Authors: signior, jmu22

CS504 Report

Project Description

At the beginning of the semester our team was interested in determining the extent to which the establishment of power plants affected CO2 emissions of that country. We wanted to learn if there would we statistically significant effects on the increase of CO2 when a power plant was established in a specific country. We eventually decided to look at power plants and vehicles as factors that affect CO2 emissions.

in this project, we gathered the following datasets related to climate change:

- Power plants established by country
  - This dataset originated from the World Resources Institute
  - It gave us the list of all power plants established globally, the country established in, the type of fuel used, and the year it was established
- Carbon Emissions by country
  - This dataset originated from the World Bank
  - It gave us the CO2 emissions in Metric Tons per Capita from years 1960 to 2014
- Number vehicles by country
  - This dataset originated from the World Bank
  - It gave us the number of vehicles per 1000 people in a specific year.
  - Years were limited in this dataset.
- Population by country each year
  - This dataset originated from the World Bank
  - Gave us the population of all countries in each year.

We used the above datasets to create the following:

- Fossil Fuel Power Plants

  - Selected all years per an arbitrary country where a fossil fuel power plant was established
- Carbon Emissions Total
  - Multiplied country and year values with country and year values of population to get a total value of Carbon Emissions (Metric tons of CO2)
- Number Vehicles by Country
  - Used population to find out how groups of 1000 people can fit into population of that year
  - Multiplied that value by the value of vehicles that year to get the total number of vehicles.

In this project we used the following techniques and algorithms:

- Constraint Satisfaction
- Linear Regression
- Hypothesis Testing

Constraint Satisfaction

Our goal in employing constraint satisfaction as a technique was mainly to determine if our intuition about the effects of power plants and vehicles was correct. We assumed that if either factor had a significant effect on CO2 emissions for that country, we should expect to see that when we analyze the CO2 emissions before a power plant was established, and after a power plant was established. For example, if a fossil fuel based power plant was established in 2006, let the change in CO2 emissions from 2005 to 2006 be X1, and the change in CO2 emissions from 2006 to 2007 be X2. If there was some effect, we would expect X2 to be greater than X1. We wanted to guarantee that at least one of these years in which a power plant was established in a country had this property.

We did the same process for the vehicle data but less strict. We wanted to just ensure that a positive change in vehicles corresponded with a positive change in CO2 emissions. This property existed in most of the years, but there were a few countries that did not experience this property. More on that in the results section.

Linear Regression/Hypothesis Testing

To bolster our results from the constraint satisfaction analysis, we wanted to make sure that our model fit the data we had. We also wanted to do hypothesis testing, and actually figure out how significant our results are. We performed a linear regression analysis on both the number of power plants vs. CO2 emissions and number of vehicles per year vs CO2 emissions. In this portion our null hypothesis was that there is no relationship between the two variables. Our alternative hypothesis is that there is a linear relationship between the two variables.

Using Scipy.stats.linreg we calculated the slope, intercept, R-value, R-squared, and P-value of the two variable pairs. With an alpha value of 0.05, any P-value that was under 0.05 or close to 0.05 meant that we can accept our alternative hypothesis and reject the null. Any P-value greater than 0.05 meant that we cannot reject the null. More analysis of these results in the results section.

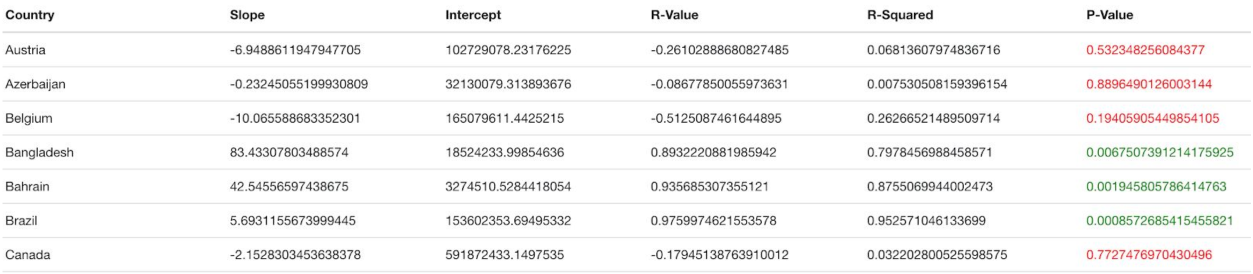
Technique Usability

In our case, the constraint satisfaction technique was a good gateway technique to making sure we were on the right track, but we weren’t super strict about finding all combinations of years that satisfied the constraint of CO2 percentage change. The other option was to do some sort of optimization but for our problem space the constraint satisfaction made more sense.

Linear regression with Scipy worked well because it gave us exactly what we needed to determine the degree of the relationship, and whether or not we should accept those results or not.

Authors: signior, jmu22

Results



Country	Slope	Intercept	R-Value	R-Squared	P-Value
Austria	-6.9488611947947705	102729078.23176225	-0.26102888680827485	0.06813607974836716	0.532348256084377
Azerbaijan	-0.23245055199930809	32130079.313893676	-0.08677850055973631	0.007530508159396154	0.8896490126003144
Belgium	-10.065588683352301	165079611.4425215	-0.5125087461644895	0.26266521489509714	0.19405905449854105
Bangladesh	83.43307803488574	18524233.99854636	0.8932220881985942	0.7978456988458571	0.0067507391214175925
Bahrain	42.54556597438675	3274510.5284418054	0.935685307355121	0.8755069944002473	0.001945805786414763
Brazil	5.6931155673999445	153602353.69495332	0.9759974621553578	0.952571046133699	0.0008572685415455821
Canada	-2.1528303453638378	591872433.1497535	-0.17945138763910012	0.032202800525598575	0.7727476970430496

The above screenshot is an excerpt of results

In our web application we rendered a table of the statistics generated from both the vehicle/CO2 test and the power plants/CO2 test. The slope and intercept help us determine the line of regression (best fit line). The main things we care about are the R-squared value and the P-value. The P-value is important because it determines whether or not we can reject the null hypothesis. The P-values that are less than our alpha (a = 0.05) we can reject the null hypothesis. This means that for those countries our model fits well. The R-squared value explains how much variation is explained by the model. We want higher values here. As you can see from the excerpt above, some countries have a P-value less than 0.05, others don't. For these results a few things could be the explanation. First, countries where we can't reject the null hypothesis tend to be well-developed/first-world countries. These countries have the resources necessary to pursue alternate forms of energy (wind, solar, hydroelectric, nuclear) that can reduce CO2 emissions or blunt the effect of vehicles and power plant establishments. These countries generally also have policies that either tax gas powered vehicles / fossil fuel based power plants which would reduce CO2 emissions. On the other side of the coin, the countries that we can accept the hypothesis tend to be countries that are underdeveloped/third/second world countries. These countries do not have the resources to pursue alternate forms of energy nor do they have policies that limit CO2 emissions. Some of these countries that end up not rejecting the null are due to lack or data.

Authors: signior, jmu22

Future Work

There is a lot of possibility with this project for future work, because we think it's interesting to consider different factors that lead to the ever increasing rate of CO2 emissions. We only considered power plants and vehicle data, but we could also consider various policies established by specific countries or policies established globally and determine the effects on CO2 emissions. We could also look into the historical rate of CFCs (Chlorofluorocarbons) and see if the decline of chemicals harmful to the ozone layer could affect CO2 emissions. Changes in forest acres by country or globally could also come into play, as trees are a vital organism that pulls CO2 from the air. Analysis of a variety of these factors could be helpful in determining which factor causes the greatest negative impact in CO2 emissions, and could help policy makers plan specific actions to counter those factors.

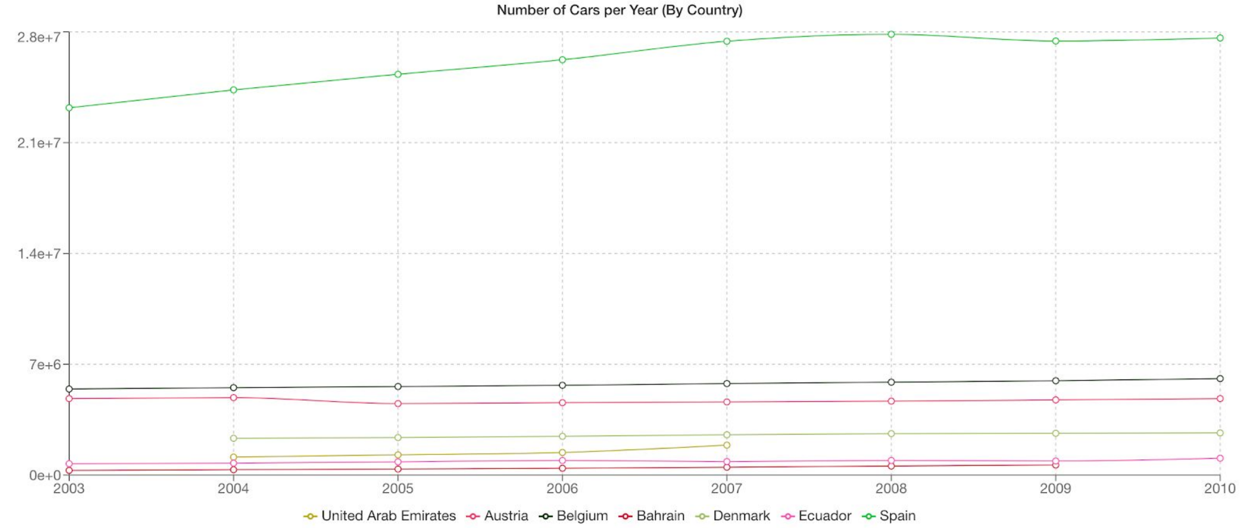
Web App description:

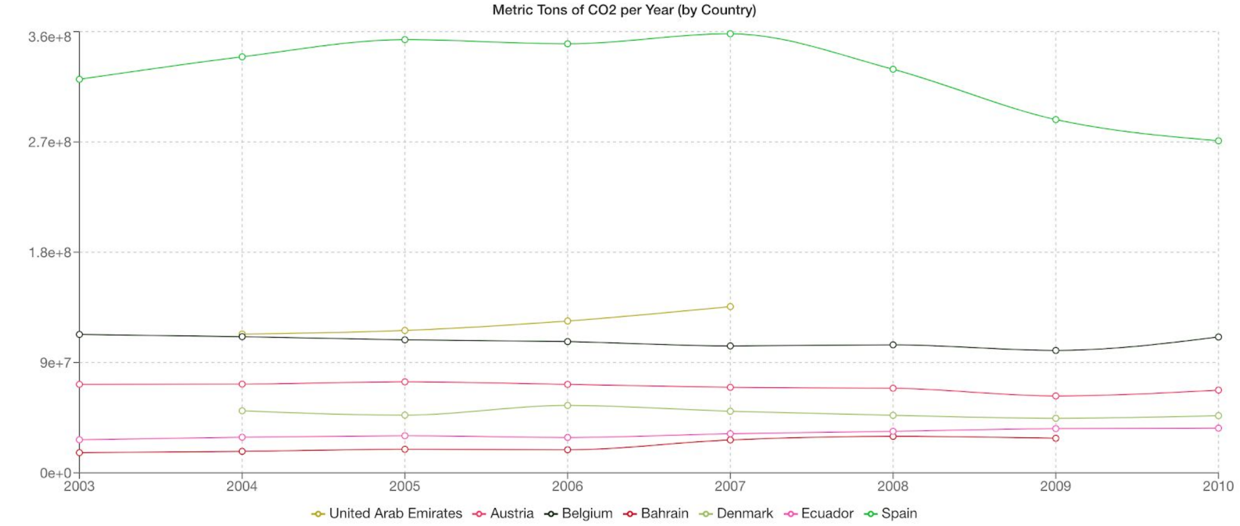
For the web application part of this project we decided to create a Flask Backend and a React Frontend. We wanted to plot the changes in the number of vehicles per year and the CO2 emissions per year by country, as well as the table of statistical data from our linear regression calculations on both the vehicle/CO2 relationship and the number of power plants/CO2 relationship.

For the backend we instantiated a simple Flask server that connects to our Mongodb repo and authenticates. We declared specific routes for the specific datasets we wanted to plot, and returned a json object pulled from the database.

On the front end we created a React application that has a home and visualizations page. The visualizations are a multi-line series of vehicles per country per year and CO2 emissions per country per year. A series of checkboxes on the left hand side allow the user to pick which countries to graph. These visualizations are built with recharts, which is a react charting library.

Authors: signior, jmu22





If you scroll to the bottom of the visualizations page you will see a table of the statistical analysis for the vehicles/CO2 emissions relationship, and the number of power plants/CO2 emissions relationship. The P-value is color coded based on the alpha value of 0.05. If the P-value is less than the alpha it is highlighted green; otherwise it is highlighted red. Green means we can reject the null hypothesis; Red means we cannot reject the null hypothesis.