# EEB C177 Final Project

Alexander Phu

February 29, 2020

## Abstract

CO2 Emissions in the United States are increasing each year. It is important that we monitor the release of CO2 Emissions as CO2 Emissions affect many health related risks.

# Contents

# List of Figures

# 1 Introduction

For this project, I am analyzing the CO2 Emissions of different fossil fuels from the United States in metric tons. This data is important to analyze because as CO2 emissions increase, there are increased health risks and issues that affect entire communities (Parshall et. al, 2009). Some of these health risks include increased cardiovascular disease, increased risk of asthma, and increased cancer risks (Gerber et. al, 2013). This makes this research also biologically relevant because if we can analyze the amount of CO2 that is in the air in a quantifiable way, we can create ways to reduce emissions in consumption of specific fossil fuels.

Present the issue: Understanding CO2 emissions from different fossil fuels from the United States. CO2 emissions has the potential to cause a variety of health defects in in the community.

Purpose: To assess historical trends between CO2 emissions and the type of fossil fuel used in the United States in order to track historical trends and create methods to improve fossil fuel use and consumption.

Methods: Obtain data on the fossil fuel consumption separated by fossil fuel type that the United States used each year.

Major Findings: There is a trend that although fossil fuel consumption increases each year, it is not linear and certain fossil fuel types are used more depending on historical time period.

Biological Impact: The use of fossil fuels creates pollutants and over time, we will reach the ecosystem threshold in which the ecosystem can sustain CO2 emissions.

# 2 Methods

## 2.1 Importing the data file

Public data on CO2 Emissions from the United States found online in the form of a csv file. Data includes CO2 emissions from the beginning of the United States until 2018. The data also includes specific breakdowns of the type of gas emission fuel used. A function was defined within a python script to open the csv file, import the data using the pandas library, and convert the data to a numpy file for future numerical analysis.

```
import pandas as pd
import numpy as np

def np_from_csv(csv_file):
    temp_data = pd.read_csv(csv_file, header = None)
    data = temp_data.to_numpy()
    return data

data = np_from_csv('EEB C177 Revised Data Set.csv')
```

## 2.2 Analysis of the total amount of CO2 emitted by the United States in totality.

A function was defined to calculate the total CO2 emitted by the United States for any given category. Given an input of the column that would like analyzed, a for loop as well as a .split function is used to cut that specific data and then add the values of each data element in the column together. This is then outputted to give the total CO2 emission of a given fossil fuel type.

```
1  def total_CO2():
2      total=0
3      for line in open('EEB C177 Revised Data Set .csv'):
4          csv_row=line.split()
5          temprow=csv_row[0].split(",")
6          total += int(temprow[2])
7
8      print(total)
9      return total
```

A second function takes an input of the columns and compares the differences in CO2 emissions based on the columns given in the form of a plot. The plot function of the matplotlib.pyplot library is used to graph this relationship between the columns.

```
1  import matplotlib.pyplot as plt
2
3  def plot_column__vs_column(str(x), str(y)):
4      x=input( "What column would you like to compare as x value (exact name of column):")
5      y=input("What column would you like to compare as y value (exact name of column):")
6      plt.plot(str(x), str(y))
7      plt.axis([0, 12, 0, 100])
8      z= input("What is the name of the x axis:")
9      plt.xlabel(str(z))
10     w=input("What is the name of the y axis:")
11     plt.ylabel(str(w))
12     v=input("What is the name of the plot:")
13     plt.title(str(v))
14     plt.show()
15     return plot_column_vs_column (str(x), str(y)):
```

A third function takes the csv file and will count the number of occurrences of years in the csv file. This will help determine the historical impact of CO2 Emissions. This function utilizes a for loop that will count the amount of occurrences in the csv file and store that count within a variable and print out the variable.

```
1
2  with open('EEB C177 Final Project .csv', mode = 'r', encoding = 'utf-8-sig') as csvfile:
3      reader = csv.DictReader(csvfile)
4      occurrences = 0
5      for row in reader:
6          occurrences = occurences +1
7      print('we have {} occurrences of values to a key.\n'.format(occurrences))
```

## 2.3  Using Regex

A list reflecting the CO2 emissions in totality by year was created in order to gain a better sense of the total impact that the United States has on CO2 Emissions. The data set was imported as a txt file. The list was then analyzed to see where what the total value was in the list using regex functions. The re.findall function was used to analyze all of the integers in the list.

```
1
2  import csv
3  import re
4
5  def extractMax("total_CO2_Emisisons.txt"):
6    numbers = re.findall ('\d+',input)
7    numbers = map(int,numbers)
8    print max(numbers)
9
10 if  max  == "__main__":
11     extractMax("total_CO2_Emissions.txt")
```

## 2.4 Analysis of history of total CO2 Emission and Break Down of Carbon Fuels through United States History

The dataset was imported in RStudio and the libraries dplyr and ggplot2 were installed for data analysis and visualization, respectively. In order to evaluate the relationship between income status and distance from the supermarket, the dataset was filtered to include only the census tracts flagged for low income status.

```
1  install.packages(ggplot2)
2  library(ggplot2)
3  library(dplyr)
4  #Before graphing using our assigned data, we have to import the ggplot2 and dplyr libraries.
5  #This will allow me to utilize the necessary functions to make a line graph to show my data
6  Final_Data_2 <- read.csv("~/Desktop/EEB C177 Final Revised Data Set .csv", header=FALSE)
7  #This line helps me read in my data file and save it as "Final_Data_2". By having the header=
       FALSE, the
8  #header will not be taken into account when making the graphs.
9  p1 <- ggplot() + geom_line(aes(y = V2, x = V1), data = Final_Data_2) + scale_x_continuous(breaks=
       seq(1800,2014,20)) + theme(text=element_text(family="Tahoma"))
10 #In this line of code, I define p1 as a ggplot in which I am setting my y axis to hold all of the
        measurements
11 #from V2 and the x axis to hold all my measurements from column V1. Additionally, I set the x
       axis breaks to
12 #the range of the data in years and a divider every 20 years.
13 p1 + labs(title = "Total CO2 vs Year", x = "Year", y = "Total CO2 Emissions (Metric Tons)")
14 #In this line of code, I added axis titles and a graph title to the graph that was made from the
       prior code
15 p2 <- ggplot() + geom_line(aes(y = V5, x = V1), data = Final_Data_2) + scale_x_continuous(breaks=
       seq(1800,2014,20)) + theme(text=element_text(family="Tahoma"))
16 p2 + labs(title = "Total Gas Fuel Consumption vs Year", x = "Year", y = "Total Gas Fuel
       Consumption (Metric Tons)")
17 p3 <- ggplot() + geom_line(aes(y = V3, x = V1), data = Final_Data_2) + scale_x_continuous(breaks=
       seq(1800,2014,20)) + theme(text=element_text(family="Tahoma"))
18 p3 + labs(title = "Total Solid Fuel Consumption vs Year", x = "Year", y = "Total Solid Fuel
       Consumption (Metric Tons)")
19 p4 <- ggplot() + geom_line(aes(y = V4, x = V1), data = Final_Data_2) + scale_x_continuous(breaks=
       seq(1800,2014,20)) + theme(text=element_text(family="Tahoma"))
20 p4 + labs(title = "Total Liquid Fuel Consumption vs Year", x = "Year", y = "Total Liquid Fuel
       Consumption (Metric Tons)")
21 #Did the same code to make another line graph using other column variables for comparison.
```

# 3 Results

# 4 Discussion

# 5 References

# 6 Figures