Is Energy in Synergy with Life

Bhawneet Singh, Karan Dhir, Nikhil Gupta, Poonam Parag Thakur, Sarah Edwards

9/14/2021

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.3     v dplyr   1.0.7
## v tidyr   1.1.3     v forcats 0.5.1
## v readr   2.0.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

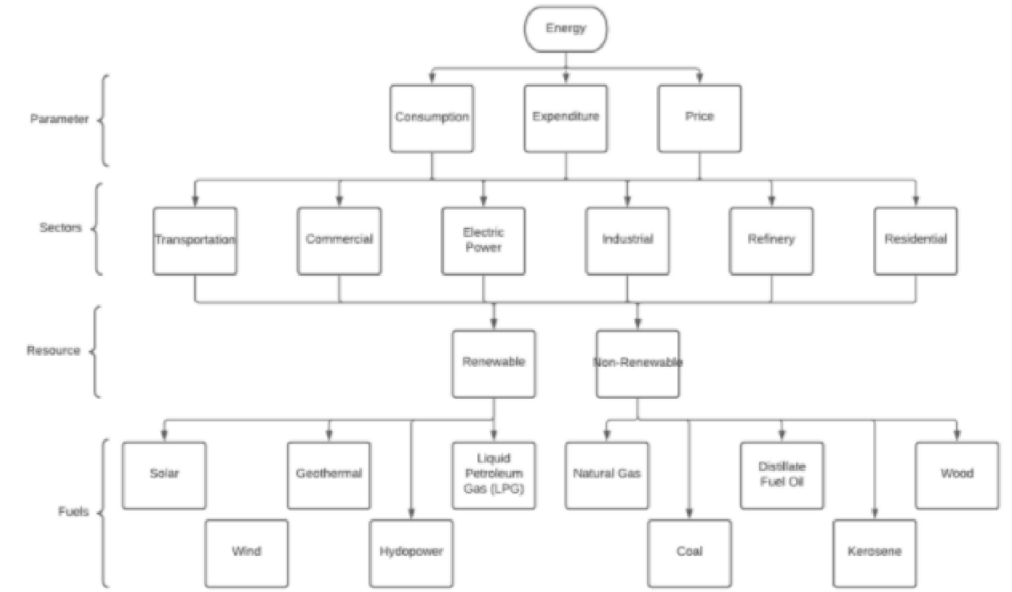

Abstract

There has been extreme global climate crises throughout the years. Ranging from record breaking heatwaves to the glaciers melting at an advance pace in the Arctic. The United States has made many head lines over the years for their extreme carbon emissions and its impact to global warming. A big reason behind the United State's carbon footprint is due to the immense amount of energy used. Behind China, the United States is the second country with the most energy consumption. To help regulate the amount of global energy consumption, the Paris Climate Agreement was made. Over 191 countries today have pledge to reduce their carbon dioxide. One of the most efficient ways to do this is to use more renewable resources, such as water and wind, over non-renewable resources, such as coal and natural gases. The United States has recently rejoined this agreement but what still stands to question is how well the United States was doing on their renewable and non-renewable energy consumption before joining the first time back in 2016. Knowing that industrial energy consumption is one of the bigger contributes to the total energy consumption, this analysis will explore what states have done the best when it comes to using renewable resources and the what states have done the worst when it comes to using non-renewable resources.

The Data

This analysis will take a look at the [corgis]https://think.cs.vt.edu/corgis/csv/energy/ dataset on energy consumption in the United States form 1960-2014. The dataset is composed of government reported energy consumption by billion BTU(british thermal units), expenditure by million dollars, and price by dollar per billion BTU. Ranging from commercial energy use to residential energy use, the fuels in the dataset can be categorized into both renewable and non-renewable resources. This flow chart describes to full detail how energy is split among the variables.

```
knitr::include_graphics("Energy_Flow_Chart.png")
```



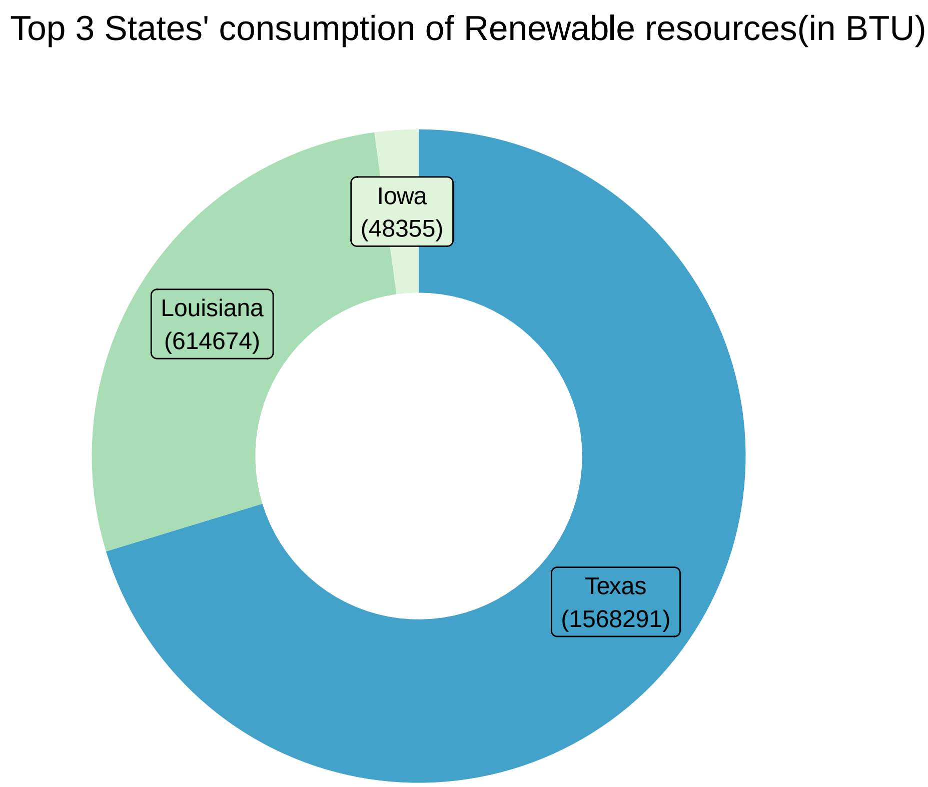
Exploring the Data

When importing the data to distinguish the renewable and non-renewable resources we created arrays for the categories where each type of resource will be combined into one table. Since we are also only interested we are making sure to only include the fuel types within the industrial section of the dataset. Then using the resources within their arrays we will create two new columns in the original energy table that will be the total consumption of each specific fuel consumption.

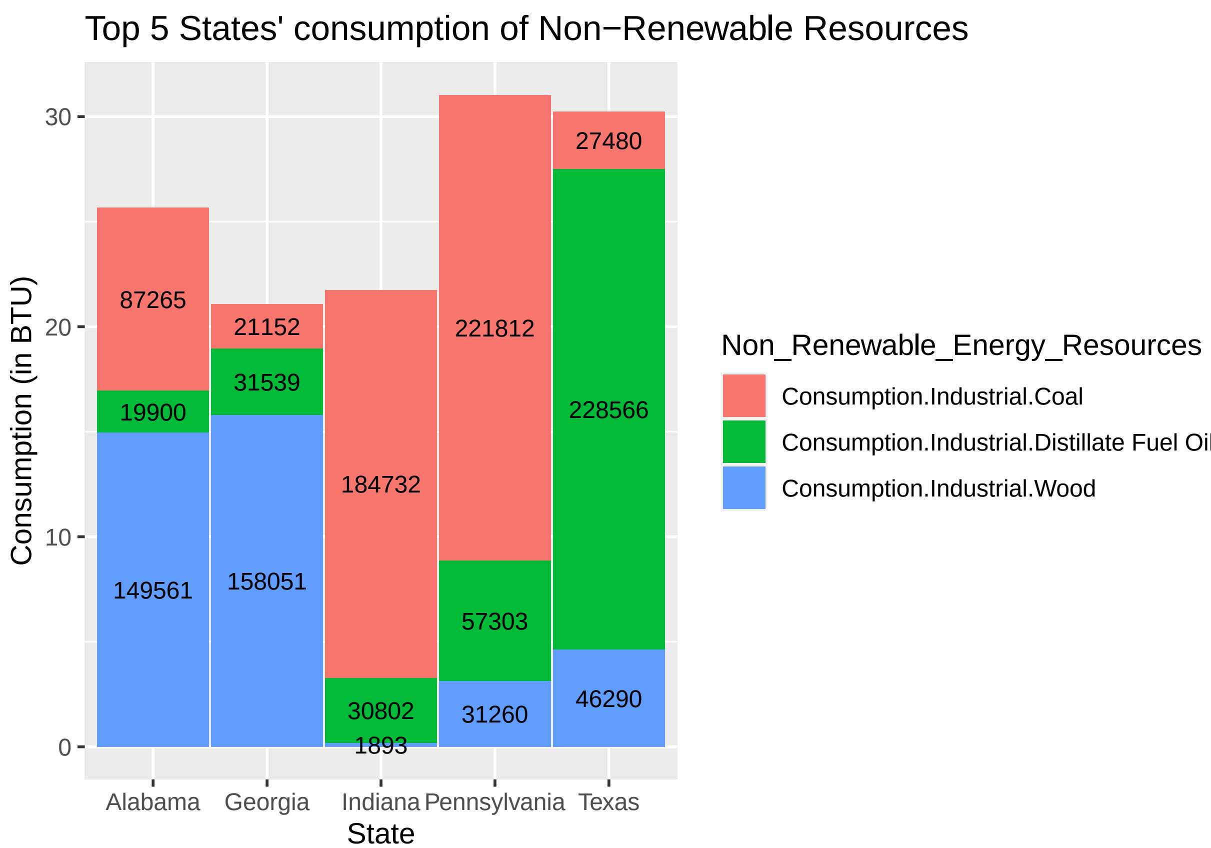
```
# Reading the energy dataset
energy_data <- read_csv("https://think.cs.vt.edu/corgis/datasets/csv/energy/energy.csv",show_col_types =
energy_df <- as_tibble(energy_data)

# Creating the arrays for renewable and non renewable resources
consumption_energy_renewable_industrial <- c('Consumption.Industrial.Solar',
                                             'Consumption.Industrial.Wind',
                                             'Consumption.Industrial.Geothermal',
                                             'Consumption.Industrial.Hydropower',
                                             'Consumption.Industrial.Liquefied Petroleum Gases')
consumption_energy_non_renewable_industrial <- c('Consumption.Industrial.Coal',
                                                 'Consumption.Industrial.Distillate Fuel Oil',
                                                 'Consumption.Industrial.Kerosene',
                                                 'Consumption.Industrial.Wood')

# Data Manipulation for Visualization
energy_mutate_df <- energy_df %>%
  mutate(Total_Consumption_renewable_industrial =
           rowSums(energy_df[consumption_energy_renewable_industrial]),
Total_Consumption_non_renewable_industrial =
  rowSums(energy_df[consumption_energy_non_renewable_industrial]))
```

With the data imported and correctly sorted based on fuel type, we can begin create visualize and do an analysis of the data. For our first question, what states have done the best using renewable energy, we will look at the top three states who were doing the best at the year before joining the Paris Agreement, 2014. We will be basing the top 3 states based on which state was using the most renewable resources. We also want to know the top five states that have done to worst when it comes to non-renewable fuel consumption. Again we will be biasing the top states who used the most non-renewable fuel over all fuel types. Since using non-renewable resources we can look into how much each state is using each type of non-renewable fuel. Specifically we will look at how much coal, distillate fuel oil, and wood resources, the top 5 states are using.

```
# Manipulating the data for Consumption of Renewable Resources
# in Industrial sector(Donut Chart)
top3_Consumption_renewable_industrial <- energy_mutate_df %>% filter(Year == 2014) %>% slice_max(Total_C
  select('State','Year','Consumption.Industrial.Solar',
         'Consumption.Industrial.Wind',
         'Consumption.Industrial.Geothermal',
         'Consumption.Industrial.Hydropower',
         'Consumption.Industrial.Liquefied Petroleum Gases',
         'Total_Consumption_renewable_industrial',
         'Total_Consumption_non_renewable_industrial')

# Visualizing the data in the form of Donut chart
data <- top3_Consumption_renewable_industrial
data_per_thou <- data$Total_Consumption_renewable_industrial
data$fraction <- data_per_thou / sum(data_per_thou)
# Compute the cumulative percentages (top of each rectangle)
data$ymax <- cumsum(data$fraction)
# Compute the bottom of each rectangle
data$ymin <- c(0, head(data$ymax, n=-1))
# Compute label position
data$labelPosition <- (data$ymax + data$ymin) / 2
# Compute a good label
data$label <- paste0(data$State,"\n(", data_per_thou,")")

# Make the plot
 top3_states_rene_chart <- ggplot(data, aes(ymax=ymax, ymin=ymin, xmax=4,
                                            xmin=3,
                                            fill=State)) +
  geom_rect() +
  geom_label( x=3.5, aes(y=labelPosition, label=label), size=3) +
  scale_fill_brewer(palette=4) +
  coord_polar(theta="y") +
  xlim(c(2, 4)) + ggtitle("Top 3 States' consumption of Renewable resources(in BTU)")+
   theme_void() +
  theme(legend.position = "none")
top3_states_rene_chart
```



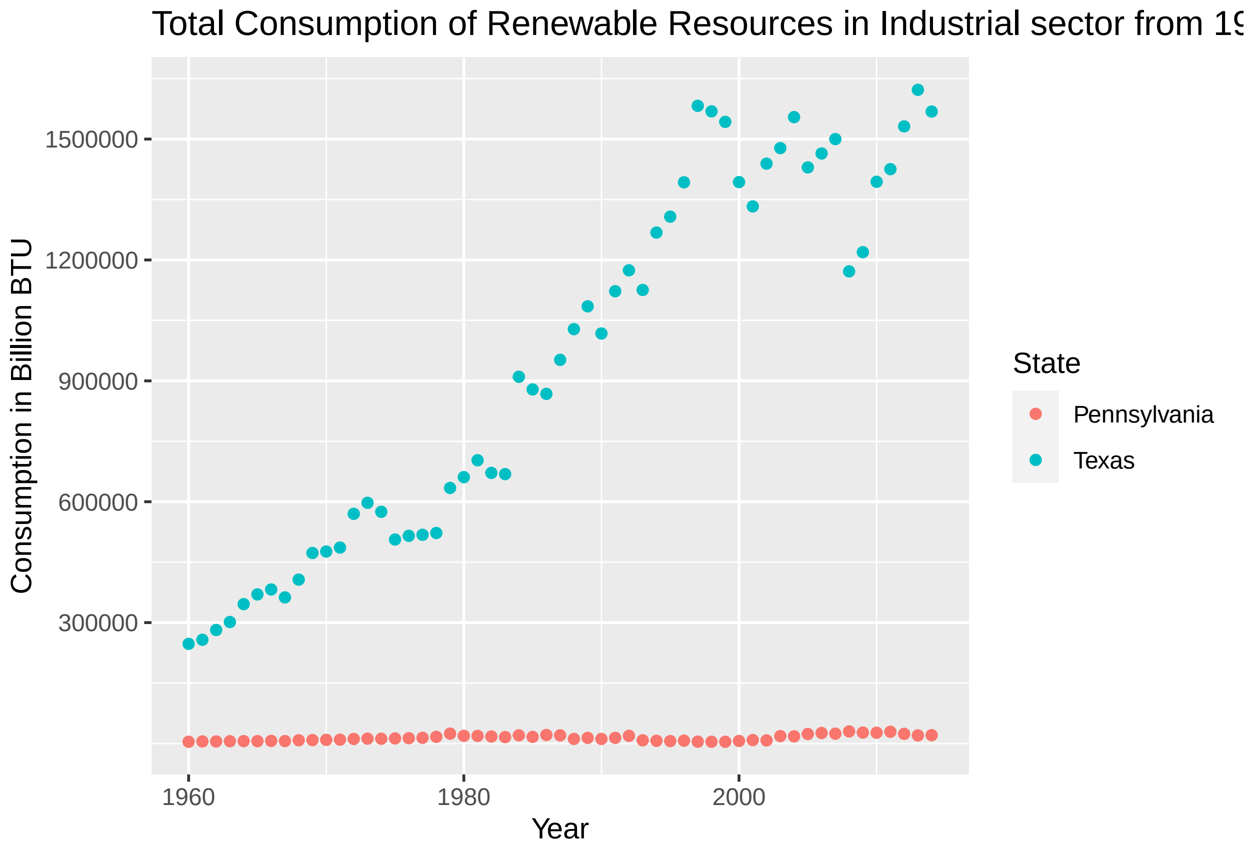

```
# Manipulating the data for Consumption of Non-Renewable Resources in Industrial sector
top5_Consumption_non_renewable_industrial <- energy_mutate_df %>% filter(Year == 2014) %>%
  slice_max(Total_Consumption_non_renewable_industrial, n = 5) %>%
  select('State','Year','Consumption.Industrial.Coal',
         'Consumption.Industrial.Distillate Fuel Oil',
         'Consumption.Industrial.Wood')

# Visualizing the data in the form of Stack Bar chart
pivot_data <- top5_Consumption_non_renewable_industrial %>%
    pivot_longer(cols = -c(State, Year),
                 names_to = "Non_Renewable_Energy_Resources",
               values_to = "Consumption")
top5_states_non_rene_chart <- ggplot(pivot_data, aes(x=State, y = Consumption/10000,
                                                     fill = Non_Renewable_Energy_Resources,
                                                     label = Consumption ))+
  geom_bar(width = 0.98, stat="identity") +
  geom_text(size = 3, position = position_stack(vjust = 0.5))+ xlab("State")  +
  ylab("Consumption (in BTU)") +
  ggtitle("Top 5 States' consumption of Non-Renewable Resources")
top5_states_non_rene_chart
```



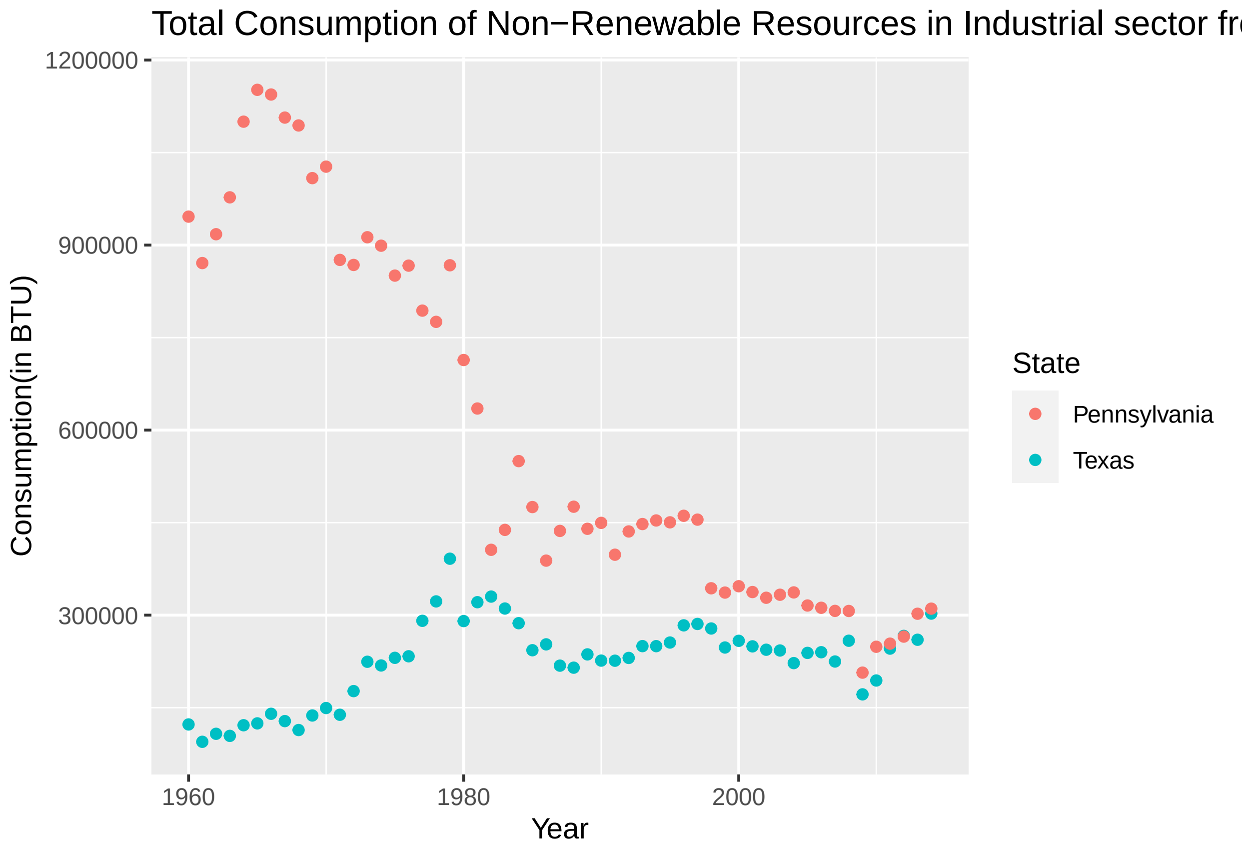
The state using the most renewable energy as in 2014 is Texas and the state using the most non-renewable energy is Pennsylvania. Coal is the biggest fuel source Pennsylvania is using out of the three types we looked at. Fuel is one of the more environmentally harmful fuel sources. This raises a few concerns considering Pennsylvania is already the top state using the most non-renewable resources. It is also important to note that even though Texas is using the most renewable resources they are still one of the top states using the most non-renewable resources. For now however we will focus on how these to states are using renewable fuel types since those are the best way to decrease to carbon emissions.

Lets take a closer look and analyze how both these states used non-renewable and renewable fuels between 1960 to 2014.

```
## Analysis of Energy Dataset
#  Who are the ideal and Worst States in consumption of energy resources?
states <- c("Texas", "Pennsylvania")
rene_non_rene_state <-filter(energy_mutate_df, State  %in% states) %>%
  select('State','Year','Total_Consumption_renewable_industrial',

# Gathering the data for both renewable and non-renewable resources usage states
# Comparing the two states based upon their total consumption of all renewable resources
c_rene_indust_graph <- ggplot(rene_non_rene_state) +
  geom_point(aes(x=Year,y=Total_Consumption_renewable_industrial,
                 fill=State,color=State),stat="identity") +
   scale_y_continuous(breaks=c(300000,600000,900000,1200000,1500000,1800000)) +
  ggtitle("Total Consumption of Renewable Resources in Industrial sector from 1960-2014") +
  xlab("Year") +  ylab("Consumption in Billion BTU")
c_rene_indust_graph
```




```r
# Comparing the two states based upon their total consumption of all non_renewable resources
c_non_rene_indust_graph <- ggplot(rene_non_rene_state) +
  geom_point(aes(x=Year,y=Total_Consumption_non_renewable_industrial,
                fill=State,color=State),stat="identity") +
  scale_y_continuous(breaks=c(300000,600000,900000,1200000,1500000,1800000)) +
  ggtitle("Total Consumption of Non-Renewable Resources in Industrial sector from 1960-2014") +
  xlab("Year") +  ylab("Consumption(in BTU)")
c_non_rene_indust_graph
```



Now we can see that throughout the years remained to consume a low level of non-renewable fuels and have continued to increase the amount of renewable fuel consumed. Even though the total energy appears to be increasing it is good that Texas has been using renewable energy versus non-renewable. Pennsylvania on the other hand has continued to have very low renewable fuel consumption. However the state has significantly reduced the amount of non-renewable resources consumed.

Taking a closer look at Pennsylvania we can use a linear model to see how much exactly the state's non-renewable energy as decreasing over the years and create a prediction on how Pennsylvania's non-renewable industrial consumption would look like even without the Paris Agreement.

```r
## Linear model for consumption of Non-Renewable source of energy for
##   Pennsylvania over the period from 1960-2025.

rene_non_rene_state_penn <- rene_non_rene_state %>% filter(State == 'Pennsylvania')
model <- lm(Total_Consumption_non_renewable_industrial ~ Year,data = rene_non_rene_state_penn)
summary(model)
```

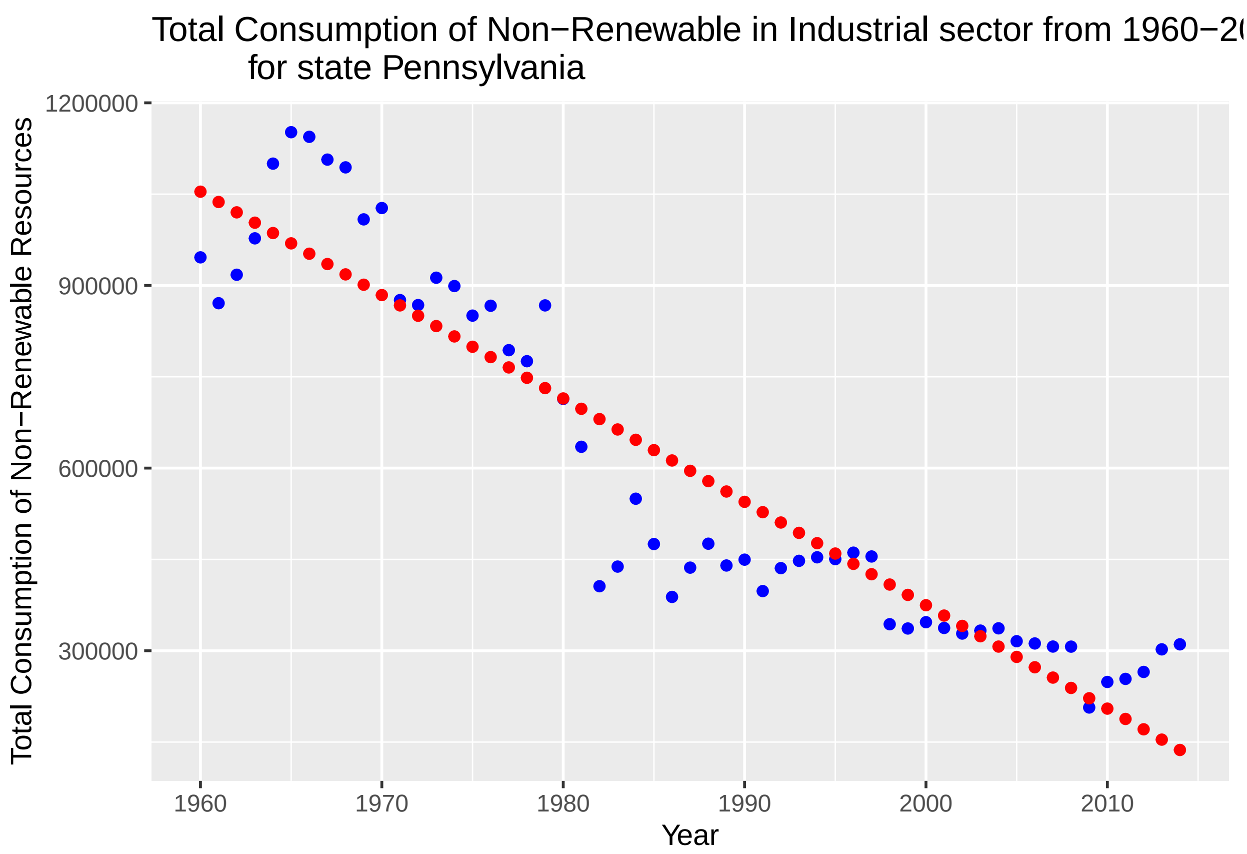
```
##
## Call:
## lm(formula = Total_Consumption_non_renewable_industrial ~ Year,
##     data = rene_non_rene_state_penn)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -274487  -70052    9284   73653  191986
##
```

```
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34339265.6  1876308.4   18.30   <2e-16 ***
## Year          -16982.3      944.3  -17.98   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 111200 on 53 degrees of freedom
## Multiple R-squared:  0.8592, Adjusted R-squared:  0.8566
## F-statistic: 323.4 on 1 and 53 DF,  p-value: < 2.2e-16
```

```
x_years <- seq(1960, 2025)
new_df <- tibble(years = x_years)
rene_non_rene_state_pred <- rene_non_rene_state_penn %>% mutate(pred = predict(model))
rene_non_rene_plot <- rene_non_rene_state_pred %>% ggplot() + geom_point(aes(x = Year, y = Total_Consum
  scale_x_continuous(breaks=c(1960,1970,1980,1990,2000,2010,2020)) +
  geom_point(aes(x = Year, y = pred), color = "red") +
  scale_x_continuous(breaks=c(1960,1970,1980,1990,2000,2010,2020)) +
xlab("Year") +  ylab("Total Consumption of Non-Renewable Resources") +
  ggtitle("Total Consumption of Non-Renewable in Industrial sector from 1960-2025
          for state Pennsylvania")
```

```
## Scale for 'x' is already present. Adding another scale for 'x', which will
## replace the existing scale.
```

```
rene_non_rene_plot
```



With this model we can see that Pennsylvania was decreasing in non-renewable energy resources by 16,982.3 billion BTU per year. Using the linear model we can build an equation of y=-16982.3x+34339265.6, where y is the total amount of non-renewable resources in billion BTU and x is the years, we can predict Pennsylvania would have kept lowering their non-renewable resources despite the United States joining the Paris Agreement.

Conclusion

Before the Paris Agreement the world had a loser control how much carbon emissions were being consumed. A big factor into carbon emissions is the consumption of fuel sources throughout the world. The United States has been known for their extreme use in energy consumption which would result in higher carbon emissions. The corgis dataset provides data on the United States energy consumption, expenditure, and prices over the years 1960 to 2014. Using this data we were able to explore how the United states was consuming non-renewable and renewable resources per state.

The state consuming the most renewable fuel was Texas and the state consuming the most non-renewable fuels was Pennsylvania in 2014. While using renewable resources is reduces the amount of carbon significantly and it is a positive sign that Texas had a high consumption in renewable energy it is important to recognize that Texas was still one of the top consumers in non-renewable energy. It would not be justifiable to say Texas was doing the best overall in energy consumption due to their high non-renewable consumption.

With the top state in non-renewable energy consumption and renewable energy consumption we began to analyze how the states were consuming these two types of energy resources over the years of 1960 to 2014. Though the state has high consumption numbers, Texas remained constant in their non-renewable energy but was increasing in the consumption in renewable energy over the years. Since the renewable energy consumption was increasing it could illiterate a sign that Texas began to favor renewable energy with increasing total energy usage. Pennsylvania however, remained to have low numbers in renewable resources but had decreasing consumption in non-renewable resources. Our linear model showed that the the state had a decreasing non-renewable energy consumption rate of 16,982.3 billion BTU per year and had the trajectory of continuing to decrease their consumption rate past 2014, without the US being apart of the Paris Agreement. Despite being apart of the Paris Agreement the United States did appear to be on its way to creating a more energy environmentally friendly atmosphere within the industrial sector, weather that be states using less non-renewable resources and/or using more renewable resources.

This conclusion does come with biases, however. The biggest bias is the last year recorded in the dataset is in 2014. Over past recent years the US could have had different energy consumption, especially being in the Paris Agreement. This could result in a completely different result than the ones drawn with this analysis. We also mainly looked at two states in the analysis. Other states could have very different industrial energy consumption and that would indicated a different conclusion. There was also missing data for other renewable resources and not all columns were used in this report.Moreover, we did not consider how the expenditure and price affected the consumption which could have played a considerable role in how and why the states consumed their fuel. Lastly we only considered the industrial data, other sectors again can have very different outputs and result in a different conclusion.

The next steps to be taken with this analysis is to look into these basis and do a deeper analysis than what was done here. To look into this analysis our report can be found on github, https://github.com/karan1dhir/CU-Boulder-DataScienceAsField.git.