Global Warming, Global Warning

Lyu Ruiqi, Yao Zhiqian, Zhao Chengyang

Abstract-Global warming has been increasingly serious in recent years. Among the greenhouse gases, carbon dioxide in the atmosphere is the main cause of global warming. To warn people and alert the world to face the crisis that global warming has caused, we intend to show the carbon dioxide emission and temperature trend of each country in the past decades and classified them into different groups based on various emission indexes and methods. On the other hand, predictive analysis will be employed to forecast the future global temperature and carbon dioxide emission by using multiple models. After comparison and evaluation, the best fit one will be eventually selected.

I. Introduction

Over the past 50 years, the average global temperature has increased at the fastest rate in history. Among the greenhouse gases, the carbon dioxide in the atmosphere is the main cause of global warming. When people burn fossil fuels like coal, oil and natural gas, vast amount of carbon dioxide is being pumped into the air, and It will absorb sunlight and solar radiation that have bounced off the earth's surface. Normally, this radiation would escape into space—but these pollutants, which can last for years to centuries in the atmosphere, trap the heat and cause the planet to get hotter. Consequently, it will lead to extreme weather such as hotter heat waves, more frequent droughts, heavier rainfall, and more powerful hurricanes. Additionally, as the Earth's surface temperature becomes hotter, the sea level becomes higher due to the melted glaciers and ice caps. In an alarming sign of events to come, Antarctica has been losing about 134 billion metric tons of ice per year since 2002. This rate could speed up if we keep burning fossil fuels at our current pace, some experts say, causing sea levels to rise several meters over the next 50 to 150 years. If we still not to pay a high attention and take necessary action toward global warming, the whole humanity will eventually suffer the consequence.[5]

II. motivation and objectives

The mainstream research believes that 2°C is the "red line" for temperature increase. Raise beyond that number will be extremely dangerous and leads to disastrous result. [4]
In order to warn people and alert the world to face crisis that global warming have caused, we intend to
employ simple visualization and describe the historical data of carbon emission, temperature anomaly, glacier volume, and sea level change in the past years.
forecast the future global average CO2 emissions and temperature anomaly and to see how they will change in the next 10 years by using various prediction models.
evaluate and compare the results generated from different models and identify which of them are most accurate with least deviation.

On the other hand, 189 countries have signed the Paris Agreement, a long-term temperature goal to keep the increase in global average temperature to limit the increase to 1.5°C. As a result, all member state must lower their annual carbon dioxide. However, a lot of developing countries think that emissions right is equal to development right, and it's not fair for them to pay the price for the developed countries and set emission limit simply based on their annual carbon emission in recent years. To test this view, we decide to
draw a map to show not only the total CO2 emission, but also the emission per capita and GDP for each country.
apply clustering analysis to classify countries into different groups based on these categories and to check the correlation between carbon emission, population and GDP.

III. Critics of Existing Work

In the notebook "Exploring the CO2 and GHG data", the author firstly reviews the history data of the global CO2 emission from 1751 to 2017 and visualizes it with a line chart by applying the time series analysis revealed how the global CO2 emission changed during different time period [2]. Then, a world map was created to indicate the CO2 emission in each country. And audiences can compare the values between different country or continent by easily observing their color. Even though the visualization has successfully showed a plenty of

important information that author trying to tell, there are still some drawbacks. For example, both graphs have no unit of measurement, which makes operators difficult to know the accurate implication of axis labels. And the lack of title and annotation might make people hard to catch the most important aspects of the chart and identify what is relevant. Besides, the world map is not completely interactive to user. Audiences are not free to view the regional CO2 emission in any single year they want but in year 2017 only.

IV. Dataset and Data Preparation
A. Dataset
There are several datasets of global warming in DataHub and The World Bank, but they are all separated, which requires us to download the required data and then organize it in a unified way.

Through the data downloaded in datahub, we can draw an analysis of the general trend of global warming, and give a visual chart of the current temperature, sea level, and melting glaciers.

Through the data in the world bank, we can analyze the growth value of these factors that affect the temperature change in the future, and the future temperature change trend.

Finally, we collected annual carbon dioxide, population, and GDP for each country from the year 1970 to 2015 from the world bank as well. And they are combined as a single document with CSV format.

B. Data Preparation
1. Combining variables glacier mass, sea-level rise changes, temperature anomalies of the northern and southern hemispheres, land and ocean from 1880 to 2014 into one table
2. Combining variables Carbon dioxide, Gas fuel, Solid fuel, Liquid fuel, Methane, Nitric oxide and temperature anomalies into one table named from 1970 to 2014.
3. Using the 'trend' function in Excel to fill the missing values of Methane and Nitric oxide from 1960 to 1970, making our data table completely.
4. Creating a pivot table to show the sum of carbon dioxide emission, population, and GDP for each country from the year 1970 to 2015 as a single table.

V. EDA
A. Temperature Anomalies Trend
For exploratory data analysis, we have created line charts to show the variation trend of the overall global temperature anomalies, or in the northern or southern hemisphere, or for land or ocean respectively. As the figure shows below, we can notice that even though there is some seasonal fluctuation, the overall temperature anomaly index has experienced a continuous rise from 1880 to 2014 by 0.86°C. Besides, it indicates that the glacier volume has decreased 28.65 cubic meters and the sea level has increased by 8.98 millimeters.

### B. Correlation between Temperature and Sea Level/Glacier Mass
On the other hand, we also generated a graph to reveal the relationship between temperature anomaly and sea level or glacier mass. Histograms outside the line chart show the distribution of two variables. For example, the diagram below indicates that there is a strong positive correlation between temperature anomaly and sea level from the year 1950 to 1990 with a Pearson index 0.77 and 0.95 confidence interval 0.61 and 0.87. Because the p-value is less than 0.001, we can say that there is no enough evidence to reject this conclusion.
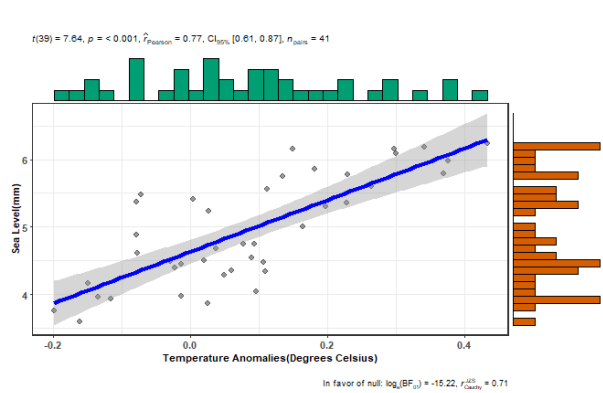

VI. Mapping Method
In the first section, we created a world map to show the total carbon dioxide emission, emission per capita, and

emission per GDP for each country every year from 1975 to 2015 by selecting different categories or years as the figure indicates below.
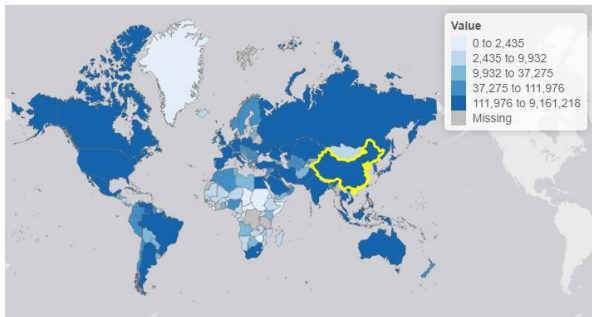
It reveals that for example in 2015, the total consumption emission of carbon dioxide in countries including China, the USA, Canada, and Russia is much higher than others, due to the dark blue on the map representing an emission value between 111,976 to 9,161,218 thousand ton. Besides, a line chart displays and compares the trend of the carbon dioxide for consumption and production from 1970 to 2015 respectively in a specified country. Based on the graph, we can notice that they kept a relatively steady trend and were equal before the year 1990 in Afghanistan. After that, both have experienced an extremely sharp increase, and the value of emission for consumption is way higher than that for production.
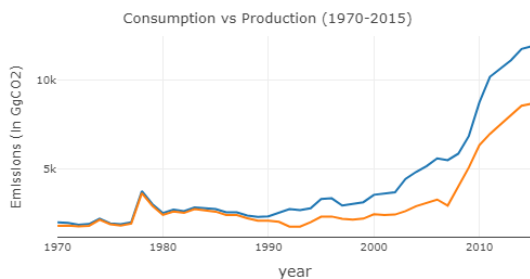

Last but not least, we also created a bar chart to reveal the top 20 countries for selected CO2 emission index in a specific year. For example, the graph below indicates that China, America, and India ranked top 3 in 2015.
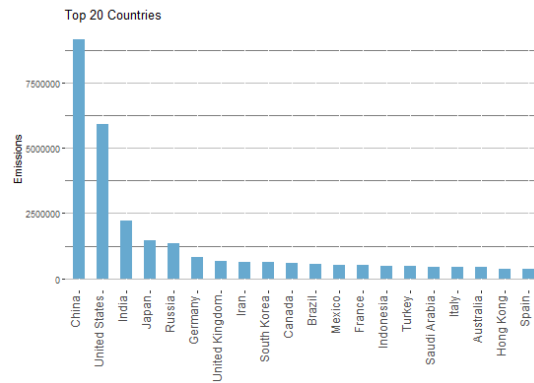

Cluster analysis refers to a multivariate method that aims to classify a sample of subjects based on a set of measured variables into several different groups such that similar subjects are placed in the same group. In this part, we tried to conduct cluster analysis and group all countries based on their carbon dioxide emission, emission per capita, emission per GDP for consumption and production.

Firstly, we can obtain a correlation graph based on the variables we have selected. If two variables have a strong positive or negative (near 1 or -1) correlation, then the clustering will be meaningless.
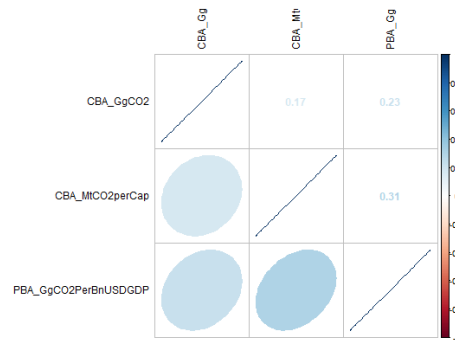

By applying hierarchical clustering with a 'Optimal Leaf Ordering' (OLO) method and pick '4' as the desired number of final groups as an instance, the result is shown below. The dendrogram along the side reveals how the variables and countries are independently clustered. The heat map displays the data value for both and possibly standardized so they all fit in the same range. The darker color, the stronger correlation between the specific country and variable, and vice versa. The rectangular area having the same color suggests a group of countries is highly correlated with the corresponding group of

variables. As a result, we can easily see which of them share common features.

World CO2 Emissions by Country
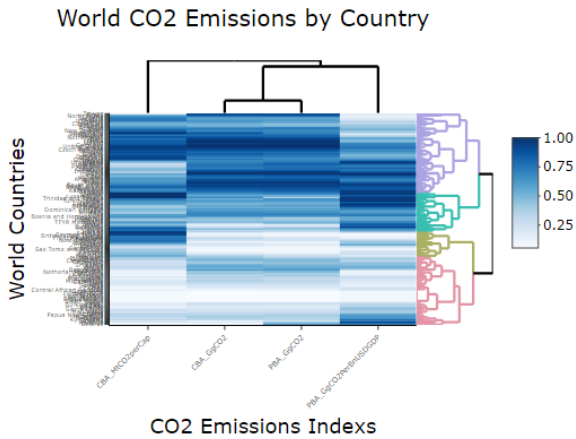
And a world map beside shows countries with different colors based on their corresponding clustering group.
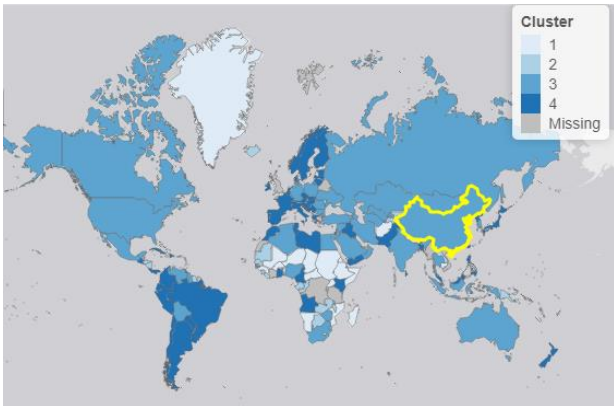
By selecting different method including Euclidean, maximum, Ward.D, Single and so on, we can gain a summary report for each cluster group showing their quantile and range.

```
[1] "Cluster 1 summary"
   CBA_GgCO2      CBA_MtCO2perCap    PBA_GgCO2PerBnUSDGDP    Cluster
Min.   :     0   Min.   :  0.000   Min.   :    0    Min.   :1
1st Qu.: 13797   1st Qu.:  5.584   1st Qu.: 1039    1st Qu.:1
Median : 48304   Median :  8.948   Median :13185    Median :1
Mean   : 66429   Mean   :101.340   Mean   :12363    Mean   :1
3rd Qu.: 85861   3rd Qu.: 49.474   3rd Qu.:18501    3rd Qu.:1
Max.   :291892   Max.   :932.193   Max.   :32604    Max.   :1
[1] "Cluster 2 summary"
   CBA_GgCO2      CBA_MtCO2perCap    PBA_GgCO2PerBnUSDGDP    Cluster
Min.   :  3116   Min.   :  0.3213   Min.   : 33112   Min.   :2
1st Qu.: 90342   1st Qu.: 36.2542   1st Qu.: 39281   1st Qu.:2
Median :161996   Median : 67.0870   Median : 44455   Median :2
Mean   :312029   Mean   : 91.5797   Mean   : 51673   Mean   :2
3rd Qu.:351385   3rd Qu.: 98.7180   3rd Qu.: 58673   3rd Qu.:2
Max.   :1540139  Max.   :387.7271   Max.   :105988   Max.   :2
[1] "Cluster 3 summary"
   CBA_GgCO2        CBA_MtCO2perCap   PBA_GgCO2PerBnUSDGDP   Cluster
Min.   :   137418  Min.   :  19.92   Min.   : 38491   Min.   :3
1st Qu.:  1409128  1st Qu.: 123.41   1st Qu.: 53675   1st Qu.:3
Median :  3018887  Median : 238.76   Median : 72877   Median :3
Mean   : 13619826  Mean   : 424.66   Mean   : 85847   Mean   :3
3rd Qu.:  9360370  3rd Qu.: 561.79   3rd Qu.: 94152   3rd Qu.:3
Max.   :250937397  Max.   :2914.50   Max.   :246672   Max.   :3
[1] "Cluster 4 summary"
   CBA_GgCO2       CBA_MtCO2perCap    PBA_GgCO2PerBnUSDGDP   Cluster
Min.   :  135430   Min.   :   0.00   Min.   :    0    Min.   :4
1st Qu.:  379915   1st Qu.:  48.12   1st Qu.:21142    1st Qu.:4
Median : 1080964   Median : 214.30   Median :28052    Median :4
Mean   : 4669665   Mean   : 254.38   Mean   :25582    Mean   :4
3rd Qu.: 3765476   3rd Qu.: 398.09   3rd Qu.:32825    3rd Qu.:4
Max.   :59678844   Max.   :1240.68   Max.   :40151    Max.   :4
```

VII. Forecasting Method

Since the data we obtained has no seasonal changes, we use an automatic time series prediction algorithm to do future forecast.

A. ETS Model

introduction of automatic exponential smoothing algorithm:

Exponential smoothing is a rule of thumb for smoothing time series data using an exponential window function. In a simple moving average, past observations are weighted equally, and the exponential function is used to assign weights that decrease exponentially with time. The raw data sequence is often represented by **{Xt}** beginning at time **t=0**, and the output of the exponential smoothing algorithm is commonly written as **{St}**, which may be regarded as a best estimate of what the next value of **X** will be. The simplest form of exponential smoothing is given by the formulas:

$$S_0 = X_0$$
$$S_t = \alpha X_t + (1-\alpha) S_{t-1}, \ t > 0$$

where **α** is the smoothing factor, and **0<α<1**. [3]
As our data all are annual data, our exponential smoothing method is **ETS(M,A,N)**.

outcome of ETS model:

- ETS(M,A,N) for annual Total Co2 emission (α = 0.9999, β = 0.0975, l= 8566785.6994, b =509076.9212);
- ETS(M,A,N) for annual Liquid Fuel (α = 0.9999, β = 0.149, l= 2683588.1766, b =334447.3337);
- ETS(M,A,N) for annual Solid Fuel (α = 0.9999, β = 0.3162, l= 4953836.8902, b =66674.698);
- ETS(M,A,N) for annual Gas Fuel (α = 0.9999, β = 0.0001, l= 699139.2053, b =96630.89);
- ETS(M,A,N) for annual Methane (α = 0.6222, l= 4557825.5839);
- ETS(M,A,N) for annual Nitric oxide (α = 0.864, l=2303348.7964);
- ETS(M,A,N) for annual Temperature anomaly (α = 0.2472, β = 0.0001, l= -0.0197, b =0.013);

B. ARIMA Model

introduction of Auto Arima method:

ARIMA is a class of models that 'explains' a given time series based on its own past values, that is, its own lags and the lagged forecast errors, so that equation can be used to forecast future values and it is very suitable for non-seasonal time series.

A non-seasonal **ARIMA (p, d, q)** process is given by

$$\varphi(B)(1 - B\ d\ )\ yt = c + \theta(B)\varepsilon t$$

The main task in automatic **ARIMA** forecasting is selecting an appropriate model order, that is the values **p, q, P, Q, D, d**. If **d** and **D** are known, we can select the orders **p**, **q**, **P** and **Q** via an information criterion such as the **AIC**:

$$AIC = -2\log(L) + 2(p + q + P + Q + k)$$

where **k = 1** if **c ≠ 0** and **0** otherwise, and **L** is the maximized likelihood of the model fitted to the differenced data **(1−Bm) d (1−B) dyt**. [1]

outcome of Auto Arima Model:

- ARIMA (1,1,0) for Total CO2 emission ($\varphi 1 = 0.3331$, c=489292.49);
- ARIMA (1,2,2) for Liquid Fuel ($\varphi 1 = -0.6586$; $\theta 1 = 0.3325$; $\theta 2 = 0.5997$);
- ARIMA (0,1,0) for Solid Fuel (c=165093.98);
- ARIMA (0,1,0) for Gas Fuel (c=108380.22);
- ARIMA (0,1,1) for Methane ($\varphi 1 =$
- -0.3378, c=64533.38);
- ARIMA (0,1,0) for Nitric oxide;
- ARIMA (0,1,1) for Temperature Difference ($\varphi 1 = 0.7817$, c=0.0132);

C. Model Comparison

1. Residual analysis suggests estimation model quality.
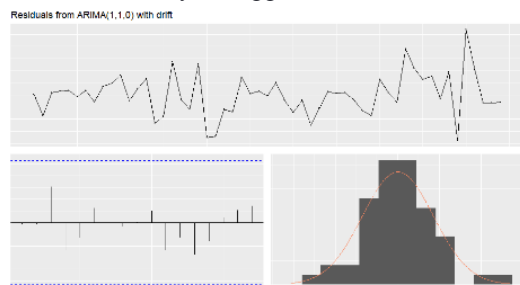
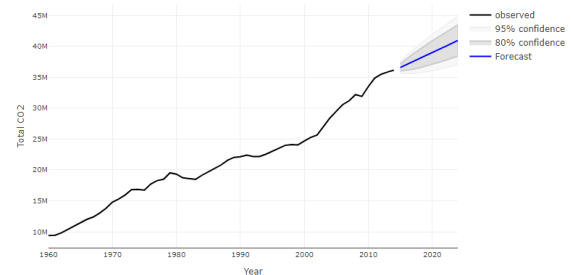2.Calculating the difference between consecutive observations shows a way to stabilize the non-stationary time series.
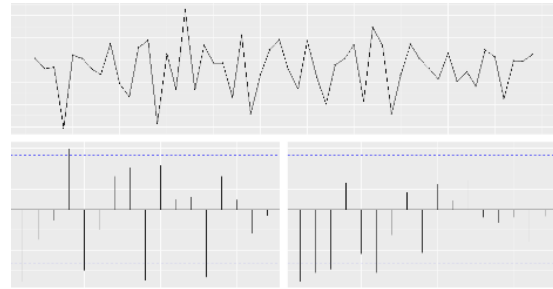
3. RMSE and MSE to check the accuracy of two model

① Total CO2 emission (Choose ARIMA)

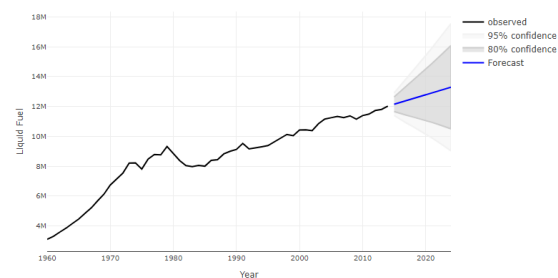ETS ---[RMSE]:450988.2 [MAPE]:1.708637 [MASE]:0.6326368

ARIMA---[RMSE]:42.3747.5 [MAPE]:1.495187[MASE]:0.5602159

② Liquid Fuel (Choose ETS)

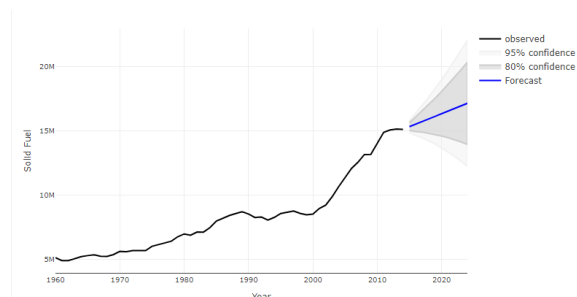**ETS ---[RMSE]:266517.9 [MAPE]:2.231191 [MASE]:0.7098989**

ARIMA---[RMSE]:267265.6 [MAPE]:2.643741 [MASE]:0.802328

③ Solid Fuel (Choose ARIMA)
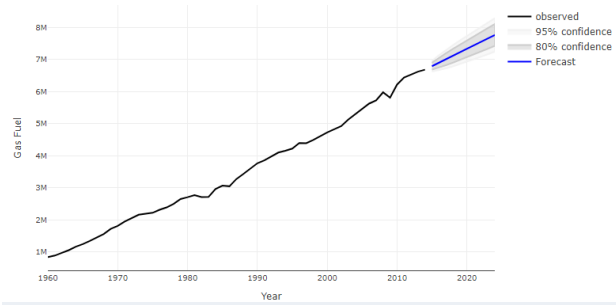
ETS ---[RMSE]:242494.4 [MAPE]:2.2104 [MASE]:0.789762

ARIMA---[RMSE]:222180.6 [MAPE]:2.065883 [MASE]:0.7177826

④Gas Fuel (Choose ARIMA)
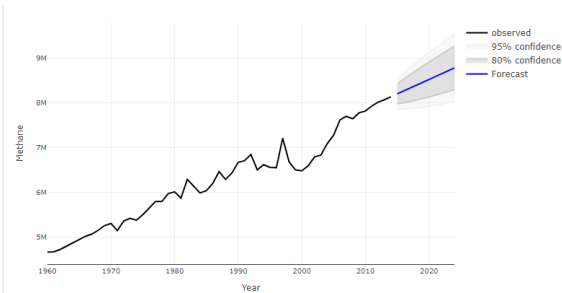
ETS ---[RMSE]: 82506.96 [MAPE]:1.734919 [MASE]:0.4802816

ARIMA---[RMSE]:82453.58 [MAPE]:1.668037 [MASE]:0.4617664

⑤Methane (Choose ARIMA)

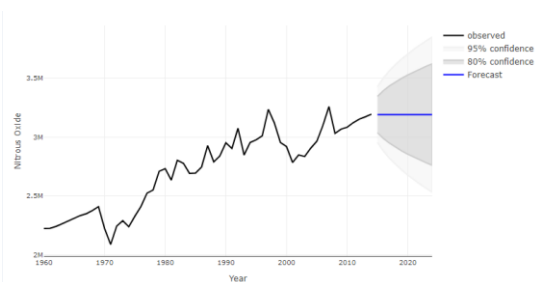ETS ---[RMSE]: 166821.5 [MAPE]:1.77536 [MASE]:0.8234555

ARIMA---[RMSE]:166652.3 [MAPE]:1.7651 [MASE]:0.8193982

⑥Nitric oxide (Choose ETS)
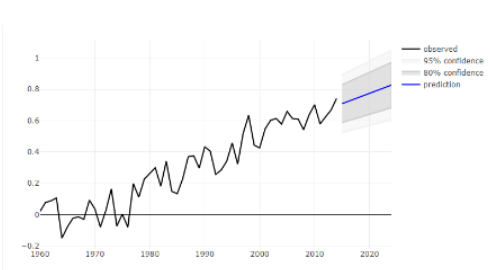
ETS ---[RMSE]: 101545.8 [MAPE]:2.932402 [MASE]:0.9943876

ARIMA---[RMSE]:102298.7 [MAPE]:2.875211 [MASE]:0.9823211

⑦Temperature Difference (Choose ARIMA)

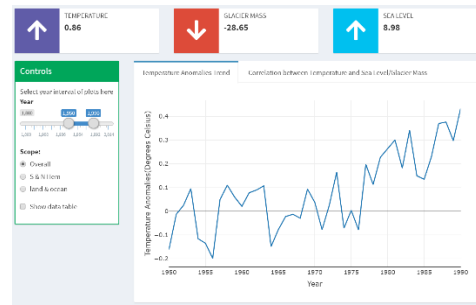ETS ---[RMSE]: 0.09186258 [MAPE]:88.66174 [MASE]: 0.8611064

ARIMA---[RMSE]:0.0917142 [MAPE]: 90.78408 [MASE]: 0.841901
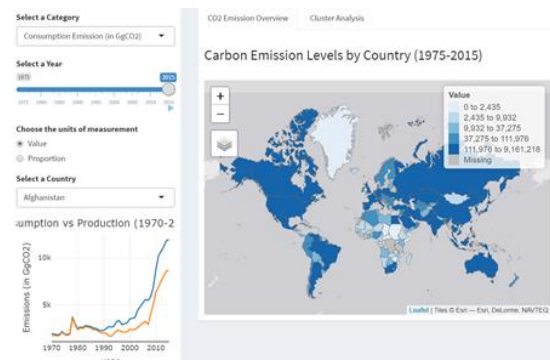

VIII. The Application
A. EDA

In the first page, users are able to gain a general look of temperature trend in the past decades.


B. World map

With our shiny application, users are free to select the specific category, year, and country in the mapping part to observe the carbon dioxide emission and its trend during a period.

C. Clustering analysis

In terms of the cluster analysis, users can select different clustering methods, the number of classified groups and measured variables to see how countries will be classified into different groups in such a way.
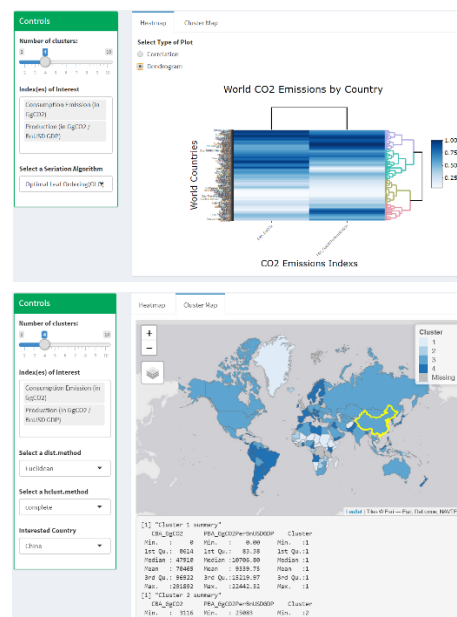
D. Predictive analysis

With our shiny application, users can choose the model to see the outcome of a different forecasting method.

1. Prediction Model

In the ETS Method interface, users can choose the year they want to predict and observe the results. Also, they can see the residual of variables and get the whole data outcome from the data panel. In the ARIMA Method interface, users can choose the variables and the diff they want to see. This plot is interactive, so the user can get the specific value just by placing the mouse on the plot.

2. Model comparison

User can choose the variable of based on their personal interest, the application will give the best model plot and the RMSE result. With this application, user can get the best fit model.
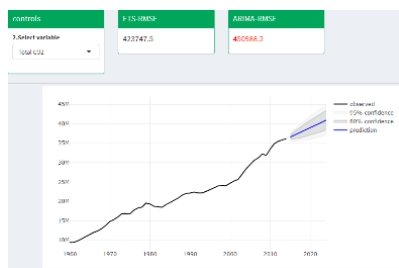
IX. Future Works and Conclusion

From the clustering analysis, we can find that if we select total CO2 emissions and CO2 emissions per capita as classification features, countries such as China are not in the group having a high overall ranking for emission. Hence, we can conclude that it is not proper to judge a country's CO2 emissions simply based on their total emissions but ignore their population and GDP. Only those with a high number in multiple categories should be strictly limited.

In terms of the predictive analysis, the best fit model ARIMA indicates that the future temperature anomaly and carbon dioxide emission in 2024 is going to be 0.8271°C and 40.9370 million ton respectively.

Since our data are recorded annually, we cannot add seasonal factors to make predictions. As a result, the predicted data will not be very accurate. In the future, we need to find more detailed data such as recorded monthly. It will be more accurate to predict based on seasonal changes and apply extra time series forecasting methods combined with seasonal data for forecasting and model comparison.

Reference

[1] Hyndman, R., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. Journal of Statistical Software, 27(3), 1 - 22. doi: http://dx.doi.org/10.18637/jss.v027.i03

[2] Kaggle.com. 2020. Exploring The CO2 And GHG Data. [online] Available at: <https://www.kaggle.com/martijnjschut/exploring-the-co2-and-ghg-data> [Accessed 26 April 2020].

[3] En.wikipedia.org. 2020. Exponential Smoothing. [online] Available at: <https://en.wikipedia.org/wiki/Exponential_smoothing> [Accessed 26 April 2020].

[4] En.wikipedia.org. 2020. Global Warming. [online] Available at: <https://en.wikipedia.org/wiki/Global_warming> [Accessed 26 April 2020].

[5] NRDC. 2020. Global Warming 101. [online] Available at: <https://www.nrdc.org/stories/global-warming-101> [Accessed 26 April 2020].