

Data 606 Project

Alyssa Gurkas

Can a larger Gross Domestic Product (GDP) indicate more CO2 Emissions?

This project will fit a model that uses GDP to predict CO2 emissions, exploring if there is a relationship between GDP and CO2 Emissions.

Alyssa Gurkas | December 9, 2024

Background

Carbon dioxide (CO₂) is a type of greenhouse gas, known for trapping heat, that is emitted into the atmosphere from burning fossil fuels (like coal, oil, and natural gas), and other natural processes like wildfires or volcanic eruptions.

CO₂ in the atmosphere warms the planet, and causes global warming (or climate change). Human dependence on fossil fuels raised CO₂ levels in the atmosphere. According to NASA, CO₂ content in the atmosphere increased by 50% in less than 200 years. [Source](#)

Background Cont.

The U.S. Department of Energy prepares an annual Electric Power Report that includes information about energy production, sales, consumption of fossil fuels, environmental data, and other topics related to energy. Additionally, the U.S. Department of Commerce tracks state annual summary statistics which include GDP by state from years 1998-2023.

This project uses data from the Department of Energy's (DOE) Annual Electric Power Report and the U.S. Department of Commerce (DOC) State Annual Summary Statistics.

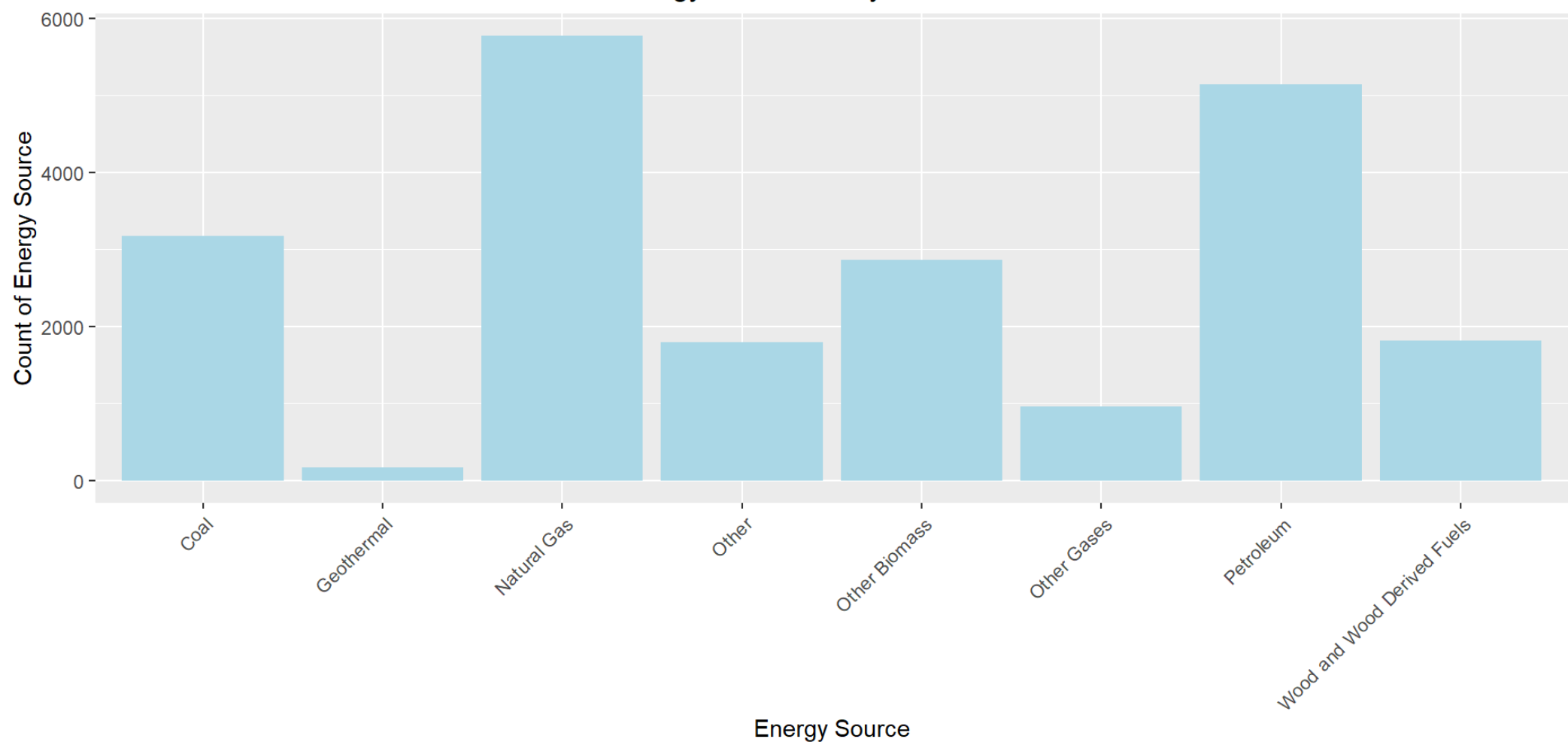
Dataset Characteristics

Each row represents the respective state's annual energy usage and GDP for the reporting year. There are 1326 rows in the data set and eight columns.

- Year
- State
- Producer Type
- Energy Source
- CO2 (Metric Tons)
- SO2 (Metric Tons)
- NOx (Metric Tons)
- GDP

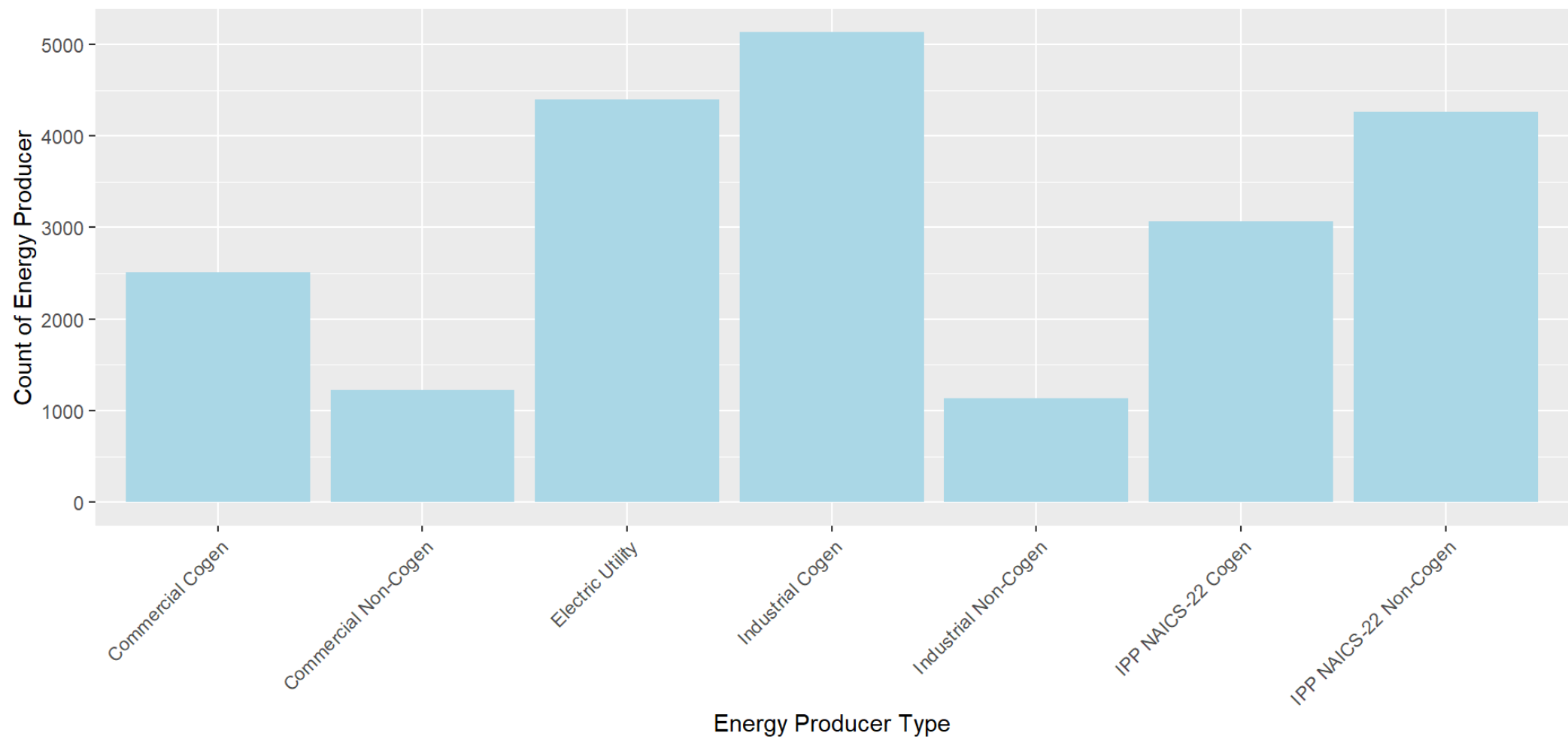
Data Exploration (Energy Production)

2023 Energy Production by Source in the U.S.



Data Exploration (Producer Type)

2023 Energy Producer by Type in the U.S.



Data Exploration (CO2 and GDP)

```
1 summary(state_gdp_emissions$CO2)
```

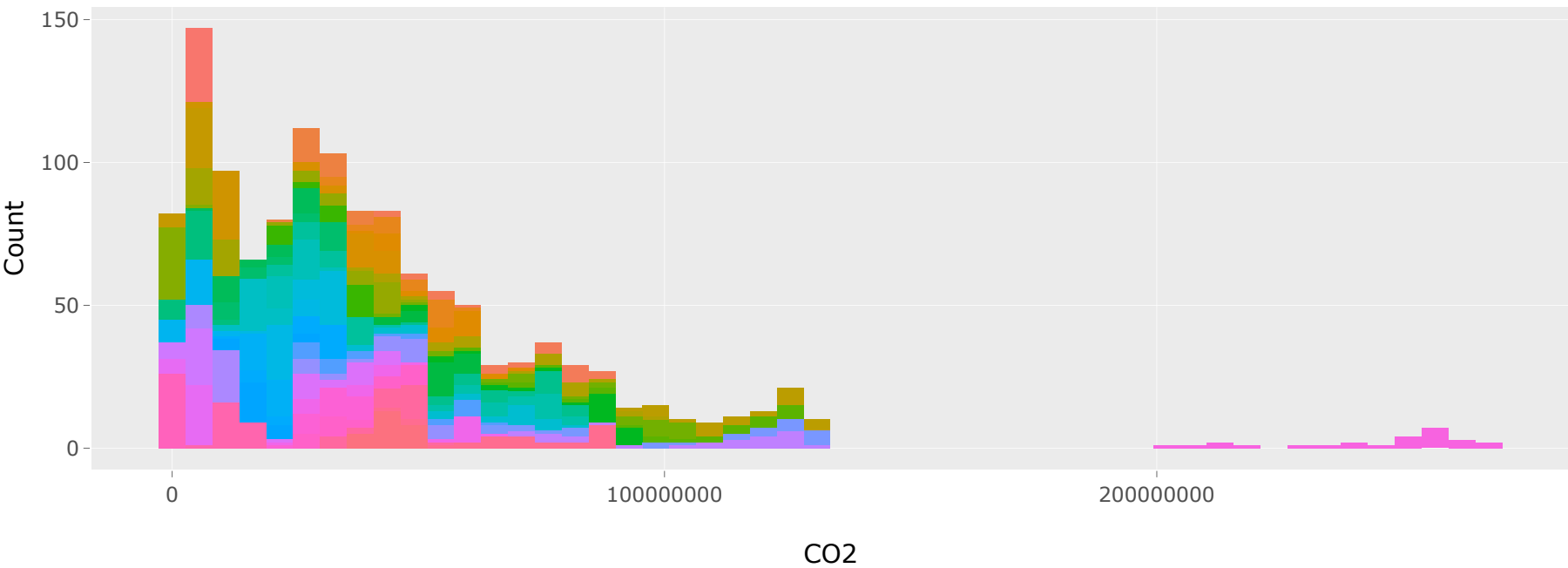
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
6583	13537718	33019610	43277808	57791802	267464092

```
1 summary(state_gdp_emissions$GDP)
```

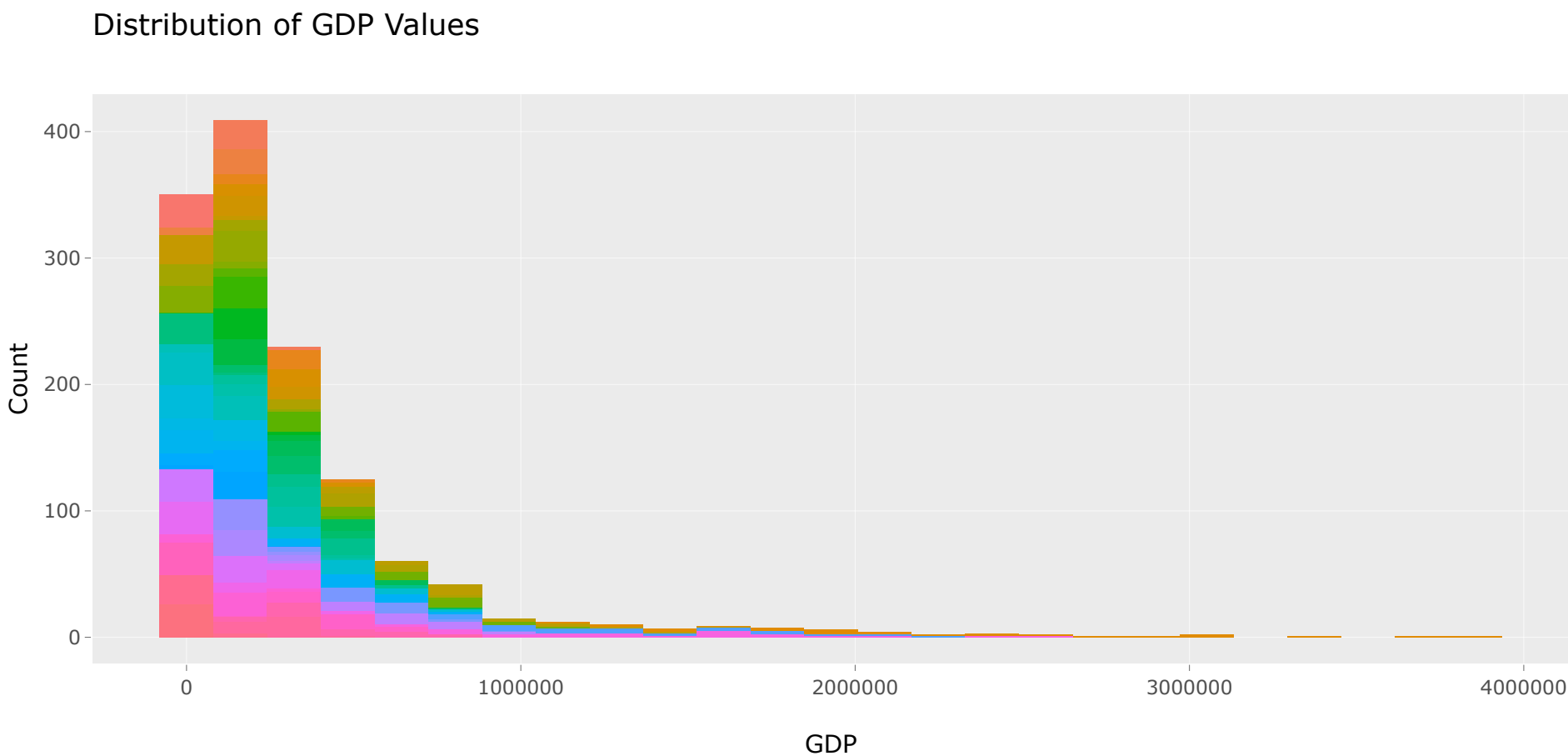
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
14833	74347	189134	321806	389481	3870379

Data Exploration - CO2 Distribution

Distribution of CO2 Values



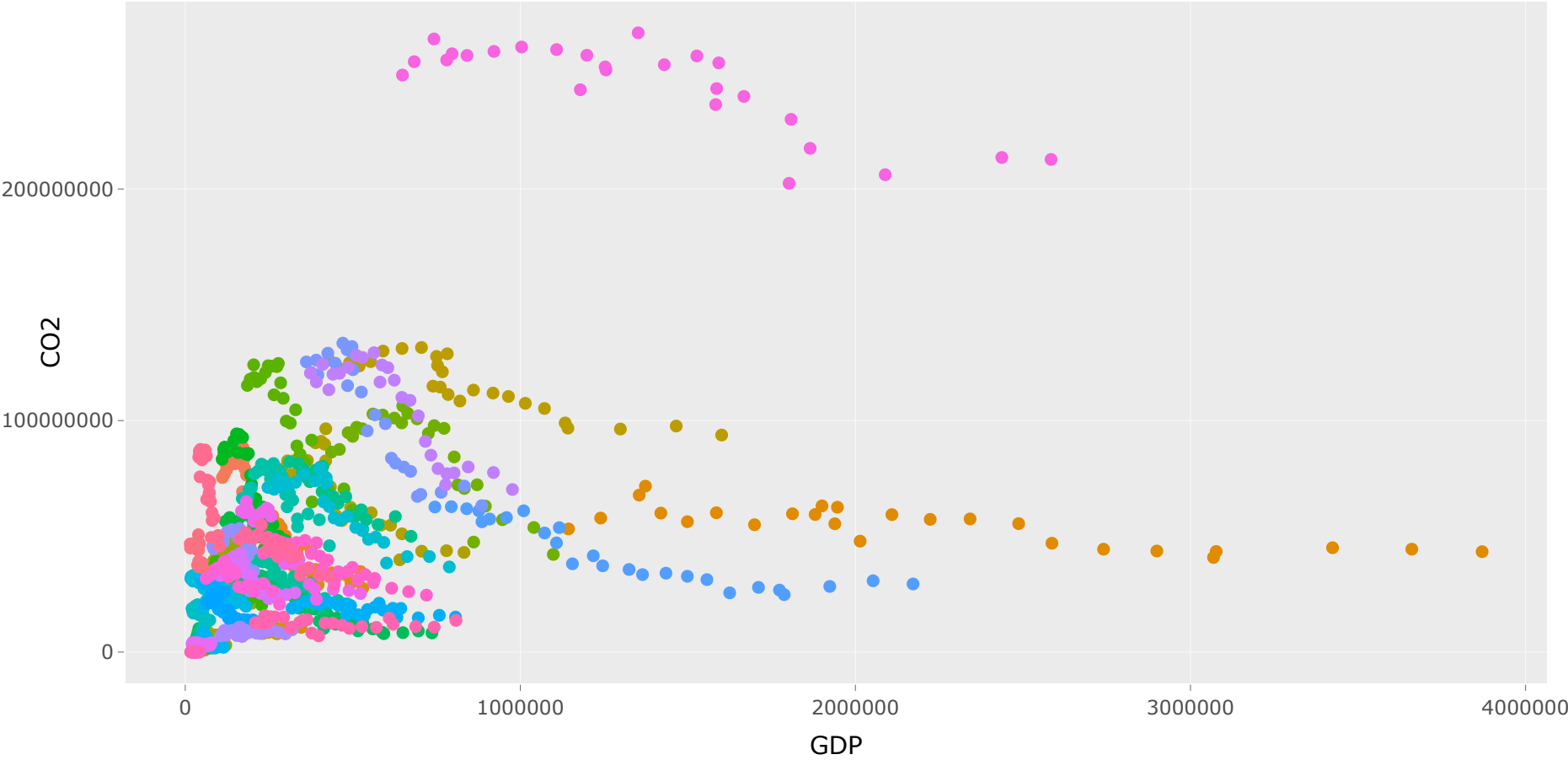
Data Exploration - GDP Distribution



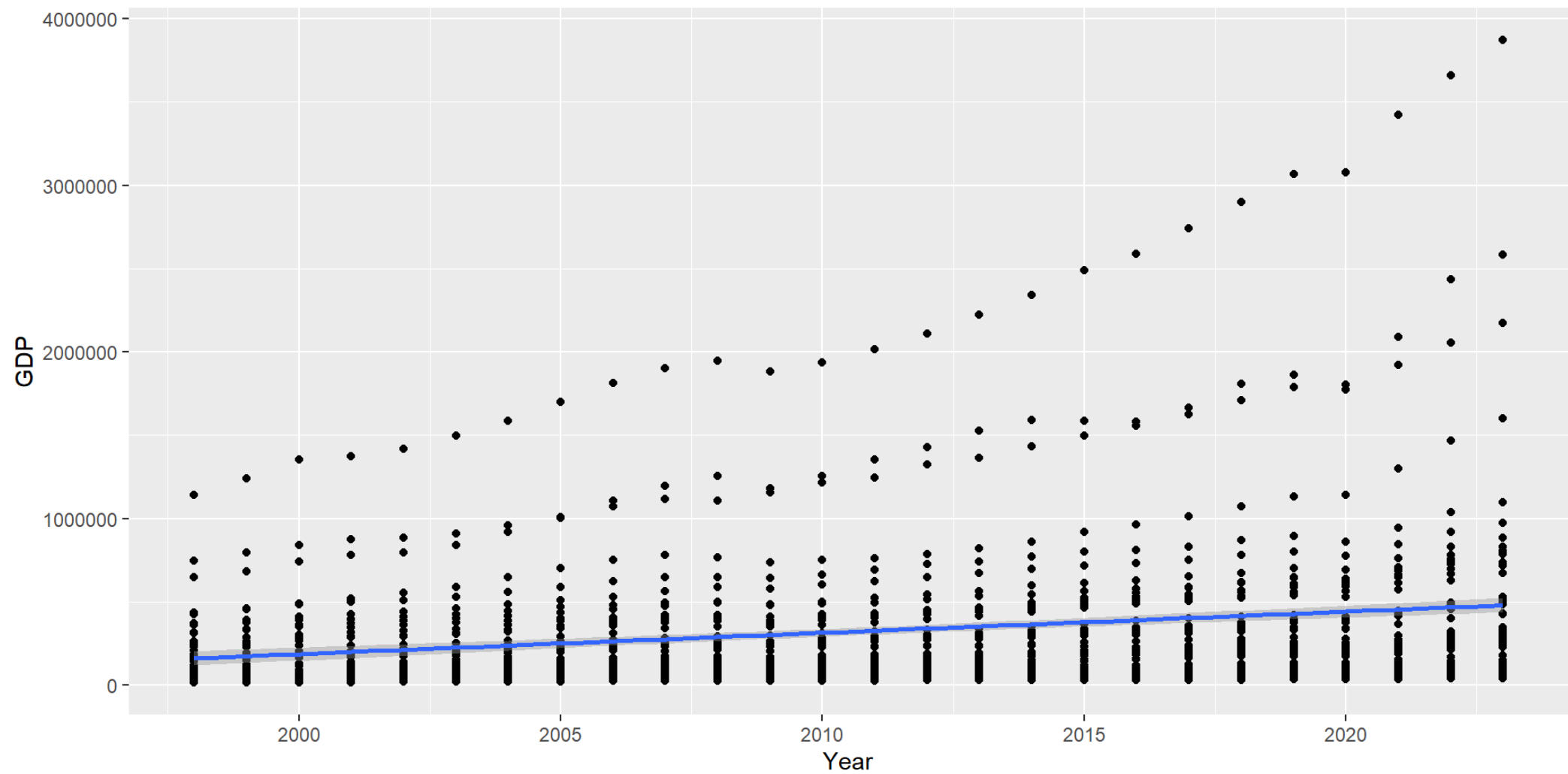
CO2 and GDP Relationship

To determine if there is a relationship between the CO2 emitted from energy production in the respective state, and the state GDP, a scatter plot can be used to visualize the linear relationship.

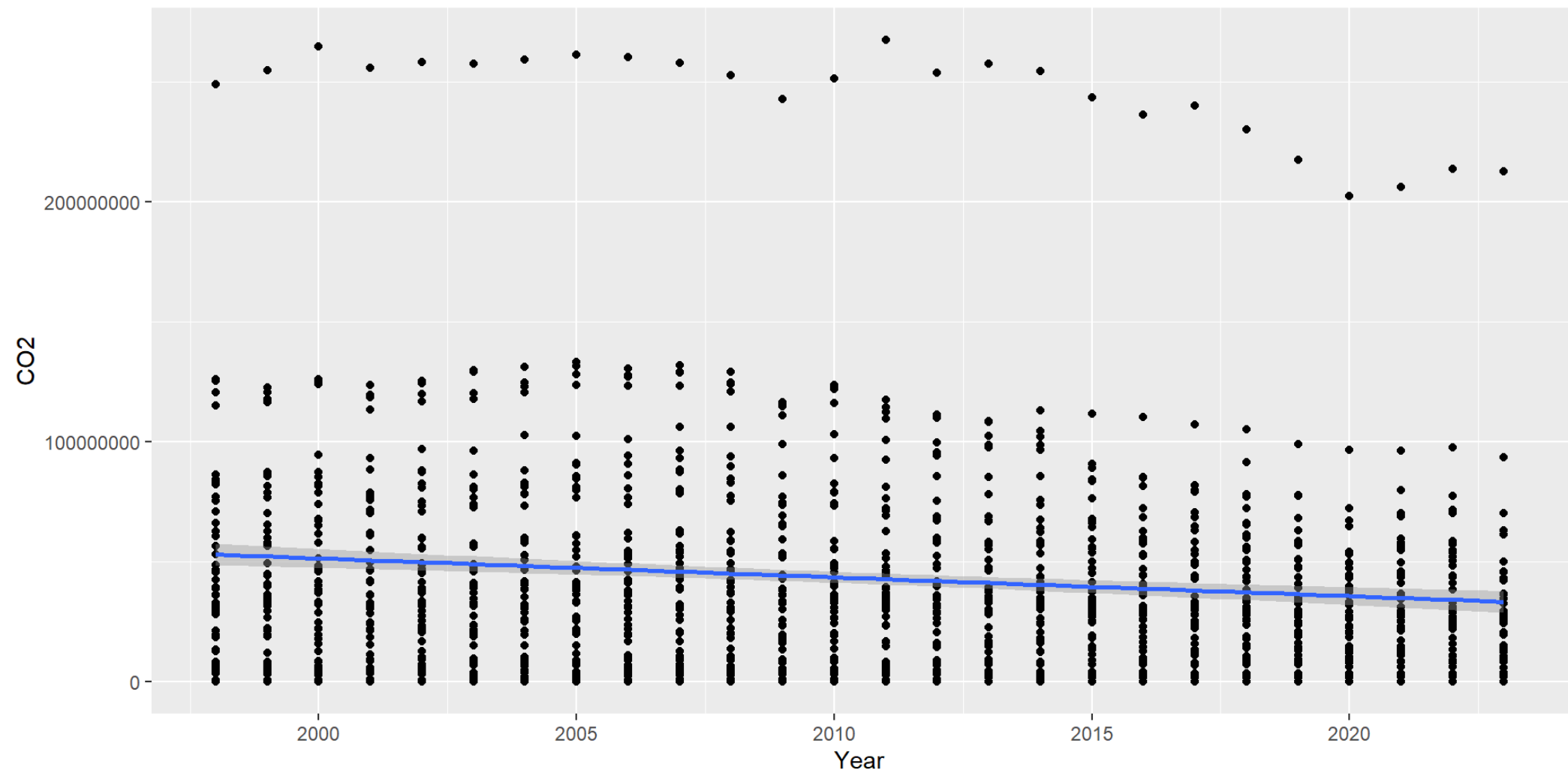
CO2 and GDP Scatterplot



GDP and Year Scatterplot



CO2 and Year Scatterplot

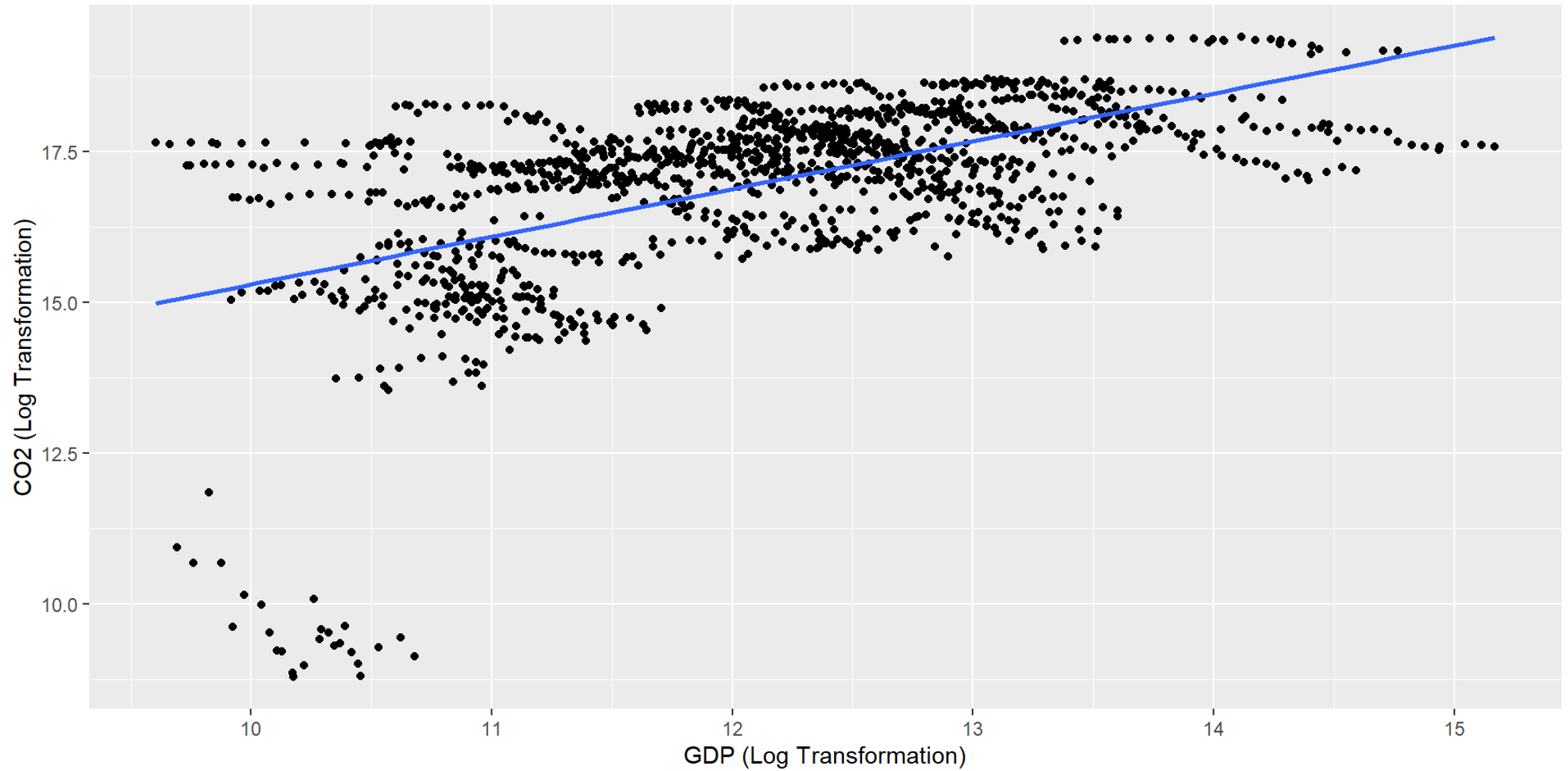


Dealing with Skewness and Outliers

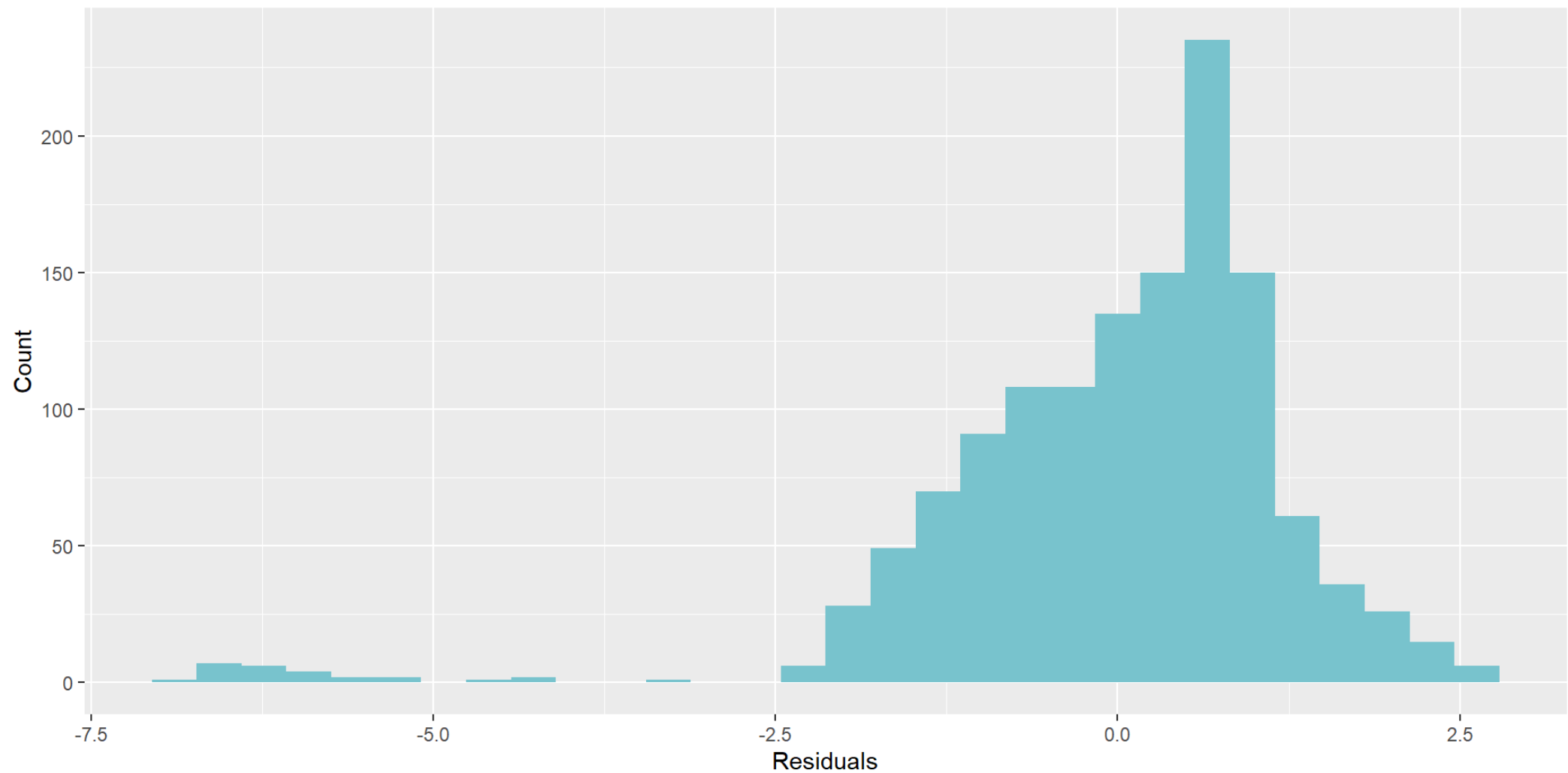
Texas and California seem to be outliers in this dataset. They have very large GDPs, and emit more CO2 than the other States. The data is also positively skewed. Due to this, a log transformation may be necessary to run a linear regression.

```
1 trans_state_gdp_emissions <- state_gdp_emissions |>
2   mutate(log_gdp = log(GDP),
3          log_CO2 = log(CO2))
4
5 log_m1 <- lm(log_CO2 ~ log_gdp, data = trans_state_gdp_emissions)
6 log_m2 <- lm(log_CO2 ~ log_gdp + Year, data=trans_state_gdp_emissions)
```

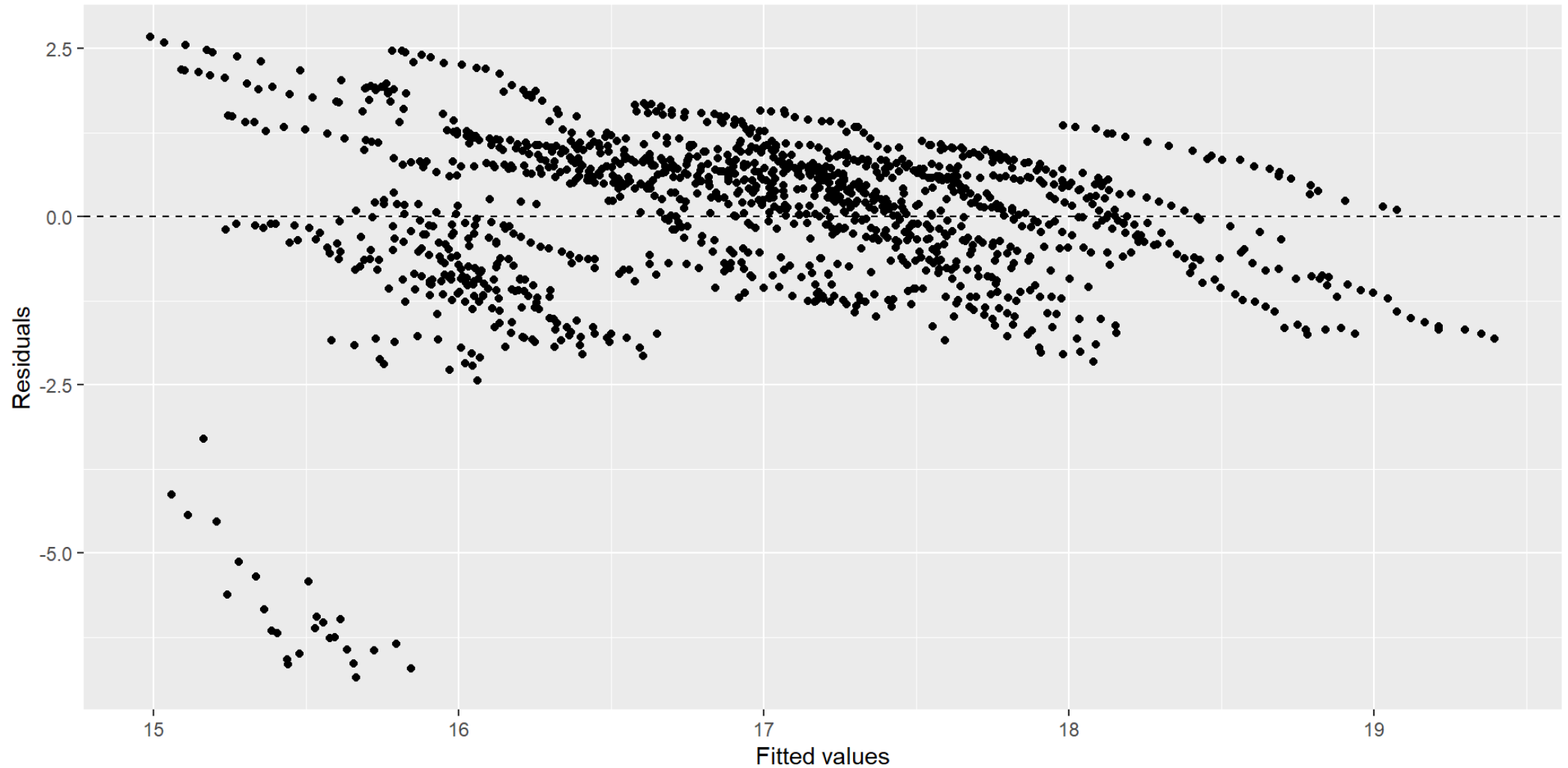
Scatterplot with Least Squares Line



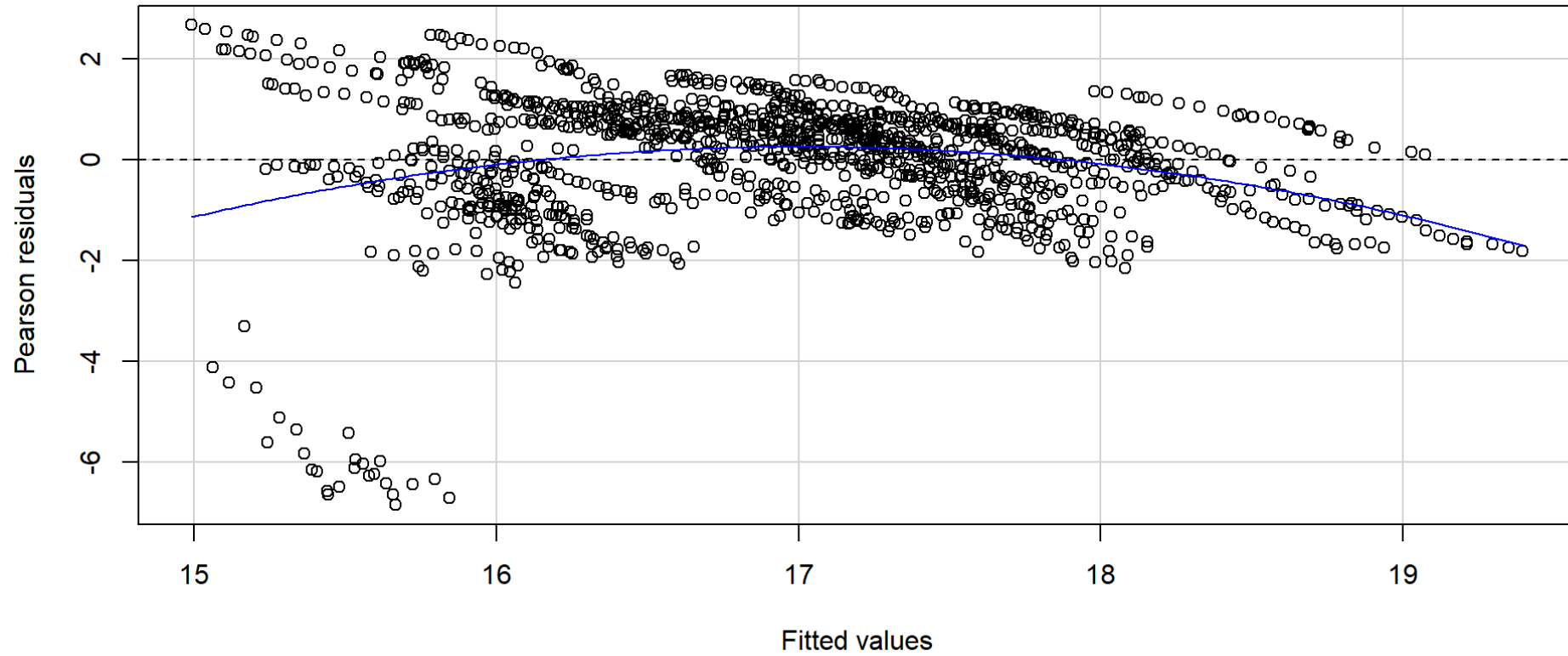
Residuals Histogram Plot



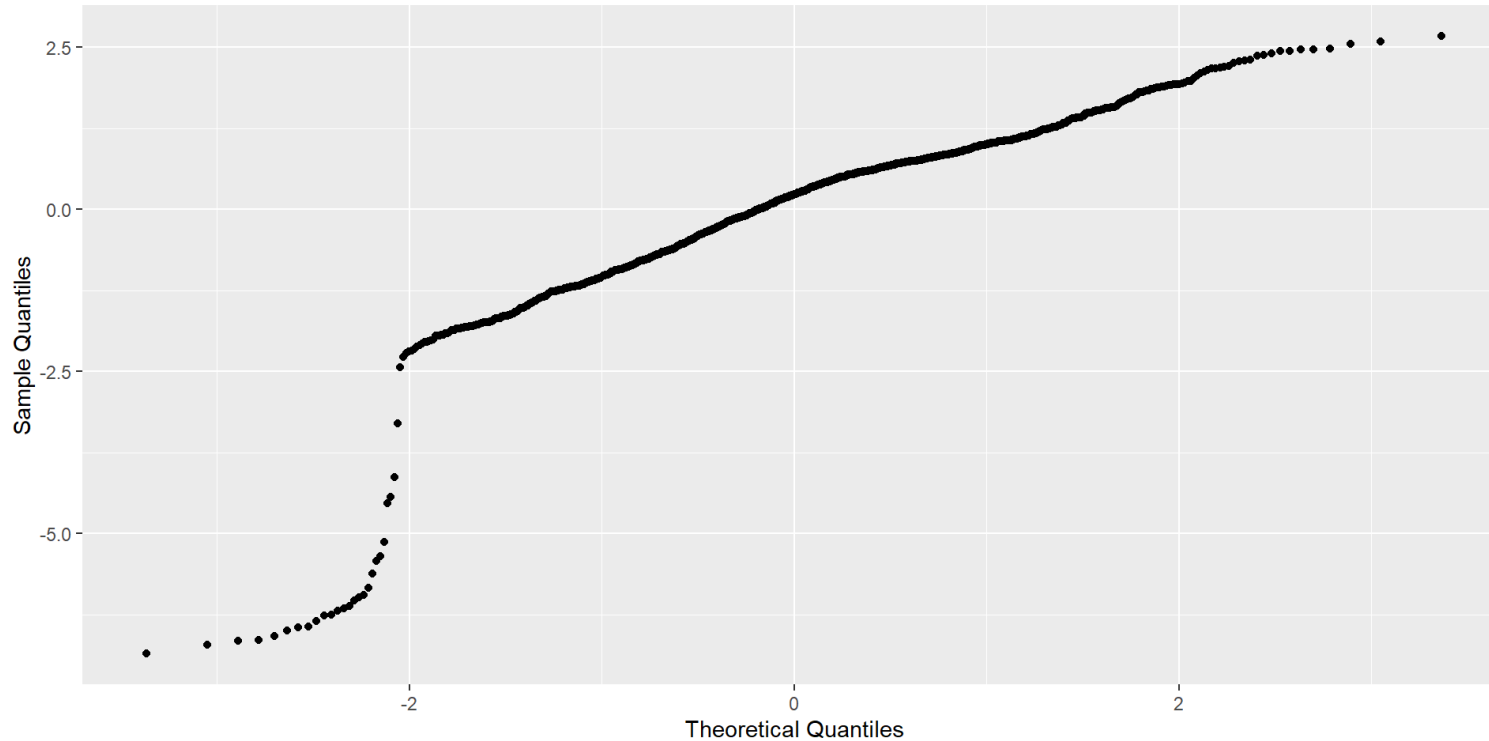
Residuals vs. Fitted (Predicted) Values



Residuals vs. Fitted (Predicted) Values with Curvature Test



Quantile-Quantile Plot



The non-linear Q-Q plot with the heavy tail indicates that there may be skewness in the residuals. This signals that the residuals are not normally distributed.

Predict CA CO2 Emissions from the Model for 2023

```
1 real_CA <- trans_state_gdp_emissions |>
2   filter(`Year` == 2023,
3         `State` == "CA")
4
5 real_CA$CO2 - exp(predict(log_m2, real_CA))
```

1

-139442428

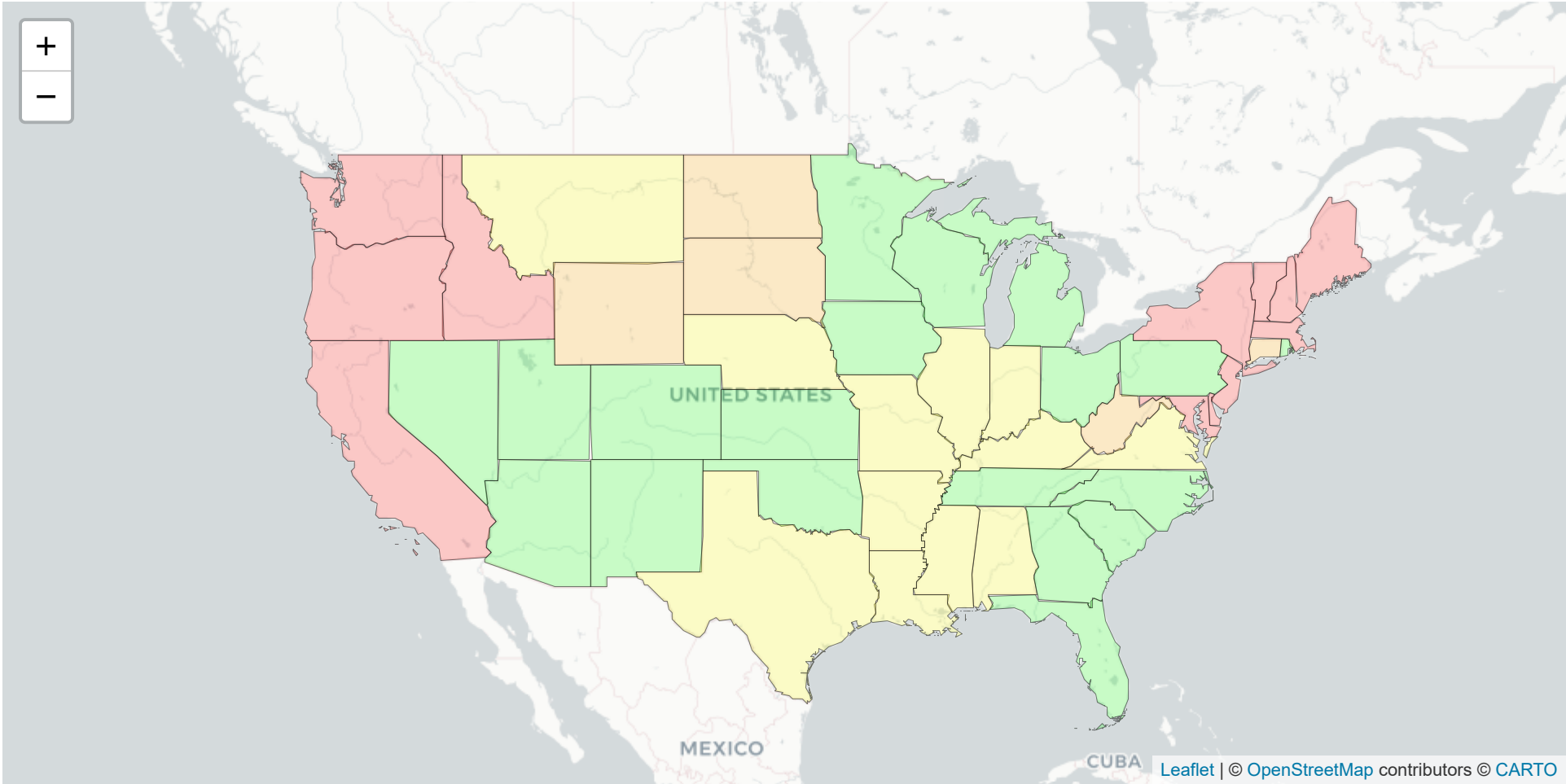
Predict PA CO2 Emissions from the Model for 2023

```
1 real_PA <- trans_state_gdp_emissions |>
2   filter(`Year` == 2023,
3         `State` == "PA")
4
5 (real_PA$CO2 - exp(predict(log_m2, real_PA)))/real_PA$CO2
```

1

0.2417473

Predicted vs. Real Values for 2023



References

1. DOC/BEA's GDP and Personal Income by State
2. DOE/EIA's Electric Power Industry Estimated Emissions by State
3. Electric Power Annual Report

Additional Notes

Producer Types (Business Classification): The EIA classifies Producer Types on page 227 of the Electric Power Annual 2023 Annual Report. On slide seven the Energy Producer Types were referenced.

Energy Source: The EIA defines renewable energy within the footnote section on page 33 of the Electric Power Annual 2023 Annual Report. On slide six Energy Sources were referenced.