****

**Visual Analysis of the Potential Impacts of Climate Change on World Food Supply**

**An examination of earth’s changing climate and its effect on agricultural factors**

**Final Project Report**

**Bryan Fendley**

**Daniel Noble**

**Fantahun Zikie**

**12/1/2020**

**Introduction**

The main purpose of this project is to create interactive visualizations on the potential impact of climate change on world food supply. Global data on crop production and yield as well as carbon dioxide (CO2) emissions and surface temperature over a period of time was collected, cleaned and analyzed. In coherence with several scientific reports, the result of our visual analysis showed an upward trend for global CO2 production and increase in surface temperature. As well, even though global crop production and yield didn’t show any sign of decrease, as it also depends on other factors, the increase in surface temperature can affect soil moisture, which can negatively impact crop production.

The source of the data used in this analysis is obtained from multiple sources including the Our world in data ([www.ourworldindata.org](http://www.ourworldindata.org)), a repository website for global data that goes a few hundred years in some cases and sponsored by Oxford University, Oxford Martin School and GCDL. The data set is developed by Cynthia Rosenzweig of the NASA Goddard Institute for Space Studies and Ana Iglesias of Universidad Politécnica de Madrid and disseminated by the NASA Socioeconomic Data and Applications Center (SEDAC), managed by CIESIN at Columbia University. This is obtained from <https://sedac.ciesin.columbia.edu/data/set/crop-climate-potential-impacts-world-food-supply>

**Data Cleaning and Processing**

After deciding on the data for the project, we downloaded the files in csv format to the local machine for further processing. The cereals Maize (Corn), Rice and Wheat are obtained for the project as they form the global essential crops for food globally. The production data is measured in tones, while the yield data is measured in hectogram per hectare. The CO2 production data is given in million tones, while the surface temperature is a difference from the global average from 1951 to 1980, and given as surface temperature anomaly. The data has three components: Meteorological variables, environmental modification variables, crop management strategies and the dependent variable crop yield.

The data are imported into SQL Server for further cleaning and processing. Each csv file is imported to a separate table. Since there are discrepancies in the range of data for each country, in a single data set and across the data sets, the range of years is identified by using SQL queries, attached here with the project as a separate file. Six countries only have very recent data for all datasets and opted out from the analysis. The countries are Montenegro has data only in the year range of 2006-2018, Serbia has data for 2006-2018, Sudan 2012-2018, Luxembourg 2000-2018, Belgium 2000-2018 and Denmark 2010-2018.

Further processing of the data is performed with python pandas by creating dataframes for all the datasets.

**Data Visualizations and Visual Analytics**

The visualization creation, and for analysis our project was created using Python and Jupyter Notebook and google collaborator. The visualizations are built using plotly and ipywidgets for interactivity.

The visualizations depict changes over time for several metrics, including surface temperature, carbon dioxide emissions, and crop production. Various metrics can be filtered by crop and country. The scatterplots, line graphs, and bar graphs serve to show changes in crop production, population, carbon dioxide levels, and surface temperature over time. By examining how indicators of global climate change affects a country's food supply and production, one can draw conclusions regarding future global agricultural health. In addition to the main goal of depicting how climate change affects the global food supply, other data were examined to provide context for the crops that were examined in our study.

After cleaning the data, importing into pandas data frames, the first information visualized is what the global CO2 production looks like for all the countries in the dataset. Figure 1 shows the CO2 production by each country from 1993 and 2018. It is evident from Fig 1 that CO2 production for most developing countries is increasing continuously, with China producing the highest overtaking US from 2006.

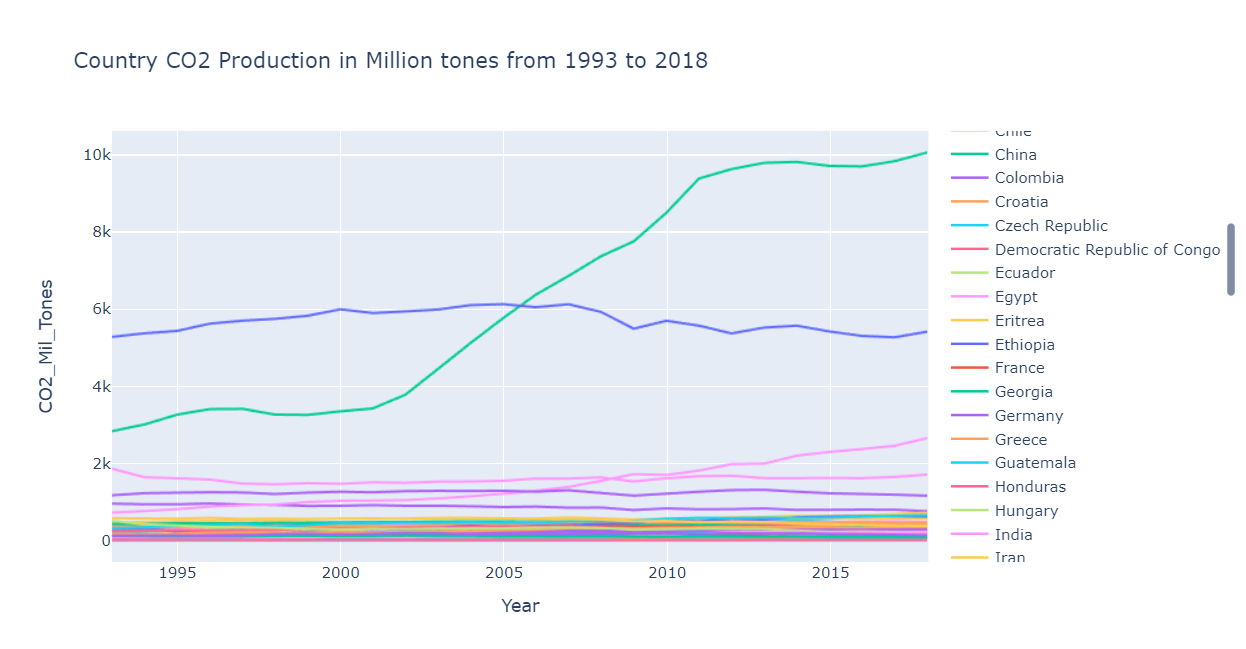
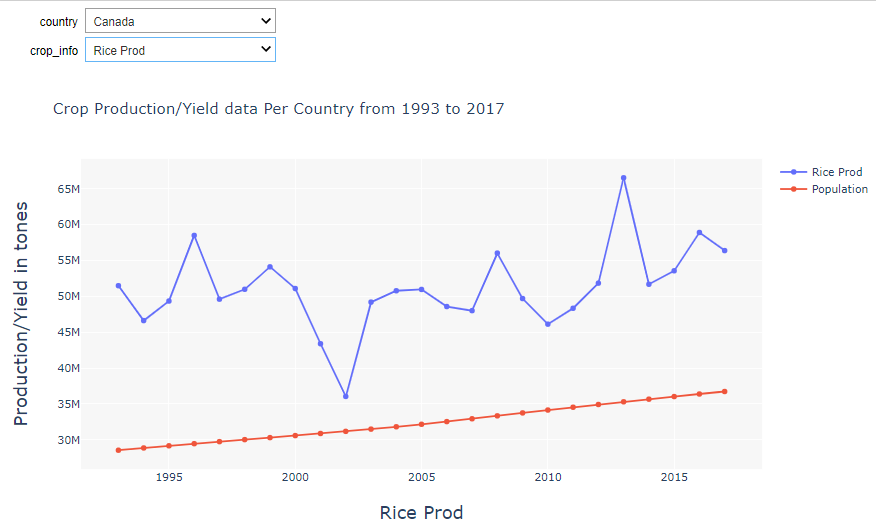


Figure 1: CO2 production from 1993 to 2018 by Countries

The next set of visualizations are for showing crop production and yield with respect to the population. Such charts help to interactively view the relationship between each country’s productions and yields with respect to population in one place. The python code for this is provided in the appendix and also attached herewith this as a separate python code document. Figure 2 to 5 show the Rice and Maize (Corn) production comparison between Canada, China, USA and Ukraine

Figure 2: Canada Rice Production Vs Population from 1993 to 2017

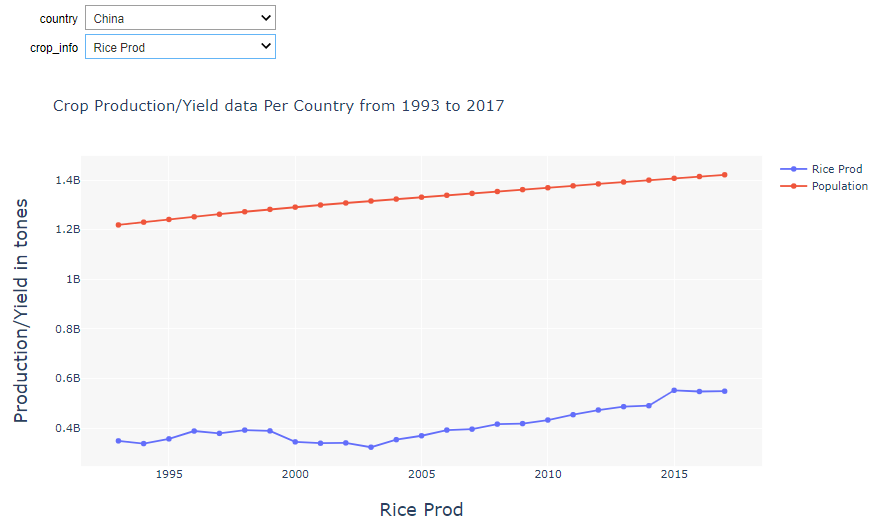


Figure 3: China Rice Production Vs Population Growth from 1993 to 2017

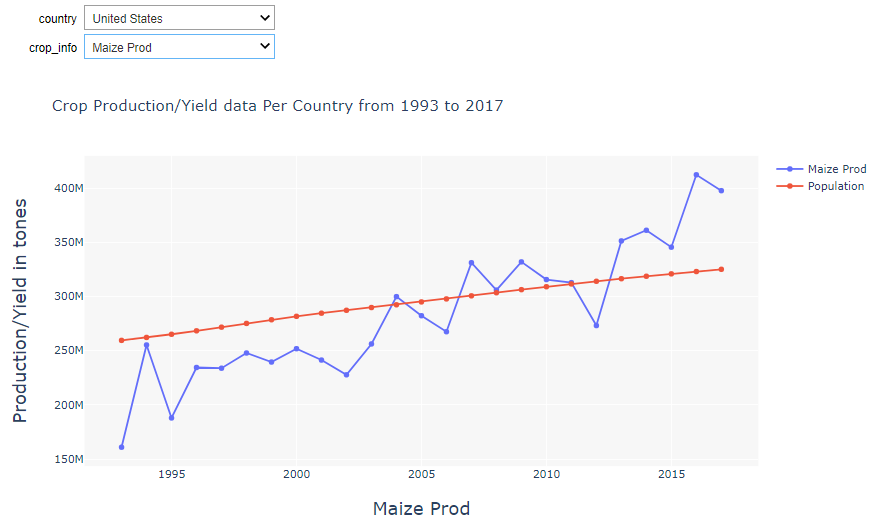


Figure 4: US Maize Production Vs Population Growth from 1993 to 2017

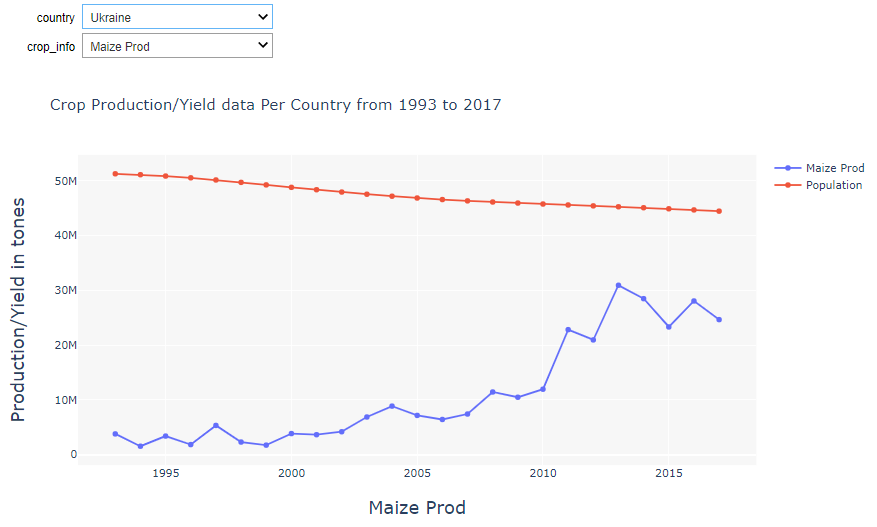


Figure 5: Ukraine Maize Production Vs Population Growth from 1993 to 2017

Barcharts are used when we want to depict distribution of data points or perform a comparison of metric values across different subgroups of datasets. It’s possible to see from a barchart which groups have the highest or most common, and how other groups compare against the others. We used barcharts to show the production of crops by each country (which can be interactively accessed). Figure 6 shows the crop production in Afghanistan in the years 1995, 2000, 2005, 2010 and 2015 for all the three crops. Figure 7 shows Ethiopian crop production while figure 8 is for Israel.

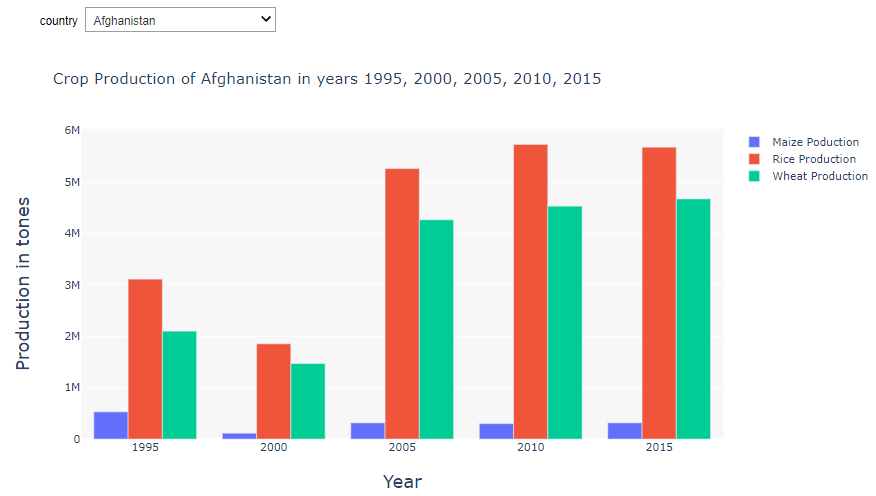


Figure 6: Crop production in Afghanistan in years 1995, 2000, 2005, 2010, 2015

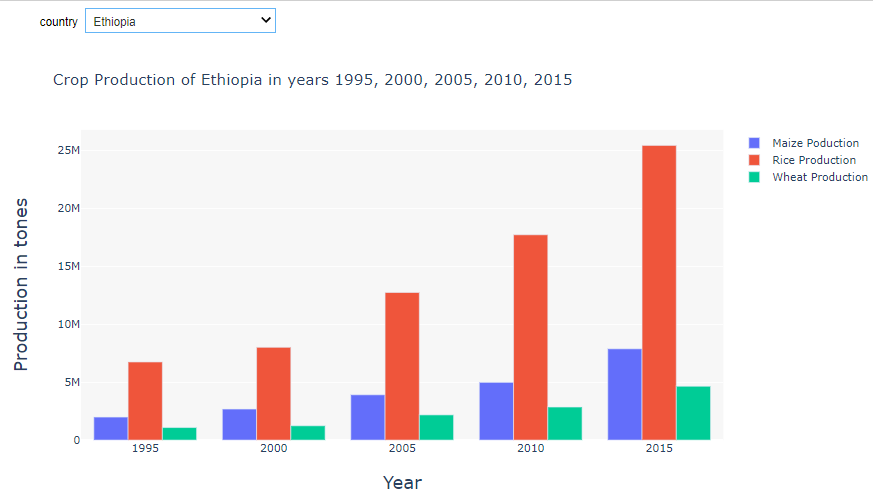


Figure 7: Crop production in Ethiopia in years 1995, 2000, 2005, 2010, 2015

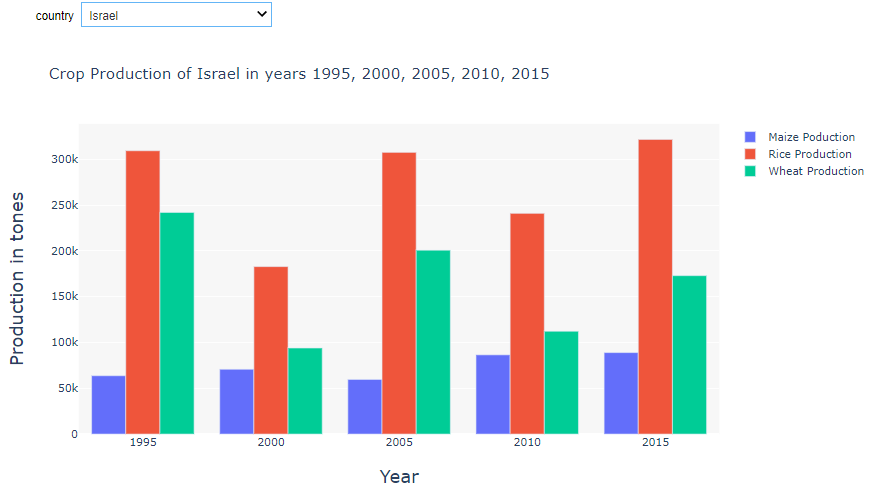


Figure 8: Crop production in Israel in the years 1995, 2000, 2005, 2010, 2015

The next set of interactive visualizations is to depict the correlation between each country’s CO2 production and surface temperature anomaly. As discussed above, most developing countries have CO2 production increase in the period under study, and few developed countries like the US and Netherlands decreased CO2 production. The general surface temperature anomaly data shows an increase in all countries, which shows global warming. For Example, even though the US CO2 production is decreasing, the surface temperature anomaly is increasing as it is affected mostly by the global trend. This shows that independent actions by different countries cannot produce intended results and thus global cooperation is required. Figure 9 shows the visualization of CO2 production Vs surface temperature anomaly of … country including the pearson correlation coefficient.

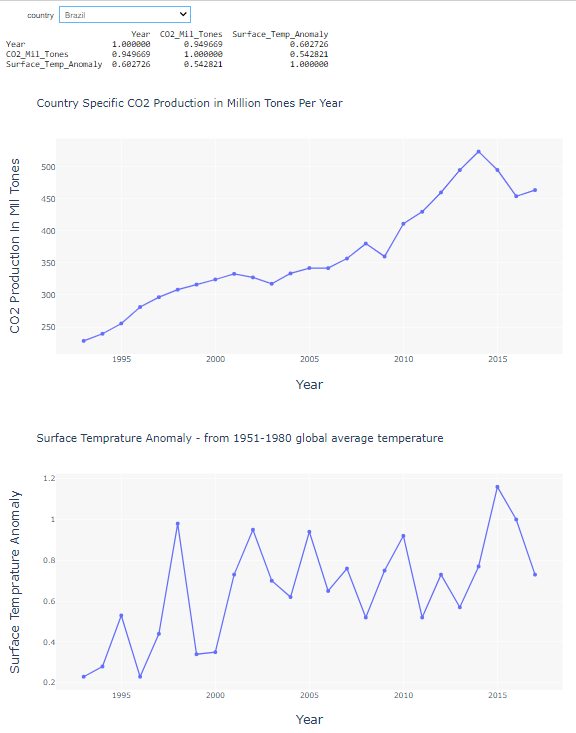


Figure 9: The correlation between CO2 production and Surface temperature anomaly for Brazil

Last but not least, is a visualization of the surface temperature anomaly per year for each country, shown over a global map. This interactive visualization of heat maps shows how each country’s and global surface temperature is changing over time. Figure 10 shows the 1993 global surface temperature anomaly information while Figure 11 shows the 2017 global surface temperature anomaly information. How the surface temperature is changing is obvious from the visualizations.



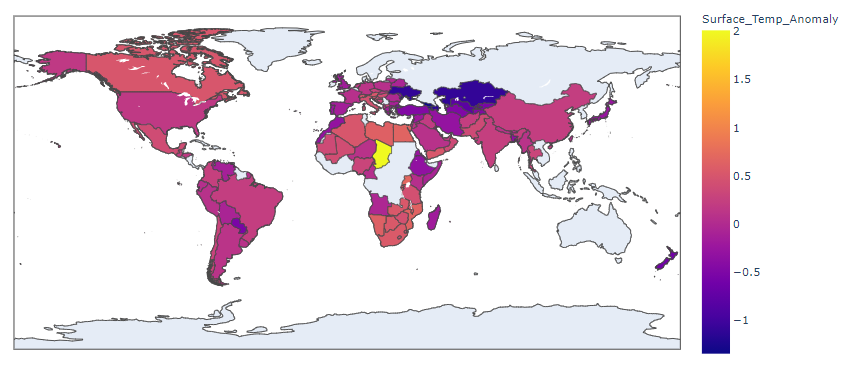


Figure 10: Global Surface temperature anomaly in 1993



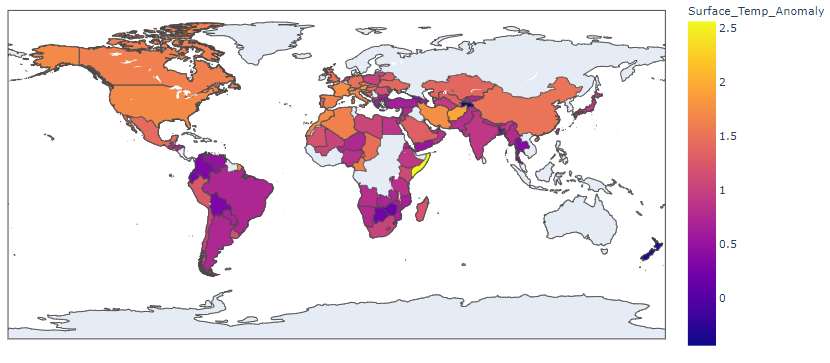


Figure 11: Global Surface temperature anomaly in 2017

**Alternative Visualizations**

Alternative visualizations were created using data from the United States Department of Agriculture Research Service. This data can be found here: [FoodData Central](https://ndb.nal.usda.gov/download-datasets.html). The caloric and macronutrient content for our chosen crops was extracted and cleaned using Microsoft Excel. The visualizations were created using Tableau and can be found here: [Crop Nutrient Data](https://public.tableau.com/views/CropNutrientData/Sheet1?:language=en&:display_count=y&:origin=viz_share_link). By comparing the nutritional benefits of the crops considered in the study one can compare the costs and benefits of shifting production from one crop to another. For example, even though a crop might be less susceptible to loss in production due to a rise in temperature, the nutritional value might be lesser than a crop that is highly susceptible to a change in climate. Treemaps depicting population and CO2 emissions were also created but it was decided there were more efficient ways of depicting these data. The treemaps were produced in Tableau and can be found here: [Alternative Treemaps](https://public.tableau.com/views/AlternativeTreemaps/CO2Emissons?:language=en&:display_count=y&publish=yes&:origin=viz_share_link). Figure 12 shows the Calories per 100 grams for the three crops under consideration in this report. Such information can later be used to evaluate if each country is producing at least the recommended amount of calories to its population at a given time. Figure 13 shows the amount of macro nutrients per 100 grams of each crop.

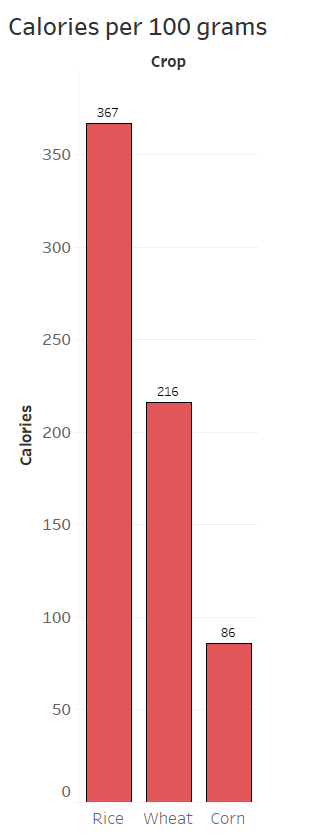


Figure 12: Calories per 100 grams of crops for Rice, Wheat and Maize (Corn)

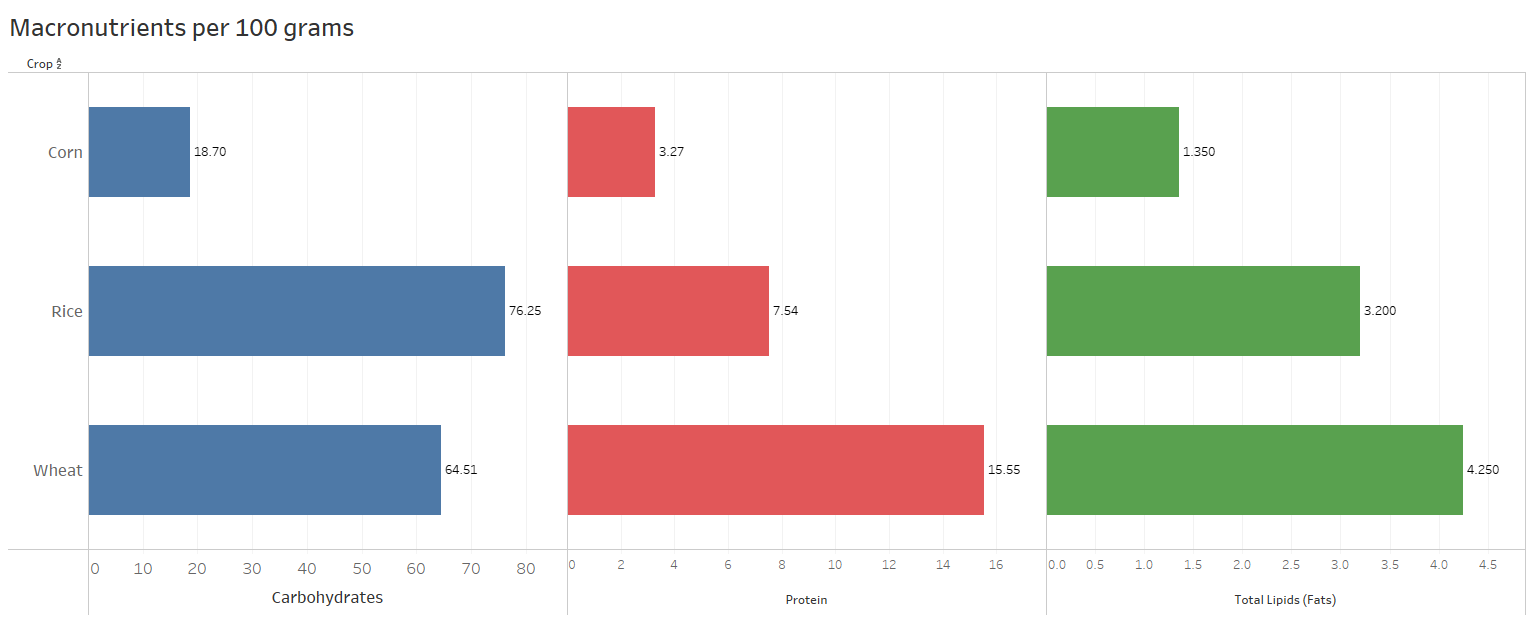


Figure 13: Macro nutrients per 100 grams of each crop.

Both figures are based on data collected by the USDA. The data is based on an average of 100 grams of raw sweet yellow corn, raw brown long-grain rice, and crude wheat bran. The first bar graph depicts the caloric content of our three crops. The second bar graph shows the amount of the three most important macronutrients for each of our respective crops. Rice generally contains more calories and carbohydrates than wheat and corn, but wheat contains more protein and fats.

**References**

Hannah Ritchie (2020) - "Agricultural Production". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/agricultural-production' [Online Resource]

Rosensweig, C. and A. Iglesias. (1999) - Potential Impacts of Climate Change on World Food Supply: Datasets from a Major Crop Modeling Study. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). <https://doi.org/10.7927/H43R0QR1>. Accessed November 2020

**Appendix:**

**Codes for Data Cleaning and Visualization**

1. SQL Code for Data Cleaning

-- Create a database called GlobalFood and Import the CSV files to the database as separate tables

USE GlobalFood;

GO

-- Get number of countries in each data set

SELECT COUNT(DISTINCT Country)

FROM [dbo].[CO2\_Data]

SELECT COUNT(DISTINCT Country)

FROM [dbo].[Maize\_Production]

SELECT COUNT(DISTINCT Country)

FROM [dbo].[Maize\_Yields]

SELECT COUNT(DISTINCT Country)

FROM [dbo].[Rice\_Production]

SELECT COUNT(DISTINCT Country)

FROM [dbo].[Rice\_Yields]

SELECT COUNT(DISTINCT Country)

FROM [dbo].[Surface\_Temperature\_Anomaly]

SELECT COUNT(DISTINCT Country)

FROM [dbo].[Wheat\_Production]

SELECT COUNT(DISTINCT Country)

FROM [dbo].[Wheat\_Yields]

-- Get countries which exist in all data sets - and also their counts

-- (We got 113 countries - common in all datasets)

SELECT COUNT(DISTINCT Country)

FROM

(

SELECT DISTINCT Country

FROM [dbo].[CO2\_Data]

INTERSECT

SELECT DISTINCT Country

FROM [dbo].[Maize\_Production]

INTERSECT

SELECT DISTINCT Country

FROM [dbo].[Maize\_Yields]

INTERSECT

SELECT DISTINCT Country

FROM [dbo].[Rice\_Production]

INTERSECT

SELECT DISTINCT Country

FROM [dbo].[Rice\_Yields]

INTERSECT

SELECT DISTINCT Country

FROM [dbo].[Surface\_Temperature\_Anomaly]

INTERSECT

SELECT DISTINCT Country

FROM [dbo].[Wheat\_Production]

INTERSECT

SELECT DISTINCT Country

FROM [dbo].[Wheat\_Yields]

)X

GO

-- To make our subsequest queries simple to read, we created a view from the above INTERSECTs

CREATE VIEW vw\_Countries

AS

SELECT DISTINCT Country

FROM [dbo].[CO2\_Data]

INTERSECT

SELECT DISTINCT Country

FROM [dbo].[Maize\_Production]

INTERSECT

SELECT DISTINCT Country

FROM [dbo].[Maize\_Yields]

INTERSECT

SELECT DISTINCT Country

FROM [dbo].[Rice\_Production]

INTERSECT

SELECT DISTINCT Country

FROM [dbo].[Rice\_Yields]

INTERSECT

SELECT DISTINCT Country

FROM [dbo].[Surface\_Temperature\_Anomaly]

INTERSECT

SELECT DISTINCT Country

FROM [dbo].[Wheat\_Production]

INTERSECT

SELECT DISTINCT Country

FROM [dbo].[Wheat\_Yields]

GO

-- To make appropriate comparisons, we opted to take countries that existed in all data sets

-- Thus remove countries not in all data sets as follows - Use the vw\_Countries created above

DELETE FROM [dbo].[CO2\_Data]

WHERE Country NOT IN (SELECT Country FROM vw\_Countries) -- Deleted 11759 rows

DELETE FROM [dbo].[Maize\_Production]

WHERE Country NOT IN (SELECT Country FROM vw\_Countries) -- Deleted 5292 rows

DELETE FROM [dbo].[Maize\_Yields]

WHERE Country NOT IN (SELECT Country FROM vw\_Countries) -- Deleted 5289 rows

DELETE FROM [dbo].[Rice\_Production]

WHERE Country NOT IN (SELECT Country FROM vw\_Countries) -- Deleted 6176 rows

DELETE FROM [dbo].[Rice\_Yields]

WHERE Country NOT IN (SELECT Country FROM vw\_Countries) -- Deleted 6142 rows

DELETE FROM [dbo].[Surface\_Temperature\_Anomaly]

WHERE Country NOT IN (SELECT Country FROM vw\_Countries) -- Deleted 13134 rows

DELETE FROM [dbo].[Wheat\_Production]

WHERE Country NOT IN (SELECT Country FROM vw\_Countries) -- Deleted 2598 rows

DELETE FROM [dbo].[Wheat\_Yields]

WHERE Country NOT IN (SELECT Country FROM vw\_Countries) -- Deleted 2586 rows

-- Check the data again

SELECT \*

FROM [dbo].[CO2\_Data]

SELECT \*

FROM [dbo].[Maize\_Production]

SELECT \*

FROM [dbo].[Maize\_Yields]

SELECT \*

FROM [dbo].[Rice\_Production]

SELECT \*

FROM [dbo].[Rice\_Yields]

SELECT \*

FROM [dbo].[Surface\_Temperature\_Anomaly]

SELECT \*

FROM [dbo].[Wheat\_Production]

SELECT \*

FROM [dbo].[Wheat\_Yields]

-- The data should be cleaned further - so that the Year range will be consistent for all data sets and countries

-- For this we will start with the range of years for which data is available for each country in all the data sets

SELECT MAX([Year])

FROM

(

SELECT Country, MIN(Year) [Year]

FROM [dbo].[CO2\_Data]

GROUP BY Country

--ORDER BY Country

)X

SELECT MAX([Year])

FROM

(

SELECT Country, MIN(Year) [Year]

FROM [dbo].[Maize\_Production]

GROUP BY Country

--ORDER BY Country

)X

SELECT MAX([Year])

FROM

(

SELECT Country, MIN(Year) [Year]

FROM [dbo].[Maize\_Yields]

GROUP BY Country

--ORDER BY Country

)X

SELECT MAX([Year])

FROM

(

SELECT Country, MIN(Year) [Year]

FROM [dbo].[Rice\_Production]

GROUP BY Country

--ORDER BY Country

)X

SELECT MAX([Year])

FROM

(

SELECT Country, MIN(Year) [Year]

FROM [dbo].[Rice\_Yields]

GROUP BY Country

--ORDER BY Country

)X

SELECT MAX([Year])

FROM

(

SELECT Country, MIN(Year)[Year]

FROM [dbo].[Surface\_Temperature\_Anomaly]

GROUP BY Country

--ORDER BY Country

)X

SELECT MAX([Year])

FROM

(

SELECT Country, MIN(Year)[Year]

FROM [dbo].[Wheat\_Production]

GROUP BY Country

--ORDER BY Country

)X

SELECT MAX([Year])

FROM

(

SELECT Country, MIN(Year) [Year]

FROM [dbo].[Wheat\_Yields]

GROUP BY Country

--ORDER BY Country

)X

---- Five countries - don't have data starting 1993 - all the rest (106 countries do)

--

SELECT \*

FROM

(

SELECT Country, MIN(Year) [Year]

FROM [dbo].[Wheat\_Yields]

GROUP BY Country

--ORDER BY Country

)X

WHERE [Year] > 1993

----------------------------------------------------

SELECT \*

FROM

(

SELECT Country, MIN(Year) [Year]

FROM [dbo].[Maize\_Production]

GROUP BY Country

--ORDER BY Country

)X

WHERE [Year] > 1993

SELECT \*

FROM

(

SELECT Country, MIN(Year) [Year]

FROM [dbo].[Maize\_Yields]

GROUP BY Country

--ORDER BY Country

)X

WHERE [Year] > 1993

SELECT \*

FROM

(

SELECT Country, MIN(Year) [Year]

FROM [dbo].[Rice\_Production]

GROUP BY Country

--ORDER BY Country

)X

WHERE [Year] > 1993

SELECT \*

FROM

(

SELECT Country, MIN(Year) [Year]

FROM [dbo].[Rice\_Yields]

GROUP BY Country

--ORDER BY Country

)X

WHERE [Year] > 1993

SELECT \*

FROM

(

SELECT Country, MIN(Year)[Year]

FROM [dbo].[Surface\_Temperature\_Anomaly]

GROUP BY Country

--ORDER BY Country

)X

WHERE [Year] > 1993

SELECT \*

FROM

(

SELECT Country, MIN(Year)[Year]

FROM [dbo].[Wheat\_Production]

GROUP BY Country

--ORDER BY Country

)X

WHERE [Year] > 1993

SELECT \*

FROM

(

SELECT Country, MIN(Year) [Year]

FROM [dbo].[Wheat\_Yields]

GROUP BY Country

--ORDER BY Country

)X

WHERE [Year] > 1993

/\*

Final: Cleaned data

1 - Taking Crop production between years 1993 and 2017 as most countries have data in that year range

2 - We're eliminating countries which don't have crop yield and production information between 1993 and 2017

3 - As such from the above identified 113 countries, five of them, viz. Montenegro, Serbia, Sudan, Luxembourg, Belgium and Denmark don't have

crop production and yield info between 1993 and 2017 - they only have in the below range

Country Year range

Montenegro 2006-2018

Serbia 2006-2018

Sudan 2012-2018

Luxembourg 2000-2018

Belgium 2000-2018

Denmark 2010-2018

As such removing them from the list

Also, while we have CO2 and Surface temprature anomaly well before 1993, that will not help us in our analysis and thus will be removed from

our range

The below deletes will remove the out of bound data and those not satisfiying the date range requirements

\*/

DELETE FROM [dbo].[CO2\_Data]

WHERE [Year] < 1993 OR Country IN ('Montenegro', 'Serbia', 'Sudan', 'Luxembourg', 'Belgium', 'Denmark')

DELETE FROM [dbo].[Maize\_Production]

WHERE [Year] < 1993 OR Country IN ('Montenegro', 'Serbia', 'Sudan', 'Luxembourg', 'Belgium', 'Denmark')

DELETE FROM [dbo].[Maize\_Yields]

WHERE [Year] < 1993 OR Country IN ('Montenegro', 'Serbia', 'Sudan', 'Luxembourg', 'Belgium', 'Denmark')

DELETE FROM [dbo].[Rice\_Production]

WHERE [Year] < 1993 OR Country IN ('Montenegro', 'Serbia', 'Sudan', 'Luxembourg', 'Belgium', 'Denmark')

DELETE FROM [dbo].[Rice\_Yields]

WHERE [Year] < 1993 OR Country IN ('Montenegro', 'Serbia', 'Sudan', 'Luxembourg', 'Belgium', 'Denmark')

DELETE FROM [dbo].[Surface\_Temperature\_Anomaly]

WHERE [Year] < 1993 OR Country IN ('Montenegro', 'Serbia', 'Sudan', 'Luxembourg', 'Belgium', 'Denmark')

DELETE FROM [dbo].[Wheat\_Production]

WHERE [Year] < 1993 OR Country IN ('Montenegro', 'Serbia', 'Sudan', 'Luxembourg', 'Belgium', 'Denmark')

DELETE FROM [dbo].[Wheat\_Yields]

WHERE [Year] < 1993 OR Country IN ('Montenegro', 'Serbia', 'Sudan', 'Luxembourg', 'Belgium', 'Denmark')

-- The remaining data contains - Maize, Rice and Wheat yield and production data from 1993 to 2018

-- And CO2 and Surface temperature data of 106 countries from 1993 t0 2017

-- Now data is ready for analysis and visualization

1. Python Code for Data processing and Visualization

#!/usr/bin/env python

# coding: utf-8

# # University of Arkansas at Little Rock

#

# IFSC 4345/5345 Information Visualization Project

#

# Team H

#

# - Bryan Fendley

#

# - Daniel Noble

#

# - Fantahun Zikie

#

# November-December 2020

# In[1]:

# Import Necessary Modules

from \_\_future\_\_ import print\_function

import pandas as pd

import numpy as np

import plotly.offline as po

import chart\_studio.plotly as py

import plotly.graph\_objects as go

import plotly.express as px

from ipywidgets import interact, interactive, fixed, interact\_manual

import ipywidgets as widgets

# In[2]:

# Create dataframes from source data

CO2\_df = pd.read\_csv("C:/Users/Fantahun/UALRProj/CO2\_Data.csv")

Maize\_Prd\_df = pd.read\_csv("C:/Users/Fantahun/UALRProj/Maize\_Production.csv")

Maize\_Yld\_df = pd.read\_csv("C:/Users/Fantahun/UALRProj/Maize\_Yields.csv")

Rice\_Prd\_df = pd.read\_csv("C:/Users/Fantahun/UALRProj/Rice\_Production.csv")

Rice\_Yld\_df = pd.read\_csv("C:/Users/Fantahun/UALRProj/Rice\_Yields.csv")

Surf\_Temp\_df = pd.read\_csv("C:/Users/Fantahun/UALRProj/Surface\_Temperature\_Anomaly.csv")

Wheat\_Prd\_df = pd.read\_csv("C:/Users/Fantahun/UALRProj/Wheat\_Production.csv")

Wheat\_Yld\_df = pd.read\_csv("C:/Users/Fantahun/UALRProj/Wheat\_Yields.csv")

# In[3]:

CO2\_df.head()

# Maize\_Prd\_df

# Maize\_Yld\_df

# Rice\_Prd\_df

# Rice\_Yld\_df

# Surf\_Temp\_df

# Wheat\_Prd\_df

# Wheat\_Yld\_df

# In[4]:

fig = px.line(CO2\_df, x="Year", y="CO2\_Mil\_Tones", title="Country CO2 Production in Million tones from 1993 to 2018",

color="Country")

fig.show()

# In[3]:

all\_data\_df = CO2\_df

for i in [Maize\_Prd\_df, Maize\_Yld\_df, Rice\_Prd\_df, Rice\_Yld\_df, Surf\_Temp\_df, Wheat\_Prd\_df, Wheat\_Yld\_df]:

all\_data\_df = pd.merge(all\_data\_df, i, on=['Code','Country', 'Year'], how='inner')

years = all\_data\_df.Year.unique()

# In[4]:

# Interactive plot - showing country production and yield information

def plot\_country\_info(country='World', crop\_info ='All'):

# First merge all data (inner join) to create one data source all\_data\_df

all\_data\_df = CO2\_df

for i in [Maize\_Prd\_df, Maize\_Yld\_df, Rice\_Prd\_df, Rice\_Yld\_df, Surf\_Temp\_df, Wheat\_Prd\_df, Wheat\_Yld\_df]:

all\_data\_df = pd.merge(all\_data\_df, i, on=['Code','Country', 'Year'], how='inner')

years = all\_data\_df.Year.unique()

layout = go.Layout(plot\_bgcolor='rgba(192,192,192,0.12)')

fig = go.Figure(layout = layout)

x\_data = years

var = {}

y\_data = []

y\_ppl = []

if country == 'world' or country == 'World':

if crop\_info == 'Maize Prod':

for y in years:

var[y] = np.sum(np.asarray(all\_data\_df[all\_data\_df['Year']== y].Maize\_Prod\_Tones))

elif crop\_info == 'Maize Yield':

for y in years:

var[y] = np.mean(np.asarray(all\_data\_df[all\_data\_df['Year']== y].Maize\_Yield\_Hg\_Per\_Ha))

elif crop\_info == 'Rice Prod':

for y in years:

var[y] = np.sum(np.asarray(all\_data\_df[all\_data\_df['Year']== y].Rice\_Prod\_Tones))

elif crop\_info == 'Rice Yield':

for y in years:

var[y] = np.mean(np.asarray(all\_data\_df[all\_data\_df['Year']== y].Rice\_Yield\_Hg\_Per\_Ha))

elif crop\_info == 'Wheat Prod':

for y in years:

var[y] = np.sum(np.asarray(all\_data\_df[all\_data\_df['Year']== y].Wheat\_Prod\_Tones))

elif crop\_info == 'Wheat Yield':

for y in years:

var[y] = np.mean(np.asarray(all\_data\_df[all\_data\_df['Year']== y].Wheat\_Yield\_Hg\_Per\_Ha))

else:

var = {}

else:

booleans = []

for c in all\_data\_df.Country:

if c == country:

booleans.append(True)

else:

booleans.append(False)

is\_country = pd.Series(booleans)

all\_data\_df\_for\_country = all\_data\_df[is\_country]

if crop\_info == 'Maize Prod':

for y in years:

var[y] = all\_data\_df\_for\_country[all\_data\_df\_for\_country['Year'] == y].Maize\_Prod\_Tones

elif crop\_info == 'Maize Yield':

for y in years:

var[y] = all\_data\_df\_for\_country[all\_data\_df\_for\_country['Year'] == y].Maize\_Yield\_Hg\_Per\_Ha

elif crop\_info == 'Rice Prod':

for y in years:

var[y] = all\_data\_df\_for\_country[all\_data\_df\_for\_country['Year'] == y].Rice\_Prod\_Tones

elif crop\_info == 'Rice Yield':

for y in years:

var[y] = all\_data\_df\_for\_country[all\_data\_df\_for\_country['Year'] == y].Rice\_Yield\_Hg\_Per\_Ha

elif crop\_info == 'Wheat Prod':

for y in years:

var[y] = all\_data\_df\_for\_country[all\_data\_df\_for\_country['Year'] == y].Wheat\_Prod\_Tones

elif crop\_info == 'Wheat Yield':

for y in years:

var[y] = all\_data\_df\_for\_country[all\_data\_df\_for\_country['Year'] == y].Wheat\_Yield\_Hg\_Per\_Ha

else:

var = {}

y\_data = np.fromiter(var.values(), dtype=np.int64)

y\_ppl = all\_data\_df\_for\_country['Population']

fig.add\_trace(go.Scatter(x=x\_data, y=y\_data, name = crop\_info, mode='lines+markers', connectgaps=True))

fig.add\_trace(go.Scatter(x=x\_data, y=y\_ppl, name='Population', mode='lines+markers', connectgaps=True))

fig.update\_xaxes(

title\_text = crop\_info,

title\_font = {"size": 20},

title\_standoff = 25)

fig.update\_yaxes(

title\_text = "Production/Yield in tones",

title\_font = {"size": 20},

title\_standoff = 25)

fig.update\_layout(title="Crop Production/Yield data Per Country from 1993 to 2017")

fig.show()

# print(y\_data)

# plot\_country\_info('China', 'Rice Prod')

# Add interactivity with text boxes then with drop down

countries = all\_data\_df.Country.unique()

crops = ['Maize Prod', 'Maize Yield', 'Rice Prod', 'Rice Yield', 'Wheat Prod', 'Wheat Yield']

# countries

interact(plot\_country\_info, country = countries, crop\_info = crops)

### Next: We can also add interactive graph on map

### Export the visualization as a standalone web app using Voila

## First install it using pip - at the cmd

# In[7]:

all\_data\_df

# all\_data\_df\_Prod\_5 = all\_data\_df\_Prod[all\_data\_df\_Prod['Year'].isin(yr\_list)]

all\_data\_df\_Prod = all\_data\_df[['Country', 'Year', 'Maize\_Prod\_Tones', 'Rice\_Prod\_Tones', 'Wheat\_Prod\_Tones' ]]

# all\_data\_df\_Prod\_5

all\_data\_df\_Prod

# In[8]:

# Interactive barchart - shows country crop production info for the years [1995, 2000, 2005, 2010, 2015]

all\_data\_df\_Prod = all\_data\_df[['Country', 'Year', 'Maize\_Prod\_Tones', 'Rice\_Prod\_Tones', 'Wheat\_Prod\_Tones' ]]

yr\_list = [1995, 2000, 2005, 2010, 2015]

all\_data\_df\_Prod\_5 = all\_data\_df\_Prod[all\_data\_df\_Prod['Year'].isin(yr\_list)]

def plot\_country\_prod\_chart(country):

list(all\_data\_df\_Prod.columns)

# yr\_list = [1995, 2000, 2005, 2010, 2015]

# all\_data\_df\_Prod\_5 = all\_data\_df\_Prod[all\_data\_df\_Prod['Year'].isin(yr\_list)]

all\_data\_df\_Prod\_5\_country = all\_data\_df\_Prod\_5[all\_data\_df\_Prod\_5['Country'] == country]

data = [

go.Bar( name = 'Maize Poduction', x = yr\_list, y = all\_data\_df\_Prod\_5\_country['Maize\_Prod\_Tones']),

go.Bar( name = 'Rice Production', x = yr\_list, y = all\_data\_df\_Prod\_5\_country['Rice\_Prod\_Tones']),

go.Bar( name = 'Wheat Production', x = yr\_list, y = all\_data\_df\_Prod\_5\_country['Wheat\_Prod\_Tones'])

]

layout = go.Layout(plot\_bgcolor='rgba(192,192,192,0.12)')

fig = go.Figure(data = data, layout = layout)

fig.update\_xaxes(

title\_text = 'Year',

title\_font = {"size": 20},

title\_standoff = 25)

fig.update\_yaxes(

title\_text = "Production in tones",

title\_font = {"size": 20},

title\_standoff = 25)

fig.update\_layout(title="Crop Production of " + country + " in years 1995, 2000, 2005, 2010, 2015")

fig.show()

countries = all\_data\_df\_Prod\_5.Country.unique()

interact(plot\_country\_prod\_chart, country = countries)

# In[ ]:

# In[5]:

# Evaluate the Correlation between country CO2 production and the Surface temprature

years = all\_data\_df.Year.unique()

def plot\_temp\_and\_co2(country='World'):

layout1 = go.Layout(plot\_bgcolor='rgba(192,192,192,0.12)')

layout2 = go.Layout(plot\_bgcolor='rgba(192,192,192,0.12)')

fig1 = go.Figure(layout = layout1)

fig2 = go.Figure(layout = layout2)

fig3 = go.Figure()

# Get entries for a given country

booleans = []

for c in all\_data\_df.Country:

if c == country:

booleans.append(True)

else:

booleans.append(False)

is\_country = pd.Series(booleans)

all\_data\_df\_for\_country = all\_data\_df[is\_country]

Country\_Surface\_Temp = all\_data\_df\_for\_country[['Year', 'Surface\_Temp\_Anomaly']]

Country\_CO2\_Prod = all\_data\_df\_for\_country[['Year', 'CO2\_Mil\_Tones']]

x\_data = years

CO2\_y\_data = []

Sur\_y\_data = []

CO2\_y\_data\_dict = {}

Sur\_y\_data\_dict = {}

for yr in years:

CO2\_y\_data\_dict[yr] = Country\_CO2\_Prod[Country\_CO2\_Prod['Year'] == yr].CO2\_Mil\_Tones

Sur\_y\_data\_dict[yr] = Country\_Surface\_Temp[Country\_Surface\_Temp['Year'] == yr].Surface\_Temp\_Anomaly

CO2\_y\_data = np.fromiter(CO2\_y\_data\_dict.values(), dtype=np.float)

Sur\_y\_data = np.fromiter(Sur\_y\_data\_dict.values(), dtype=np.float)

fig1.add\_trace(go.Scatter(x=x\_data, y=CO2\_y\_data, name = 'CO2 Production', mode='lines+markers', connectgaps=True))

fig2.add\_trace(go.Scatter(x=x\_data, y=Sur\_y\_data, name='Surface Temprature Anomaly', mode='lines+markers', connectgaps=True))

fig1.update\_xaxes(

title\_text = 'Year',

title\_font = {"size": 20},

title\_standoff = 25)

fig1.update\_yaxes(

title\_text = "CO2 Production in Mil Tones",

title\_font = {"size": 20},

title\_standoff = 25)

fig1.update\_layout(title="Country Specific CO2 Production in Million Tones Per Year")

fig2.update\_xaxes(

title\_text = 'Year',

title\_font = {"size": 20},

title\_standoff = 25)

fig2.update\_yaxes(

title\_text = "Surface Temprature Anomaly",

title\_font = {"size": 20},

title\_standoff = 25)

fig2.update\_layout(title="Surface Temprature Anomaly - from 1951-1980 global average temperature")

# Heatmap for the Surface temprature anomaly

# fig3.add\_trace(go.Heatmap(

# z = Country\_Surface\_Temp["Surface\_Temp\_Anomaly"],

# colorbar=dict(

# title="Surface Heat",

# titleside="top",

# tickmode="array",

# tickvals=[-1, 1, 5],

# ticktext=["Cooler", "Mild change", "Hotter"],

# ticks="outside" )))

# Calculate the Pearson Correlation Coefficient for CO2 Prod and Surface Temprature Anomaly

co2\_temp\_corr\_data = all\_data\_df\_for\_country[['Year', 'CO2\_Mil\_Tones', 'Surface\_Temp\_Anomaly']]

print(co2\_temp\_corr\_data.corr(method ='pearson'))

fig1.show()

fig2.show()

# fig3.show()

countries = all\_data\_df.Country.unique()

interact(plot\_temp\_and\_co2, country = countries)

# In[7]:

# This section is for depicting the surface temprature data on global map

# for reading geo json data of the world for world map view

import json

# First download the countries geo json data

# And read the world countries geo.json file

world\_countries = json.load(open('World\_Countries.geo.json','r'))

# check the data features

# world\_countries['features'][0]['properties']['name']

# world\_countries['features'][0].keys()

world\_countries['features'][110]['properties']

# In[8]:

country\_id\_map = {}

for feature in world\_countries['features']:

feature['id'] = int(feature['properties']['iso\_n3'])

country\_id\_map[feature['properties']['name\_long']] = feature['id']

# check the country\_id\_map content

# country\_id\_map.keys()

country\_list = Surf\_Temp\_df['Country'].unique()

#mod\_Surf\_Temp\_df['Country']

# create a dict from country\_id\_map that contains countries only from the country\_list under consideration (in Surf\_Temp\_df)

new\_country\_id\_map = dict()

for (key, value) in country\_id\_map.items():

if key in country\_list:

new\_country\_id\_map[key] = value

new\_country\_id\_map.keys()

#country\_list

# There are few countries in the Surf\_Temp\_df not in the countries geo json file - thus lets create a modified

# Surf\_Temp\_df as mod\_Surf\_Temp\_df

mod\_Surf\_Temp\_df = Surf\_Temp\_df[Surf\_Temp\_df['Country'].isin(new\_country\_id\_map.keys())]

mod\_Surf\_Temp\_df

# In[10]:

# Add one column to the Surf\_Temp\_df data

def plot\_Surf\_temp\_map(year):

mod\_Surf\_Temp\_df['country\_id'] = mod\_Surf\_Temp\_df['Country'].apply(lambda x: new\_country\_id\_map[x])

#mod\_Surf\_Temp\_df.head()

mod\_Surf\_Temp\_df\_country = mod\_Surf\_Temp\_df[mod\_Surf\_Temp\_df['Year'] == year]

mod\_Surf\_Temp\_df\_country

fig = px.choropleth(mod\_Surf\_Temp\_df\_country, locations='country\_id', geojson=world\_countries, color='Surface\_Temp\_Anomaly')

fig.show()

years = all\_data\_df.Year.unique()

interact(plot\_Surf\_temp\_map, year = years)

# In[16]:

fig = px.choropleth(mod\_Surf\_Temp\_df\_1993, locations='country\_id', geojson=world\_countries, color='Surface\_Temp\_Anomaly')

fig.show()