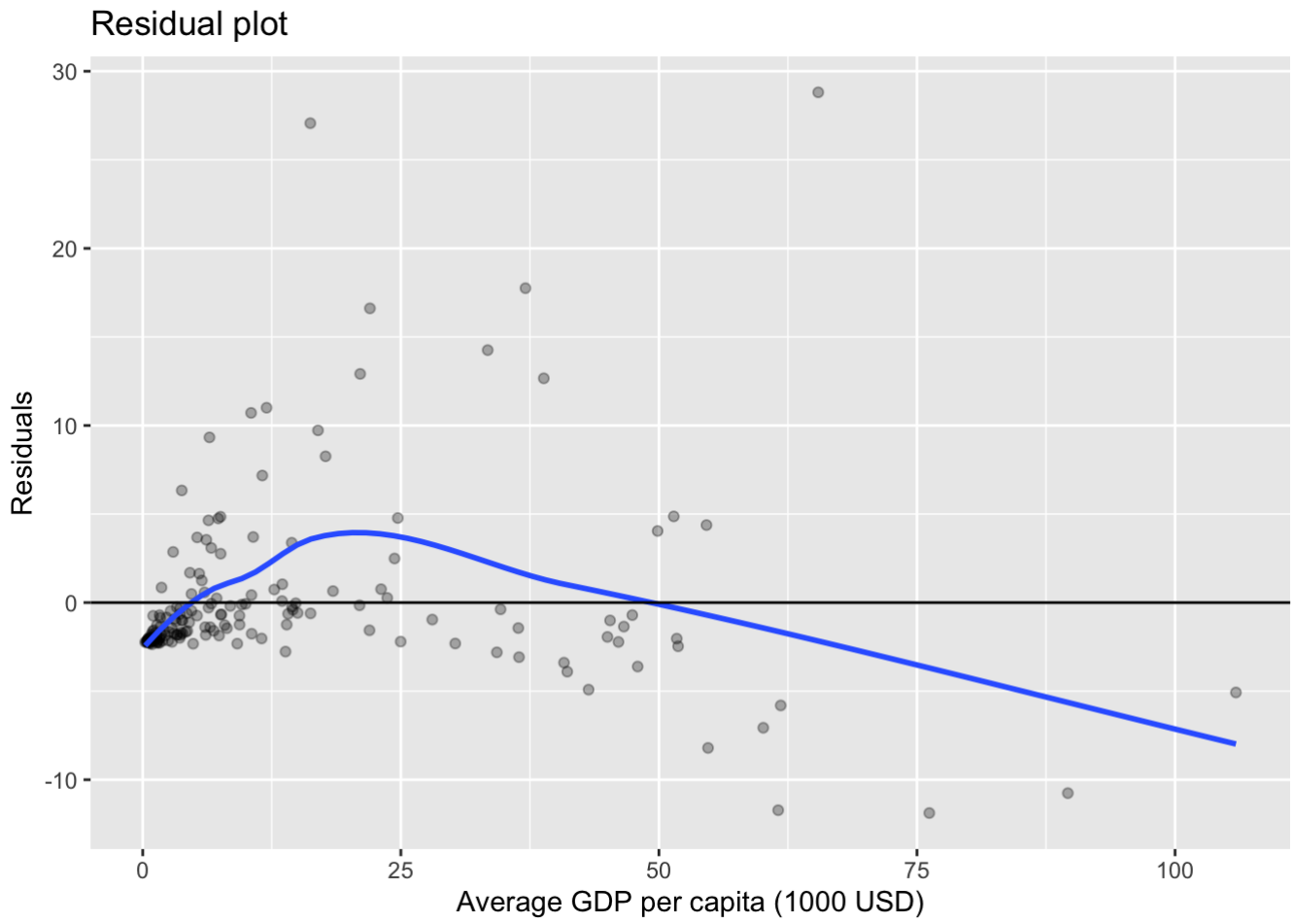
The regression line shows us that the overall relationship between GDP per capita and CO2 emissions per person is positive. However, the smooth curve tells us that this relationship is much stronger for countries with GDP per capita <= 25,000 USD and is much weaker for GDP per capita > 25,000 USD.

Evidence to support claim:

Regression model summary:

```
## # A tibble: 2 x 4
##   parameter estimate     se  p_value
##   <chr>        <dbl>  <dbl>    <dbl>
## 1 Intercept    2.23   0.485  8.49e- 6
## 2 Slope        0.191 0.0205 5.33e-17
```

Residual Plot:



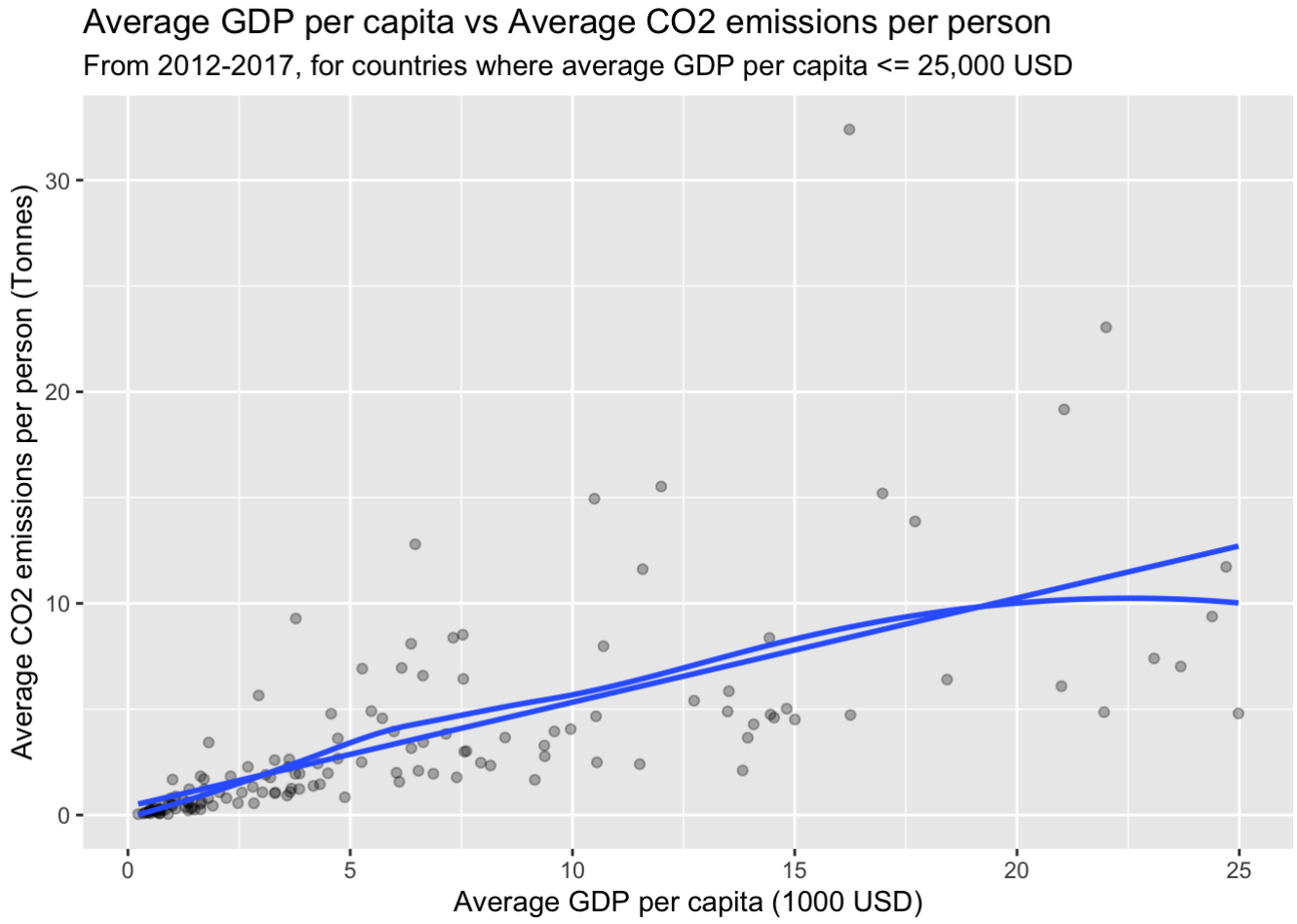
Results:

- Slope = 0.191, which means that for every increase of 1000 USD in GDP per capita there is an increase of 0.191 tonnes of CO2 emissions per person
- p-value = 5.333760e-17
- The smooth line of the residual plot moves away from y = 0, which indicates a lack of model fit as GDP per capita gets larger.

Further Analysis:

Method: A comparison between the regression line and the smooth curve, as well as the residual plot of our initial model, revealed the model to be fairly inaccurate. The smooth curve told us that the relationship between GDP per capita and CO2 emissions per person is much stronger for countries with GDP per capita <= 25,000 USD and is much weaker for GDP per capita > 25,000 USD. As a result, we have decided to split the data into two groups at this threshold, and create two separate models using the same steps as above. The first model will be for countries with average GDP per capita <= 25,000 USD, and the second model will be for countries with average GDP per capita > 25,000 USD.

Graph 1:



Results:

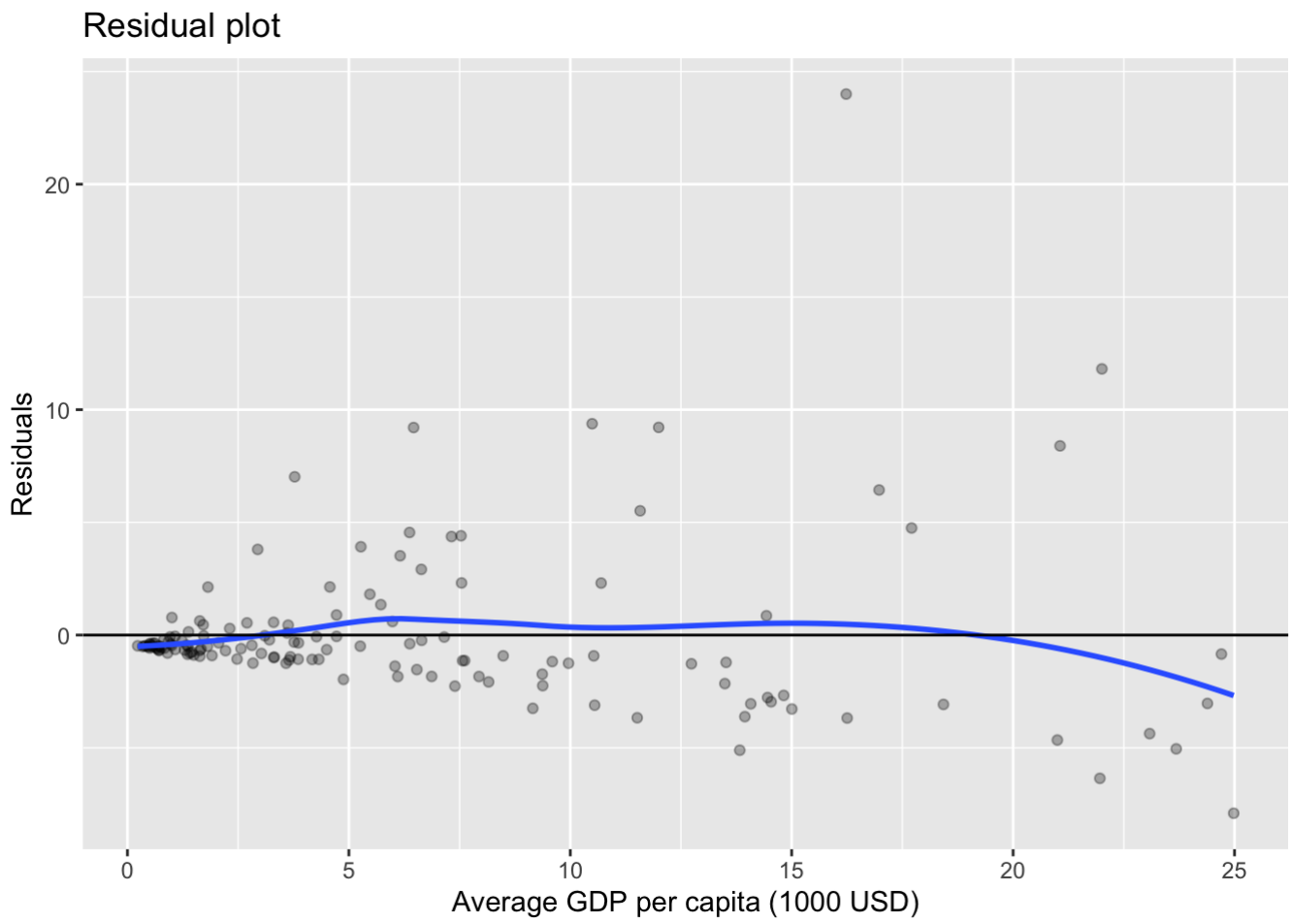
The regression line shows us that the overall relationship between GDP per capita and CO2 emissions per person is positive. The regression line and the smooth curve are very similar which indicates that this model is more accurate.

Evidence to support claim:

Regression model summary 1:

```
## # A tibble: 2 x 4
##   parameter estimate     se  p_value
##   <chr>        <dbl>  <dbl>    <dbl>
## 1 Intercept    0.404 0.418  3.36e- 1
## 2 Slope        0.492 0.0467 1.73e-19
```

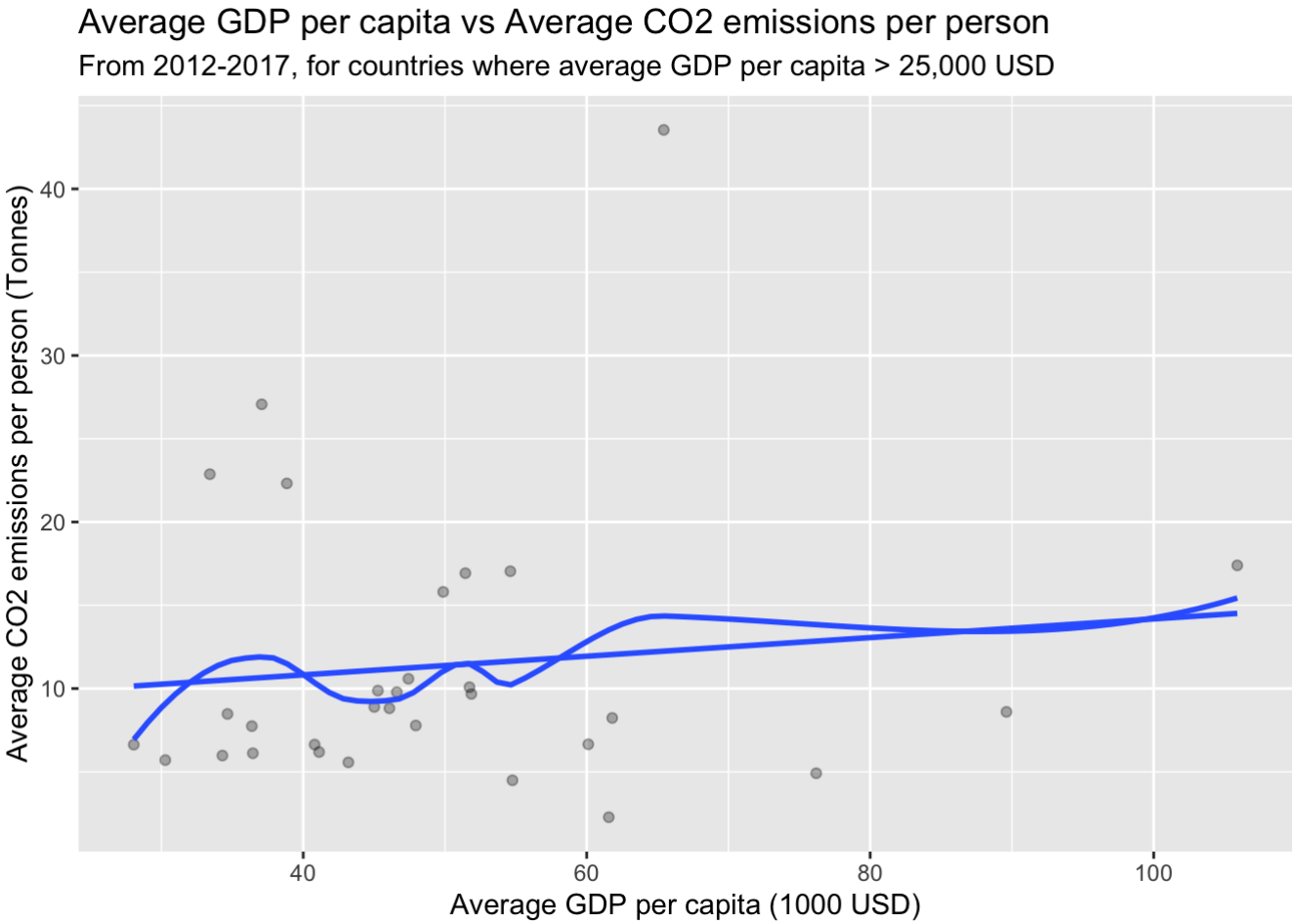
Residual Plot 1:



Results:

- Slope = 0.492, which means that for every increase of 1000 USD in GDP per capita there is an increase of 0.492 tonnes of CO2 emissions per person
- p-value = 1.731318e-19
- The smooth curve of the residual plot remains close to y = 0 as GDP per capita increases, which indicates a good model fit.

Graph 2:



Results:

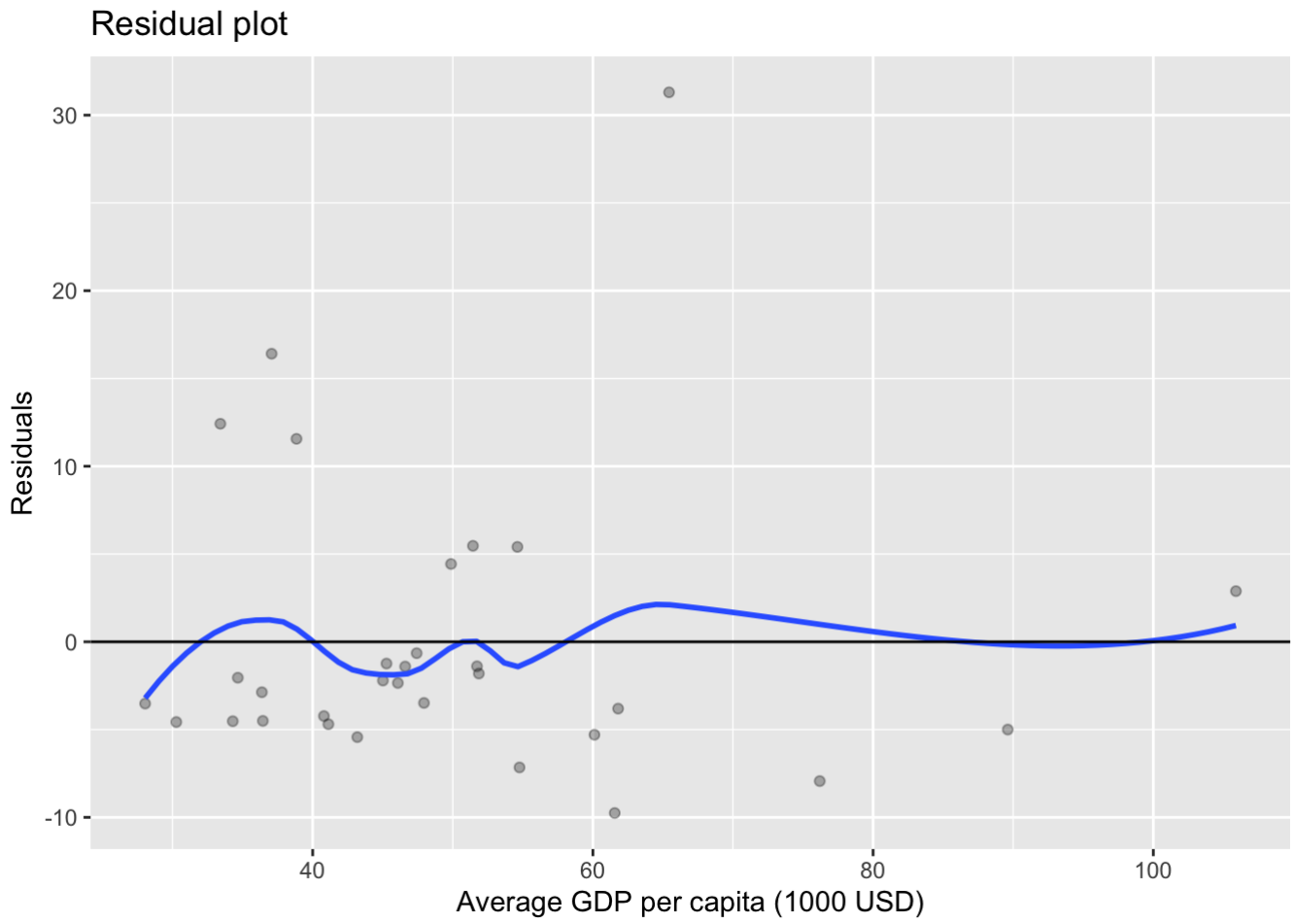
The regression line shows us that the overall relationship between GDP per capita and CO2 emissions per person is positive but very close to 0. The regression line and the smooth curve are very similar which indicates that this model is more accurate.

Evidence to support claim:

Regression model summary 2:

```
## # A tibble: 2 x 4
##   parameter estimate     se p_value
##   <chr>        <dbl>  <dbl>   <dbl>
## 1 Intercept   8.58   4.81    0.0849
## 2 Slope       0.0560 0.0914  0.545
```

Residual Plot 2:



Results:

- Slope = 0.056, which means that for every increase of 1000 USD in GDP per capita there is an increase of 0.056 tonnes of CO2 emissions per person
- p-value = 0.54471231
- The smooth curve of the residual plot stays very close to y = 0 as GDP per capita increases, which indicates a good model fit.

Discussion:

Interpretation:

Our null hypothesis states that there is no relationship between GDP per capita and CO2 emissions per person.

- For our aggregate model, there is strong evidence (p-value = 5.33e-17) to disprove our null hypothesis, suggesting that there is a positive relationship (slope = 0.191) between GDP per capita and CO2 emissions per person. However, the residual plot as well as a comparison between the smooth curve and the regression line tell us that this model is fairly inaccurate.
- For countries where average GDP per capita <= 25,000 USD, our model showed that there is strong evidence (p-value = 1.73e-19) to disprove our null hypothesis, suggesting that there is a positive relationship (slope = 0.492) between GDP per capita and CO2 emissions per person. The residual plot as well as a comparison between the smooth curve and the regression line tell us that this model is much more accurate than our aggregate model. It is also important to note that the slope we get for this model is

much greater than the slope of our initial model, which tells us that the increase in CO2 emissions for every 1000 USD increase in GDP per capita is much greater for countries where average GDP per capita <= 25,000 USD.

- For countries where average GDP per capita > 25,000 USD, our model showed that there is a lack of evidence (p-value = 0.54471231) to disprove our null hypothesis, suggesting that there is no relationship between GDP per capita and CO2 emissions per person. The residual plot as well as a comparison between the smooth curve and the regression line tell us that this model is much more accurate than our aggregate model. It is also important to note that the slope we get for this model is much smaller than the slope of our initial model, which tells us that the increase in CO2 emissions for every 1000 USD increase in GDP per capita is much smaller for countries where average GDP per capita > 25,000 USD.

Potential short-comings:

CO2 emissions for countries with an average GDP per capita greater than 25,000 USD are not consistent with the aggregate data. Therefore, the conclusions of this project can only support countries with an average GDP per capita below 25,000 USD, which leads to insufficient precision. This project only analyzes the relationship between CO2 emissions and GDP per capita for 2012-2017. In other years, the relationship between CO2 emissions and GDP per capita may be different, so the conclusions of this project are not generalizable. According to the plots obtained from the analysis, the rate of change in CO2 emissions with GDP per capita differs between countries with lower and higher GDP per capita. Therefore, although the conclusion of this project represents a positive correlation between GDP per capita and CO2 emissions in general, the degree of correlation varies from country to country, and the conclusion may not be applicable to all countries.

Potential future directions:

Because of the distinction made between countries with a lower GDP per capita than 25,000 and countries with a GDP per capita of more than 25,000 it is important to analyze the future potential findings of each independent case. To start, the countries with a GDP per capita less than 25,000 have a linear relationship between GDP per capita and CO2 emissions. Since GDP is used as a measure of standard of living or well being of a country, and both GDP and CO2 emissions are related as explained above, this would mean that if a country were to increase their GDP per capita then inevitably it would cause an increase in CO2 emissions of that country. However, this is not ideal as CO2 emissions are one of the causing factors of climate change. A potential future study into the causing factors of these CO2 emissions would yield ways countries could go about decreasing their emissions. On the other hand, countries who have a GDP per capita higher than 25,000 did not have a direct relationship with CO2 emissions. One potential future study that could be done on this data, is finding out what natural resources are exported in these countries and what is the physical, social and economic cost of obtaining these resources.

---

1. https://www.kaggle.com/yoannboyere/co2-ghg-emissionsdata (https://www.kaggle.com/yoannboyere/co2-ghg-emissionsdata)↩
2. https://www.kaggle.com/themlphdstudent/country-wise-gdp-from-1994-to-2017 (https://www.kaggle.com/themlphdstudent/country-wise-gdp-from-1994-to-2017)↩