I. Introduction

A. Motivation.

I have always been very curious on energy consumption. It is something we use every day to accomplish amazing tasks. It has the ability to save lives or simply entertain someone. Its wide range of uses and abilities make it so crucial to our everyday lives. On a related note, I have also been interested in how we produce our energy, at what cost, and how its distributed. There are consequences to our actions when we create energy, particularly in how it affects the environment. We have seen in the past few decades extreme changes to climates all over the world, with increasing extreme weather related events. This is partly due to our heavy use of fossil fuels and emission of massive amounts of CO_2, methane, and many other gasses. As a result, I seek to explore this dataset and analyze the areas concerning energy production, emissions, and agriculture as these are all affected by our production and use of energy quite greatly

B. Objective Statement

I seek to find all relatable parameters of energy production, consumption, emissions, and agriculture of every country. I then wish to analyze the data on an annual basis and compare different types of emissions to see how countries perform on the global stage. For example, I would like to see how top energy producers produce their energy, and if any are particularly more efficient than others. I am also curious to see how those who do produce methane gas from animals manage to produce energy. Perhaps there is a correlation between animal waste energy production and methane emissions. Overall, I will be seeking to find interesting trends in the creative use of our assets to produce renewable energies.

II. Data Profile

A. Initial dataset

Our dataset, as is well known, comes from the world bank and consists of “World Development Indicators” (WDI). This data set is ~422,000x62 rowsxcolumns. The data consists of several indicators of potentially valuable information. These include values for percentage of farmland usage, access to electricity, economic data, and much more. These range in values from percentages to numbers depending on the indicator it is meant to represent and are sometimes raw values or scaled. Each of these values are then provided on a yearly basis starting from 1960 to 2017. However this data is very sparse and includes many null/ missing values; as is to be expected for such large data sets. Some countries may not have existed during certain years, or just did not have the requested information. This data is clearly too large to be used for our small purposes of studying energy and agricultural practices. Accordingly, we narrow down the data.

B. Narrowing down

I first start to explore the indicator column to find those that may be of interest.

fx Electric power consumption (kWh per capita)

C	D	E	F	G	H
Electric po	EG.USE.ELEC.KH.PC				
Electric po	EG.ELC.LOSS.ZS				
Electricity p	EG.ELC.COAL.ZS				
Electricity p	EG.ELC.HYRO.ZS				
Electricity p	EG.ELC.NGAS.ZS				
Electricity p	EG.ELC.NUCL.ZS				
Electricity p	EG.ELC.PETR.ZS				
Electricity p	EG.ELC.FOSL.ZS				
Electricity p	EG.ELC.RNWX.ZS				
Electricity p	EG.ELC.RNWX.KH				

I first found some data on electric power consumption, and how much of it was produced by particular categories.I found the following relatable indicators as well.

fx CO2 emissions from solid fuel consumption (% of total)

C	D	E	F	G	H	I
CO2 emissi	EN.ATM.CO2E.KD.GD					
CO2 emissi	EN.ATM.CO2E.PP.GD.KD					
CO2 emissi	EN.ATM.CO2E.PP.GD					
CO2 emissi	EN.ATM.C(	59535.4	65119.82	74322.01	87853.79	103
CO2 emissi	EN.ATM.C(	0.643689	0.685151	0.760855	0.874941	0.99
CO2 emissi	EN.CO2.ETOT.ZS					
CO2 emissi	EN.ATM.C(	5.041292	5.521749	6.096959	5.976445	7.2
CO2 emissi	EN.ATM.CO2E.GF.KT					
CO2 emissi	EN.ATM.C(	84.85147	85.06874	86.38428	80.82192	63.2
CO2 emissi	EN.ATM.C(	50539.8	55421.99	64231.95	71037.64	653(
CO2 emissi	EN.CO2.MANF.ZS					
CO2 emissi	EN.CO2.OTHX.ZS					
CO2 emissi	EN.CO2.BLDG.ZS					
CO2 emissi	EN.ATM.C(	4.726981	4.169484	2.774487	2.297026	2.56
CO2 emissi	EN.ATM.CO2E.SF.KT					
CO2 emissi	EN.CO2.TRAN.ZS					
CO2 intens	EN.ATM.CO2E.EG.ZS					

fx Energy related methane emissions (% of total)

C	D	E	F	G	H
Employme	SL.EMP.1524.SP.NE.ZS				
Energy imp	EG.IMP.CONS.ZS				
Energy inte	EG.EGY.PRIM.PP.KD				
Energy rela	EN.ATM.METH.EG.ZS				

fx Energy imports, net (% of energy use)

C	D	E	F	G
Energy imp	EG.IMP.CONS.ZS			
Energy inte	EG.EGY.PRIM.PP.KD			
Energy rela	EN.ATM.METH.EG.ZS			
Energy use	EG.USE.PCAP.KG.OE			
Energy use	EG.USE.COMM.GD.PP.KD			

fx Nitrous oxide emissions (thousand metric tons of CO2 equivalent)

C	D	E	F	G	H	I
Nitrous oxi	EN.ATM.NOXE.ZG					
Nitrous oxi	EN.ATM.NOXE.KT.CE					
Nitrous oxi	EN.ATM.NOXE.EG.ZS					
Nitrous oxi	EN.ATM.NOXE.EG.KT.CE					

fx Renewable electricity output (% of total electricity output)

C	D	E	F	G	H	I
Renewable	EG.ELC.RNEW.ZS					
Renewable	EG.FEC.RNEW.ZS					

We clearly have a lot of relatable data. There is clearly a lot of data on CO_2 emissions, which is useful, but also some more minute information on nitrous oxide emissions, methane emissions, renewable energy production, and even imported and exported energy. This has allowed me to narrow down my topics of research quite a bit from the near 500000 rows. Clearly this is still a lot, but made me start with one topic. Are there a relationships between energy related methane emissions and renewable energy production. This will be limited to countries of large capacity, such as the United States, Russia, Germany, France, United Kingdom, and more or less depending on data availability.

III. Hypothesis 1: How does agricultural land usage relate to agriculture related methane emissions?

A. Processing the data

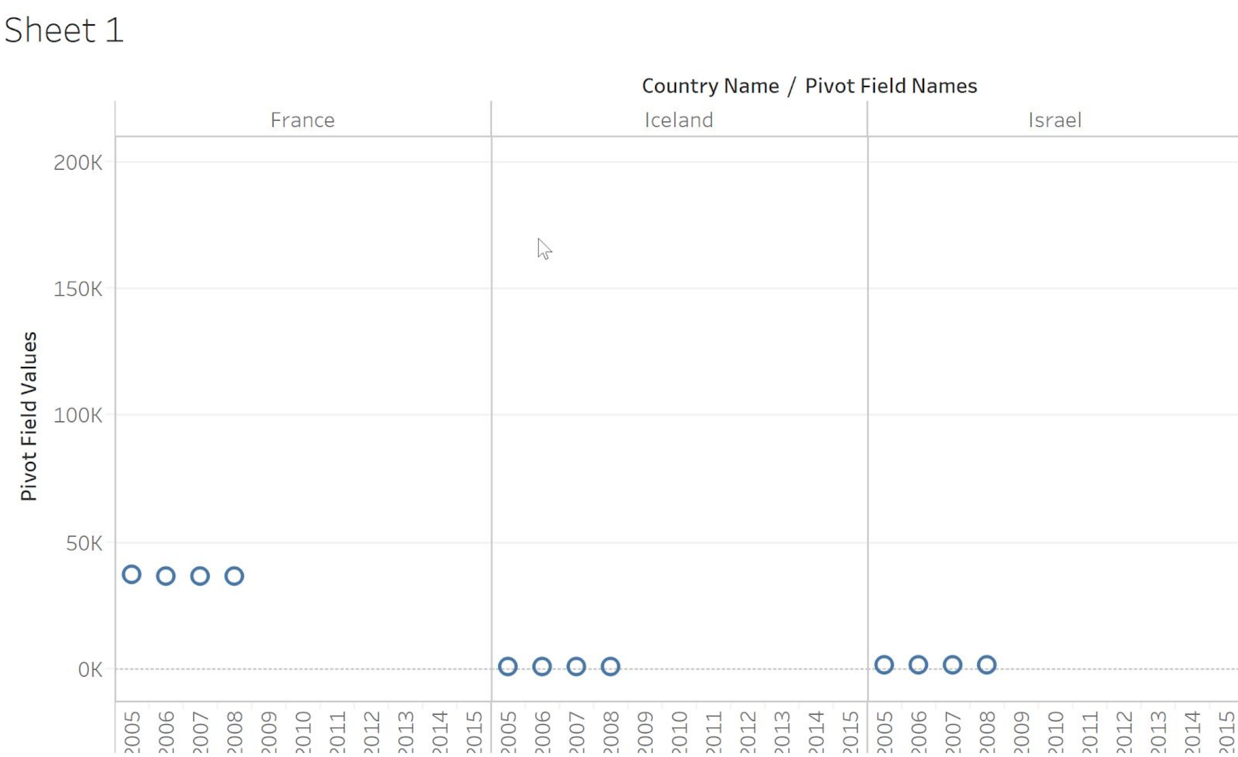
I used python scripts to grab only the rows of interest from the large data set.

```
agriculturalStuff = ['EN.ATM.METH.AG.KT.CE','AG.LND.AGRI.K2','AG.LND.AGRI.ZS'] #agriucltural meth
agrDF = newdf.loc[newdf['Indicator Code'].isin(agriculturalStuff)]
print(agrDF.head(10))
print(agrDF.shape)
print(newdf.shape)
agrDF.to_csv("/home/avanroi1/HCDE-511/A3/data/justFarmStuff.csv",",")
```

We can see that the data set is reduced by a large amount.

```
(792, 63)
(422400, 63)
```

However, there are some missing values when I first investigated



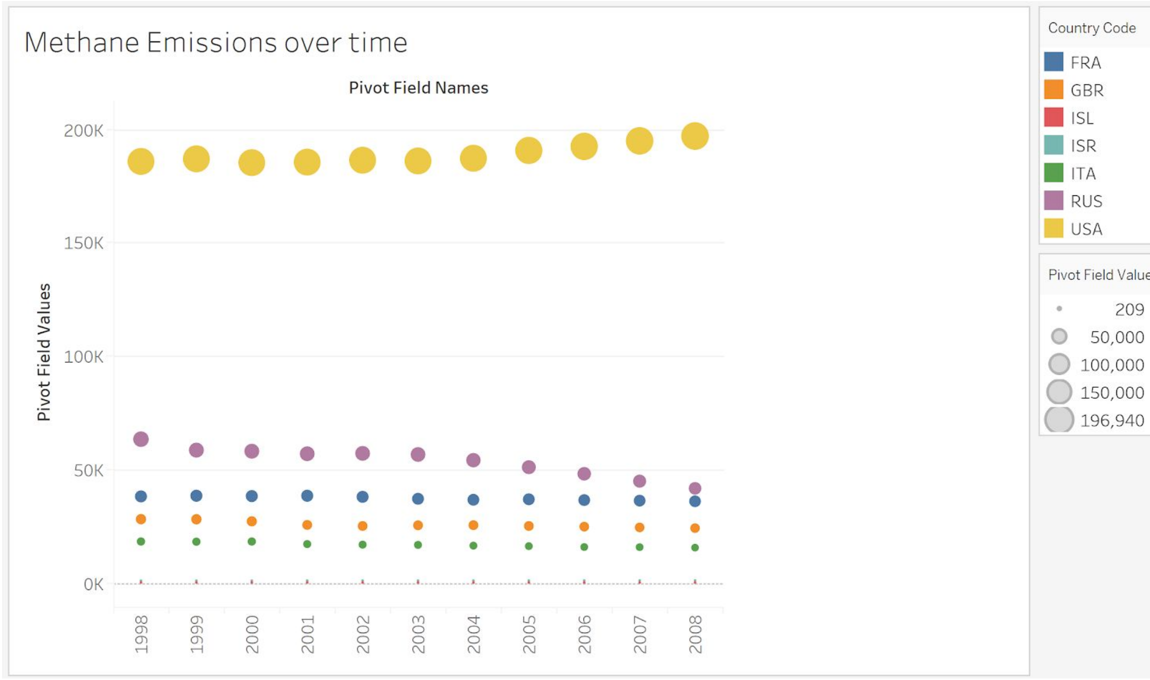
It appears that there is no relatable data past 2008 for whatever reason. This is something I may want to investigate in further detail later. Perhaps it has been absorbed into a new column.

B. Final version of Data

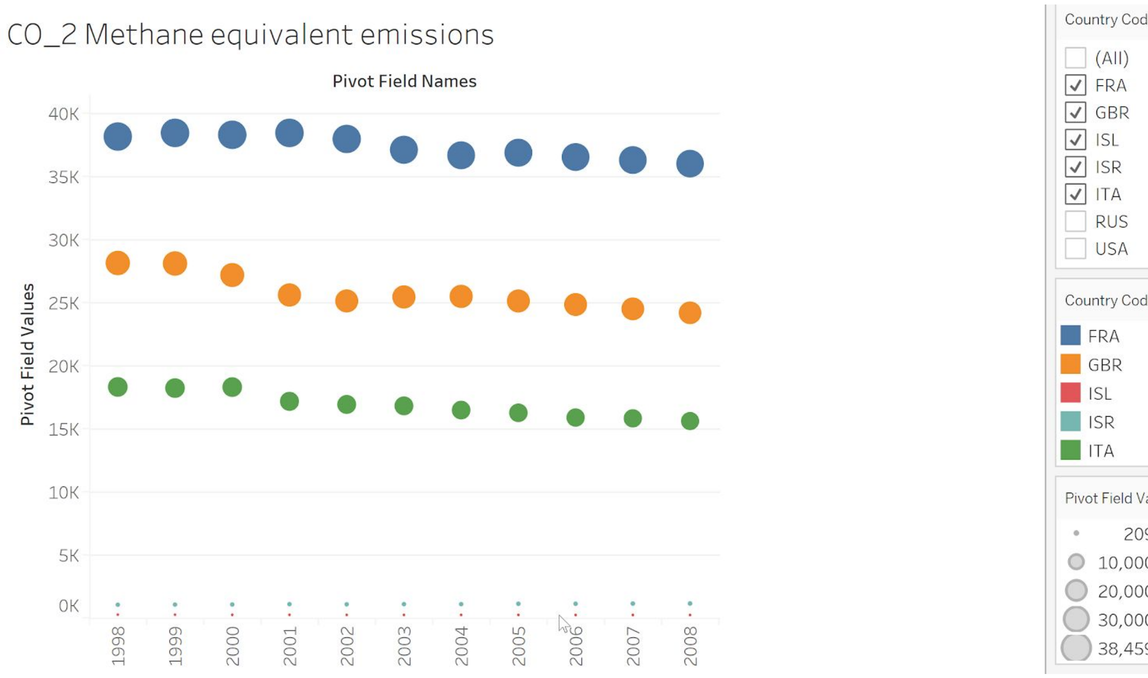
I ultimately narrowed down my data to the years between 1998 and 2008 and between larger countries such as the United States and Russia, and smaller countries including France and Israel. The resulting exploration narrowed down the data as will be seen.

C. Exploration and Results

1. Initial visualization

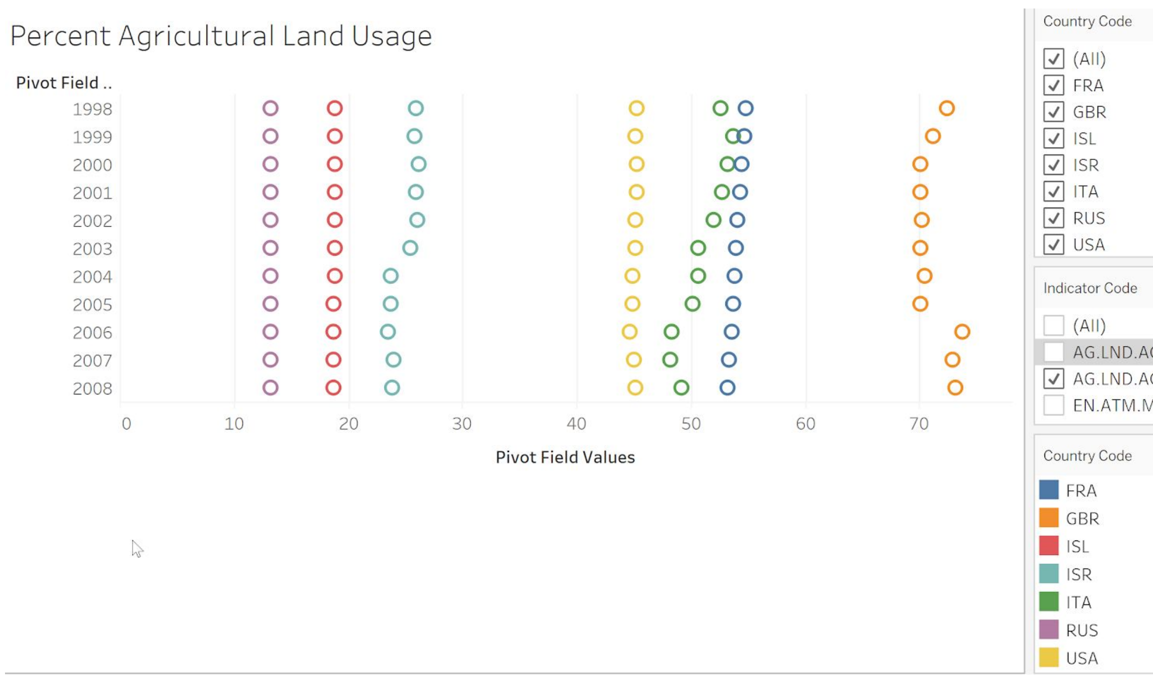


The y-axis represents the CO_2 equivalent of agricultural methane related emissions. First thing we noticed is that the USA has dominated in this category and continues to increase its output. We can see the USA increased over a decade while Russia decreased by quite an amount. Meanwhile, other countries such as France seem to stagnate, but this is simply due to our visualization! It violates a principle of apprehension and integrity as the data is unscaled, thus larger countries shadow the changes of smaller countries. Consequently, I decided to take a closer look at two different views. I wanted to look at how the relationships may relate for big countries, like the USA and Russia, and smaller countries, such as the UK. This fits nicely into our mantra of “overview -> zoom” as we seek to better understand these patterns.

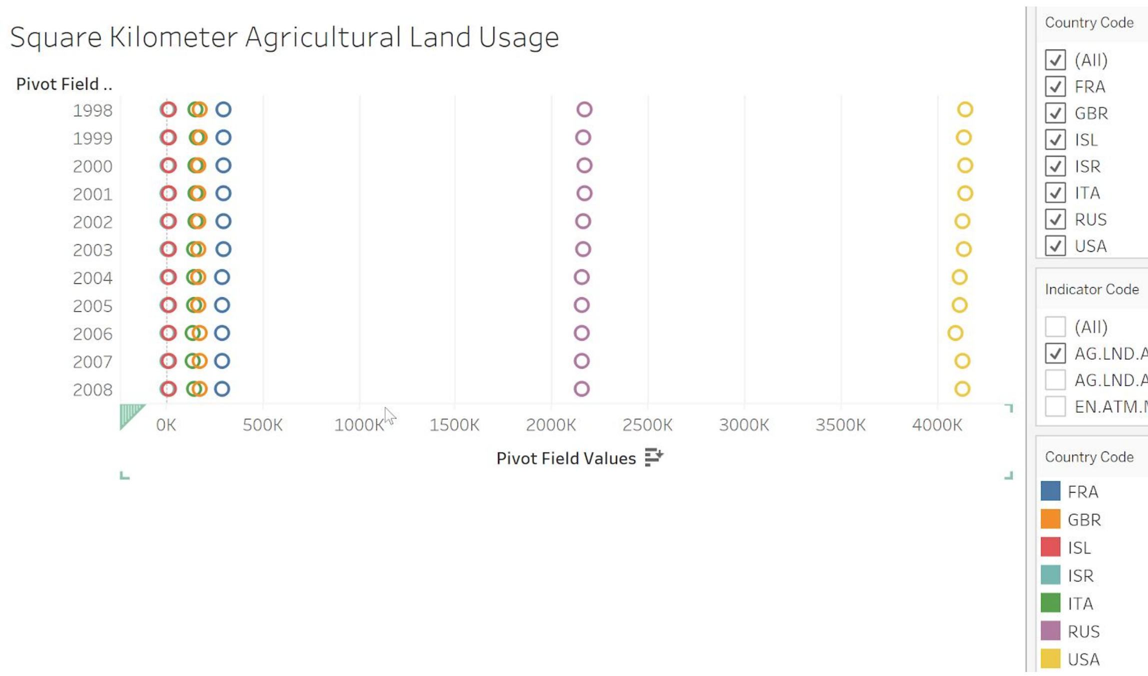


This graphic allowed me to better visualize how the methane emissions have changed for the smaller countries such as Italy and France without having the large countries diminish their changes. In particular, we can see slight decreases in France, Italy, and Great Britain. Since we couldnt see this in our previous iteration, you could say we were violating the principles of graphical integrity, as the scales were hiding the impacts of smaller countries and their changes over the years.

2. Agricultural Land Usage



Here, we can see the percentage of land used for agricultural purposes. Interestingly, we see unintuitive results. Neither the U.S. or Russia appear to change their percentage all too much. Despite their increases and decreases in methane emissions respectively. However Great Britain does, but it is an **increase**, despite the fact we see a decrease overall of methane emissions from that country, unlike Italy and France who decreased their percentage of land usage.



Looking at the raw amount of agricultural land usage, Russia and the United States dont vary all to much from year to year. If so it is perhaps in the order of tens of thousands of square kilometers compared to their million square kilometer usage, with the U.S. at about double of Russia's usage.

D. Conclusion
Overall, there seems to be no clear relationship between the amount of agricultural methane related emissions and agricultural land usage from the graphs. In large countries, no clear pattern is present as consistent land usage resulted in increases and decreases in methane emissions for the USA and Russia respectively. Meanwhile for smaller countries such as Great Britain, Italy, and France, we see contradicting results, where despite all of them decreasing emissions, their land usage increases and decreases. However, there is at least one observation we gain from this. The land that is used by a country for agriculture seems to be a strong predictor of the amount of agriculturally produced emissions. I use R to manipulate the data and create a linear model as highlighted below.

```
Call:
lm(formula = METHEMISS ~ amntLand, data = result)

Residuals:
   Min     1Q Median     3Q    Max 
-50579  -7616   3782  14731  26431 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 7.093e+03  2.640e+03   2.686  0.00889 ** 
amntLand    3.954e-02  1.493e-03  26.475  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 19190 on 75 degrees of freedom
Multiple R-squared:  0.9033,	Adjusted R-squared:  0.9021 
F-statistic: 700.9 on 1 and 75 DF,  p-value: < 2.2e-16
```

As we can see, there is a very strong positive linear relationship between these two values. For every square kilometer of land used for agriculture, we expect to see an additional .0395 metric tons of CO_2 produced. Further note that this was determined over ALL countries, not just the subsets shown above. This is a rather comprehensive statement that clearly has validity. This further agrees with our intuition which reflects the principle of apprehension.

E. Future Work

Of course, this was not a comprehensive study of all countries of particular sizes in a visual medium. This would take far too much time. I hope to have more time to group up data by size and aggregate their results on their averages to gain a better understanding of these relationships. However, for now, these results provide some interesting results on a small group of countries and their emissions.

IV. What is the relationship between percentage of renewable energy production and CO_2 emissions?

A. Processing the data

```
def getEnergyData():
    csv_file = ("/home/avanroi1/HCDE-511/A3/data/WDIData.csv")
    newdf = pd.read_csv(csv_file,',')
    netInterestEnergy = [
    'EN.ATM.CO2E.PC', #CO2 emissions (metric tons per capita)
    'EN.CO2.ETOT.ZS', #CO2 emissions from electricity and heat production, total (% of total fuel
    'EG.USE.ELEC.KH.PC',#Electric power consumption (kWh per capita)
    'EG.ELC.NUCL.ZS',#Electricity production from nuclear sources (% of total)
    'EG.ELC.HYRO.ZS',#Electricity production from hydroelectric sources (% of total)
    'EG.ELC.FOSL.ZS',#Electricity production from oil, gas and coal sources (% of total)
    'EG.ELC.RNWX.ZS',#Electricity production from renewable sources, excluding hydroelectric (% o
    ]
    energyDF = newdf.loc[newdf['Indicator Code'].isin(netInterestEnergy)]
```

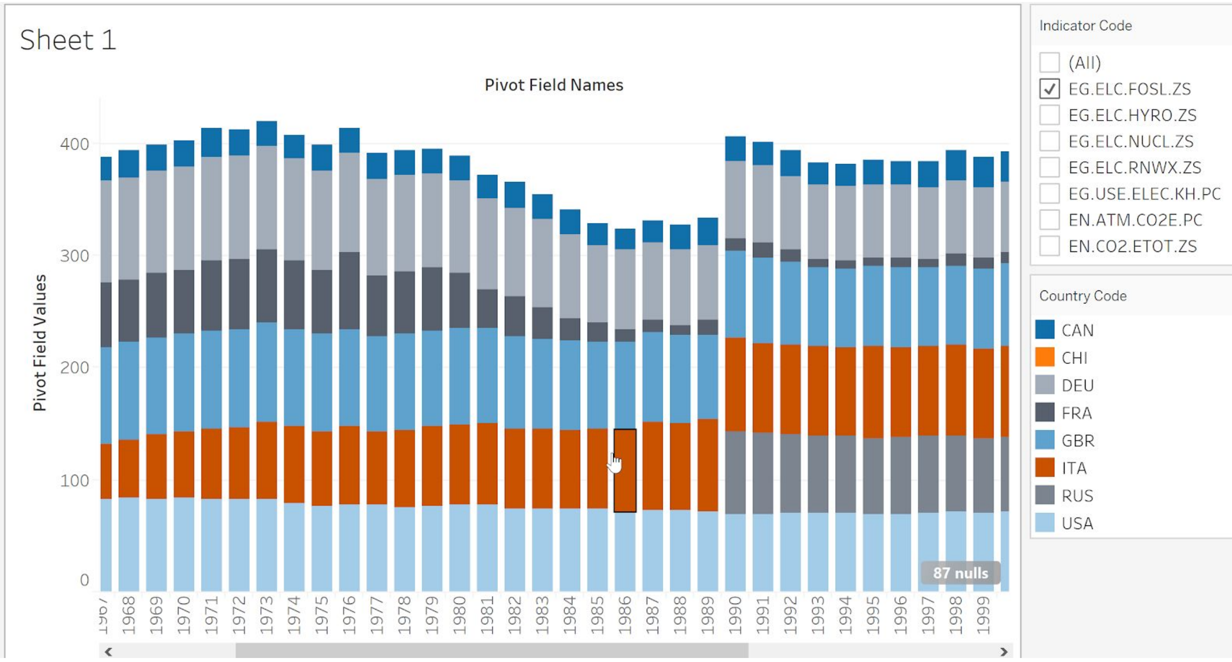
I am particularly interested in looking at how different percentages of energy produced by various sources impact CO2 emissions.

energyStuff.csv Country Name	energyStuff.csv Country Code	energyStuff.csv Indicator Name	energyStuff.csv Indicator Code	energyStuff.csv 1960	energyStuff.csv 1961	energyStuff.csv 1962	energyStuff.csv 1963	energyStuff.csv 1964	energyStuff.csv 1965
Arab World	ARB	CO2 emissions (metric ton...	EN.ATM.CO2E.PC	0.64	0.69	0.76	0.87	1.00	1.17
Arab World	ARB	CO2 emissions from electri...	EN.CO2.ETOT.ZS	null	null	null	null	null	null
Arab World	ARB	Electric power consumptio...	EG.USE.ELEC.KH.PC	null	null	null	null	null	null
Arab World	ARB	Electricity production fro...	EG.ELC.HYRO.ZS	null	null	null	null	null	null
Arab World	ARB	Electricity production fro...	EG.ELC.NUCL.ZS	null	null	null	null	null	null
Arab World	ARB	Electricity production fro...	EG.ELC.FOSL.ZS	null	null	null	null	null	null
Arab World	ARB	Electricity production fro...	EG.ELC.RNWX.ZS	null	null	null	null	null	null
Caribbean small states	CSS	CO2 emissions (metric ton...	EN.ATM.CO2E.PC	1.40	2.06	2.70	1.35	2.36	2.60

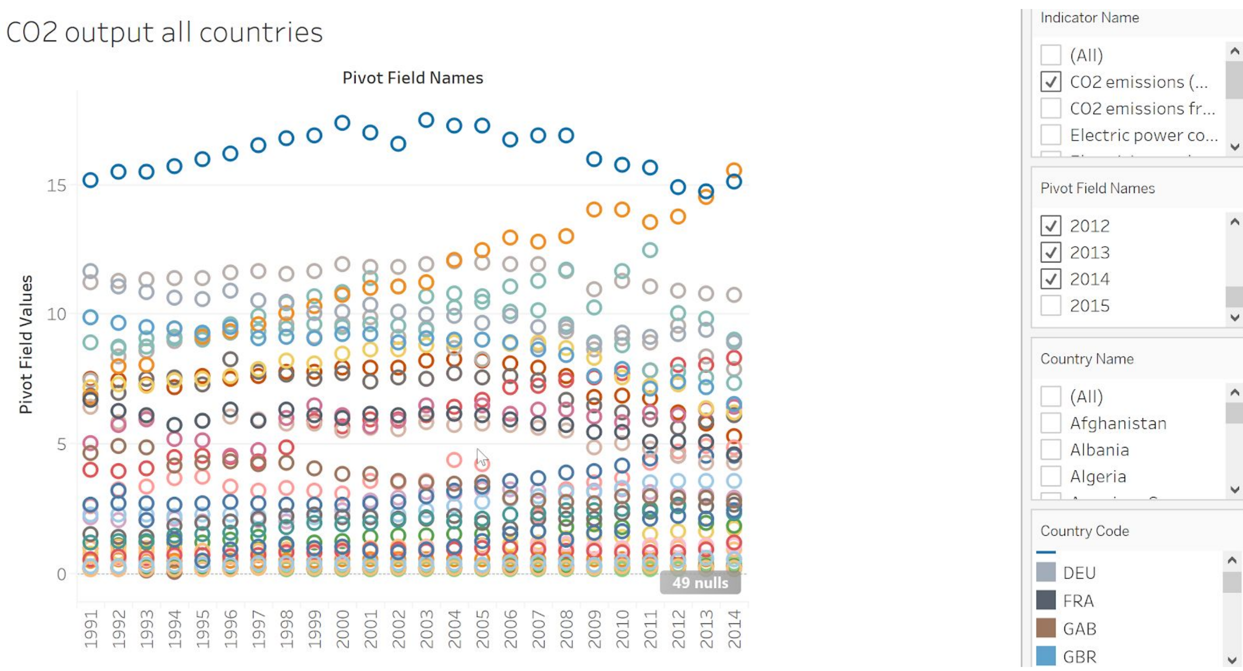
I noticed that similar to the last hypothesis, I would need to pivot on the years in order to do a better year by year analysis of this data.

energyStuff.csv Country Name	energyStuff.csv Country Code	energyStuff.csv Indicator Name	energyStuff.csv Indicator Code	Pivot Pivot Field Names	Pivot Pivot Field Values
Arab World	ARB	CO2 emissions (metric ton...	EN.ATM.CO2E.PC	1960	0.
Arab World	ARB	CO2 emissions from electri...	EN.CO2.ETOT.ZS	1960	null
Arab World	ARB	Electric power consumptio...	EG.USE.ELEC.KH.PC	1960	null
Arab World	ARB	Electricity production fro...	EG.ELC.HYRO.ZS	1960	null
Arab World	ARB	Electricity production fro...	EG.ELC.NUCL.ZS	1960	null
Arab World	ARB	Electricity production fro...	EG.ELC.FOSL.ZS	1960	null
Arab World	ARB	Electricity production fro...	EG.ELC.RNWX.ZS	1960	null

Now that our data has been formatted, I look for null values since we came across this issue last time.



As we can see here, there is no data on russian or german emissions until 1990, which makes a full analysis over time not possible between all countries.



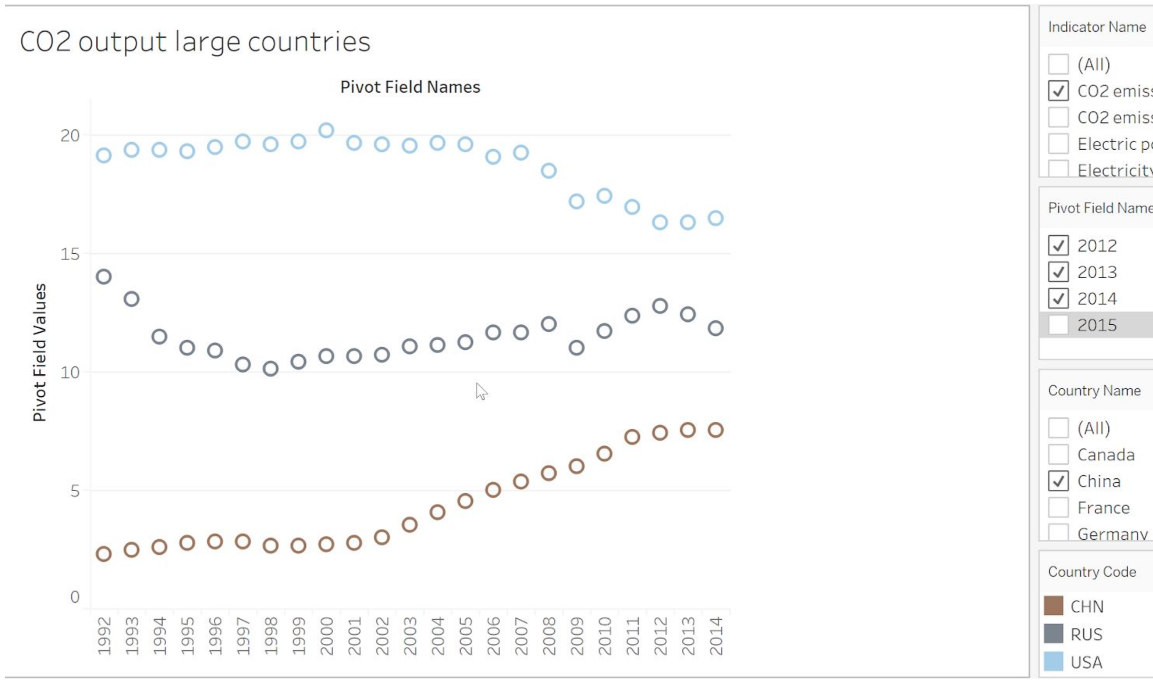
We further understand that attempting to analyze relationships between all countries is not feasible. Consequently, I will analyze the differences between small countries and large countries again from a subset of countries.

B. Final version of the data

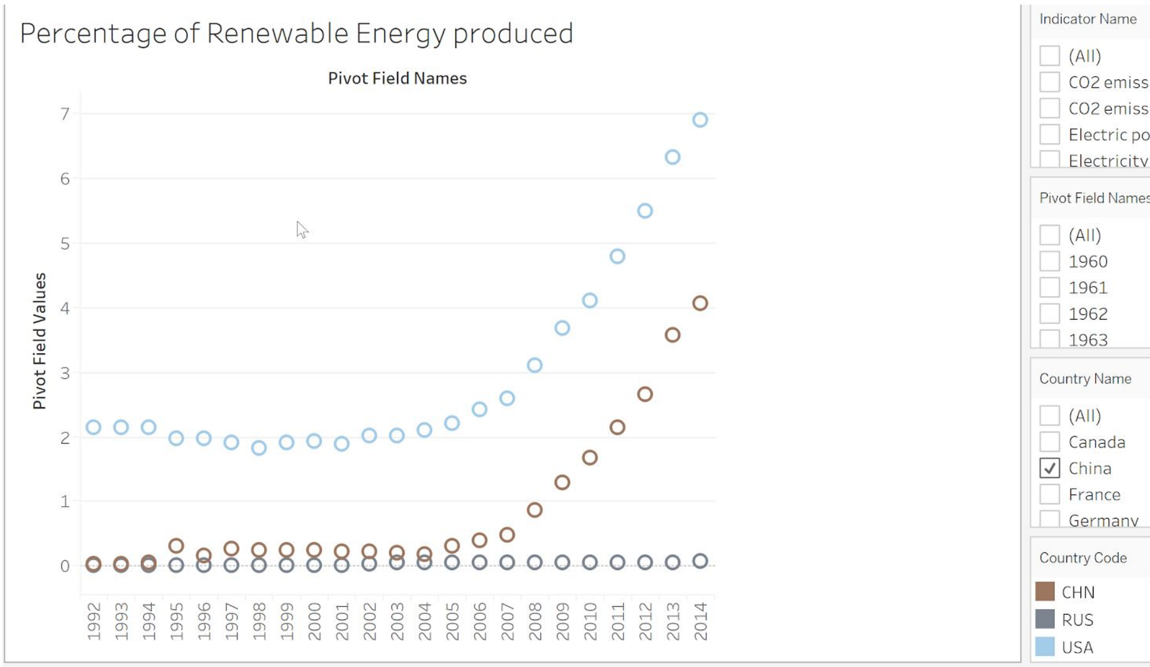
I ultimately decided to analyze the large countries of Russia, the USA, and China and the smaller countries of France, Italy, Germany, and Canada. I will analyze the decades between the 90s to early 2000s.

C. Exploration and Results

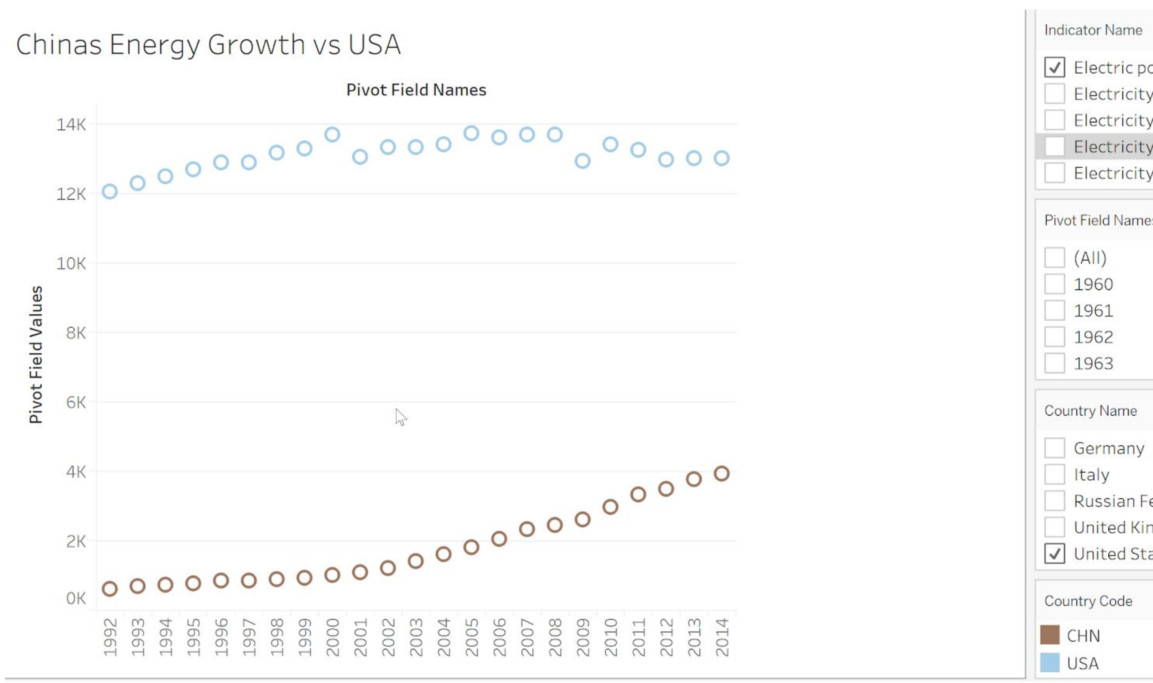
1. Large Country Emissions



Here we see over 2 decades worth of CO_2 emissions (in metric tons per capita) from our three large countries. We notice that the USA has reduced by quite a bit over these 20 some odd years, while russia has fluctuated and china has increased. In fact, China almost tripled its emission per capita since 1992. Now we seek to understand if the amount of renewable energy power produced is the explanation for these increases and decreases in emissions.

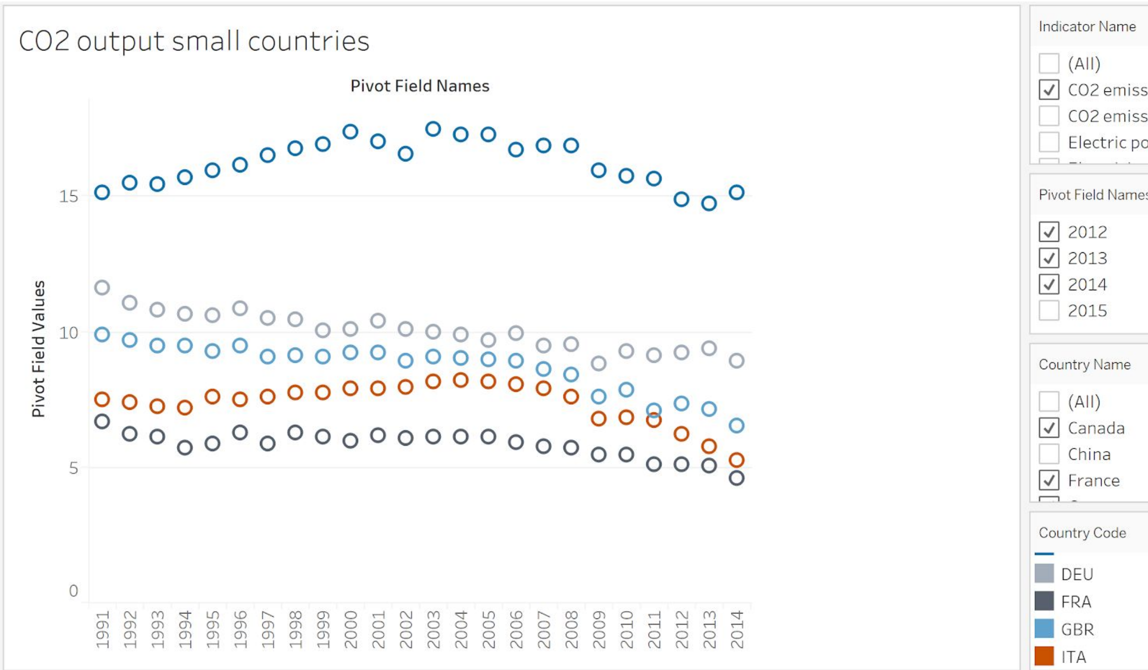


Interestingly, we can see efforts by both the United States and China to attempt to curve their impacts. Both climbing at similar times and even similar rates. Meanwhile, Russia appears to attempt very little to introduce renewable energy. However, this does not explain why China seems to have such high emission rates despite the increasing presence of renewable energies. So to achieve qualities of graphical excellence, I focus the next graphic on China and the USA and their energy production over this time.

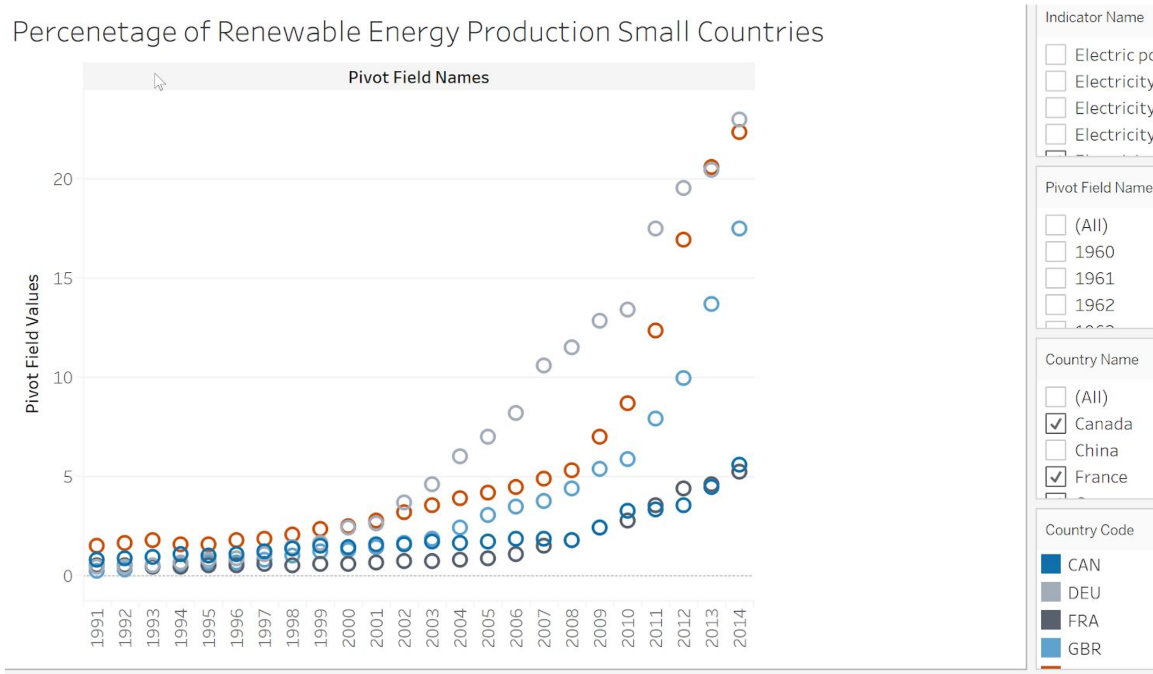


Here we gain an understanding of why our emission rates havent dropped as expected. Despite the large growth of renewable energies in China, their energy demands increased by over 700%! Unlike the United States, who produced much more energy, but has not changed much in the past 20+ years. Now we can see if similar relationships hold in smaller countries.

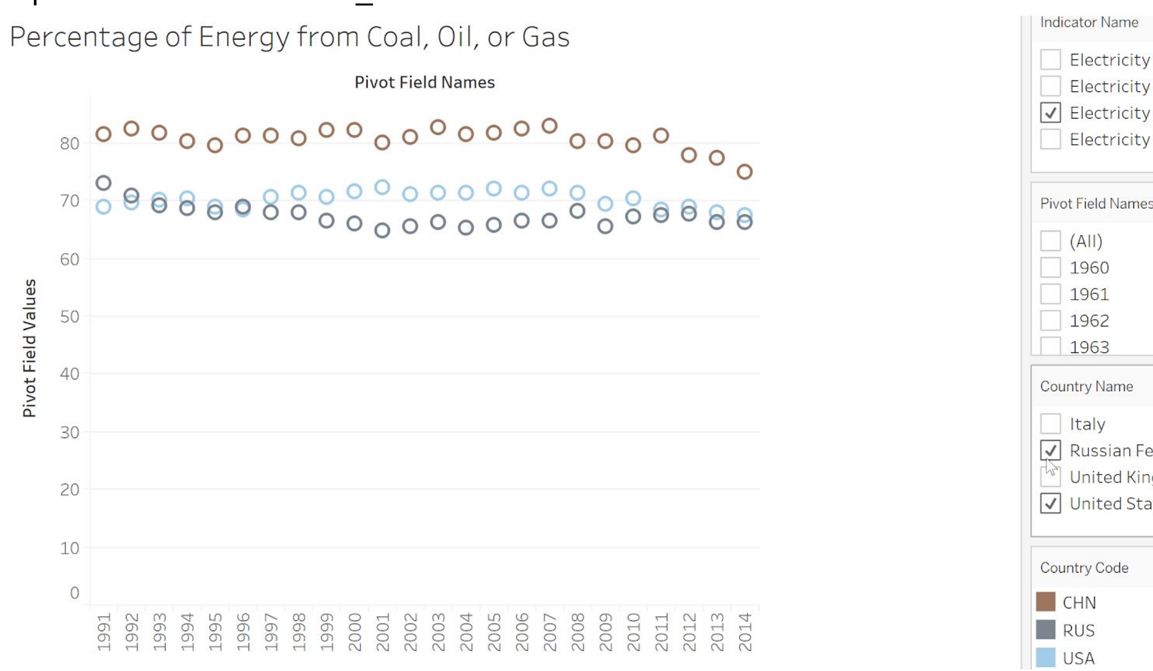
2. Small Countries

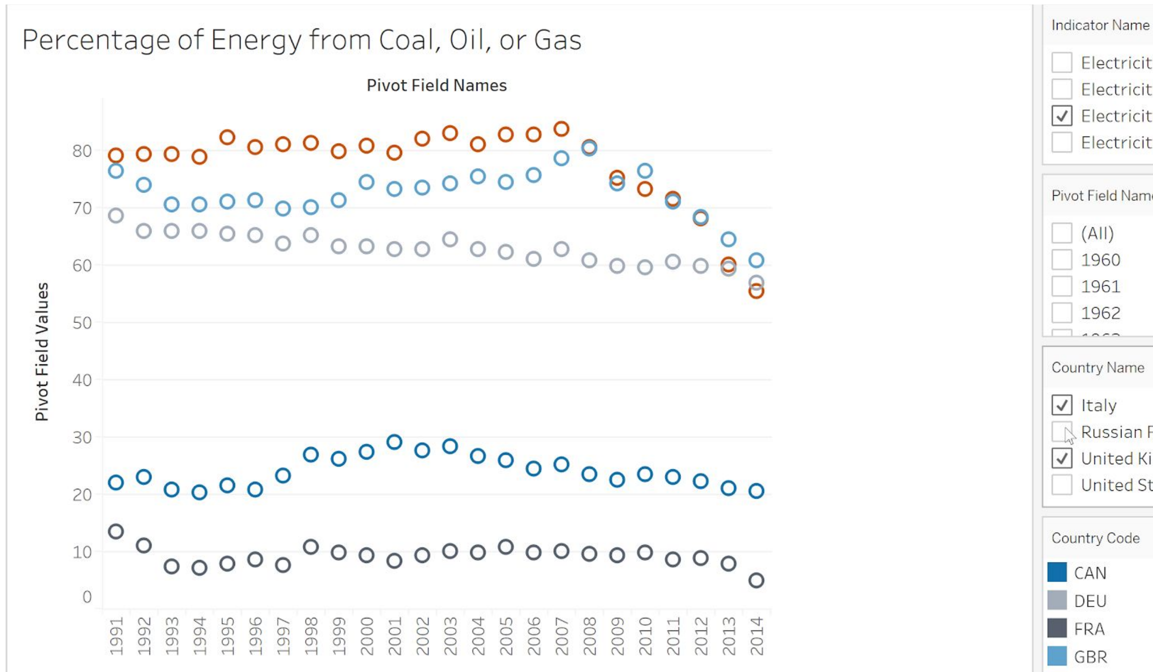


Similar to the United States, we see a trending decrease for all of our smaller countries, except for Canada, who seems to oscillate almost and is on a much higher level of emissions compared to the foreign countries.

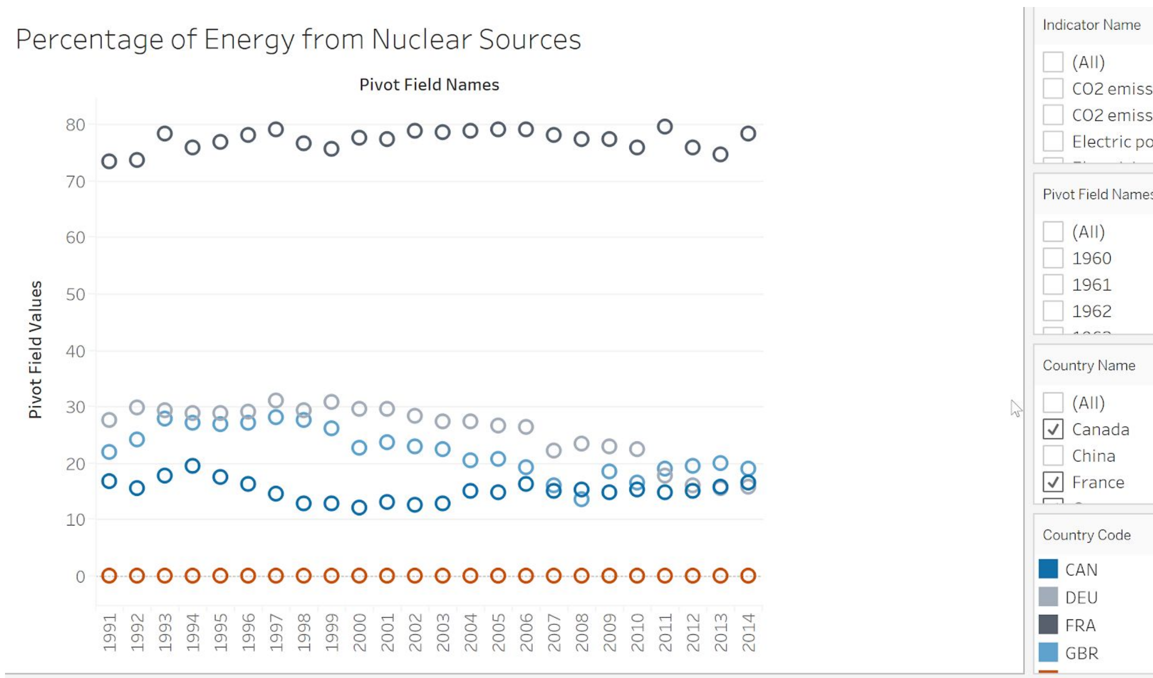


Looking at their energy production from renewable energy sources, we see great strides by countries such as Germany, Italy, and Great Britain. Excessively surpassing the larger countries efforts. Of course, this is easier done when they do not produce nearly as much energy as the larger countries. Note that the more modest growths in Canada and France are much more similar to that of the United States and China than their counterparts. We will lastly look at how much energy is produced by these countries by usage of oil, gas, and coal sources, as these will also explain a lot about their CO_2 emissions.

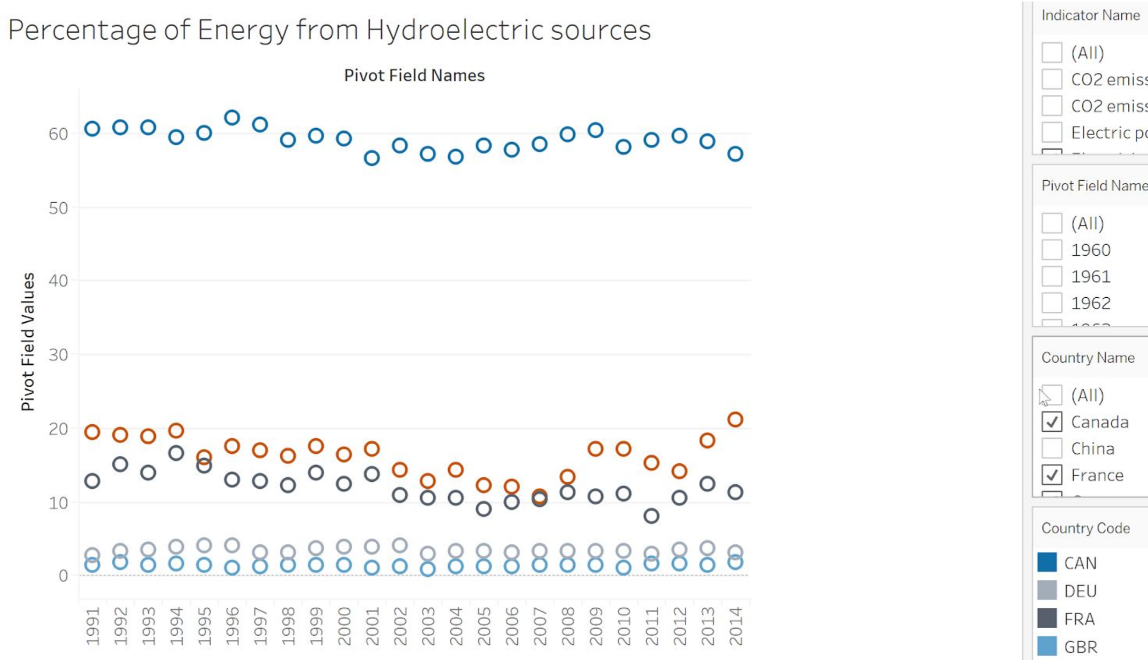




As we can see, the larger countries seem to be dominating the usage of fossil fuels for energy production. However, we do see some decreases, mainly by china. Meanwhile, the greatest strides in reducing their usage of oil, gas, or coal was achieved by Italy and the United Kingdom, reducing by ~ 20%! That is quite encouraging to see. Looking at the bottom, we see that Canada and France are still not accounting for a lot of their energy production when combined with renewable energy.



Here we can see that France gets most of its energy from nuclear sources which explains its low emission rates and its lower categories in fossil fuels. Meanwhile, Canada is only left with a large amount of hydroelectric energy.



3. Conclusion
   Overall, we see that by accounting for these different percentages of energy, we can account for trends in the CO_2 emissions in each of these countries both large and small. Our initial question was to explore mainly how renewable energy production impact these rates, but we can see that nuclear or hydroelectric energy is a large factor as well in some countries such as France or Canada for alternative sources.
   Similar to the last Instance, I attempted a linear regression on the CO_2 emissions per capita by this percentage.

```
Call:
lm(formula = co2PC ~ renewPct, data = energyresult)

Residuals:
    Min      1Q  Median      3Q     Max 
-9.7236 -3.9024 -0.8346  5.0834 11.7681 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 10.76584    0.31865  33.786   <2e-16 ***
renewPct    -0.16695    0.08096  -2.062   0.0399 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’

Residual standard error: 5.372 on 364 degrees of freedom
  (82 observations deleted due to missingness)
Multiple R-squared:  0.01155,	Adjusted R-squared:  0.008831
F-statistic: 4.252 on 1 and 364 DF,  p-value: 0.03991

> summary(newmodelEnergy)
```

However, our model seems to capture very little of the variation amongst ALL countries. Perhaps there are quite a few “outliers” in the sense of missing data, but I decided to simple add % of energy produced by fossil fuels.

```
Call:
lm(formula = co2PC ~ renewPct + foslPct, data = energyresult)

Residuals:
   Min     1Q Median     3Q    Max 
-9.451 -3.791 -0.658  4.692 12.088 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 11.59626    0.72578  15.978   <2e-16 ***
renewPct    -0.16701    0.08090  -2.064   0.0397 *  
foslPct     -0.01406    0.01104  -1.273   0.2038    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.367 on 363 degrees of freedom
  (82 observations deleted due to missingness)
Multiple R-squared:  0.01594,	Adjusted R-squared:  0.01052 
F-statistic:  2.94 on 2 and 363 DF,  p-value: 0.05412
```

However, this has contributed nothing to our model as it is not a significant predictor. This certainly warrants some further digging as our past visuals seem to indicate some success here. Overall, ***There is no***

clear relationship between renewable energy production and CO_2 emissions. We utilized the clean table format from R to indicate in a graphically apprehensive manner that there is no relationship overall, despite the patterns we easily understood from our previous graphics using Tableau.

4. Future work
   Same as the previous hypothesis, the data examined was not comprehensive. It would be best to similarly group by large and small countries by population and then examine this in more detail. Further, the results from our regression were not encouraging. Clearly there is much more to understanding CO_2 emissions than how we produce our energy.

V. Discussion

A. Summary

In both cases, we observed patterns in emissions of CO_2 into our atmosphere. This is a subject of extensive study, and understanding how it is introduced into our environment is crucial in our fight to limit the damage of global warming and climate change. We saw some patterns in both hypothesis, however, we found a strong linear relationship between land usage for agriculture and methane emissions, but not one for renewable energy production and CO_2 emissions. Overall, this was an illuminating experience as I went through the visual seeking mantra. From the large overview of all countries down to a select representative few with further filtering to particular categories. Many times we had to separate large countries from the visuals of small countries. This was due to the unscaled nature of some data and thus violated the principles of graphical excellence and integrity as we could not understand the impacts of smaller countries. Clearly, there is much more work to be done to truly understand how we impact our environment, but this was a valuable project to learn how to use several tools to analyze large datasets at various levels.

B. Tools utilized

Throughout this project, as seen in the visuals, I utilized the following:

1. R
2. Python
   a. Pandas
   b. Numpy
3. Tableau
4. WDI from http://data.worldbank.org/data-catalog/world-development-indicators.

References:

https://datacatalog.worldbank.org/dataset/world-development-indicators, for the data
https://stackoverflow.com, for many bugs and other issues