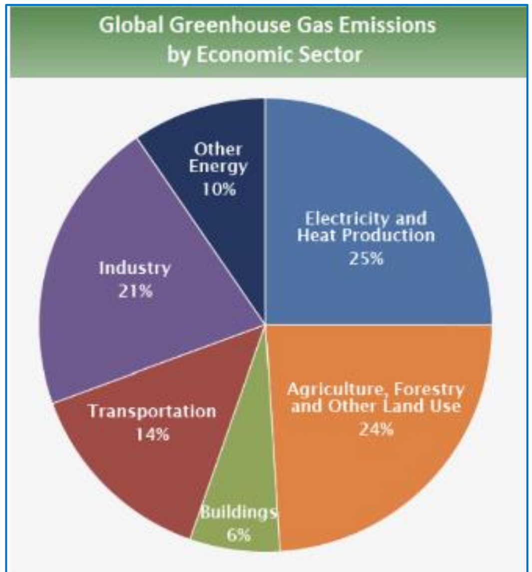
Senal Peiris, Shanel Peiris - November 2022

Economic factors which contribute to carbon dioxide Emissions in the United States & its influence on surrounding states

Summary. Classical economics assumes "Ceteris Paribus" holds in analysis & thus there is no impact of our neighbors on us & Vice Versa, however *Anselin* in his journal **Spatial Effects in Econometric Practice in Environmental and Resource Economics**(2001) used spatial econometrics to investigate spatial spillover in resources and environmental economics. Using these foundations, in our research we focus on analyzing whether, within a state, economic factors which drives emissions in carbon dioxide has a spillover effect on its surrounding neighbor states, using a Spatial Probit model.

Introduction

Traditional econometric analysis would look at how CO2 emissions in an individual state is affected by respective factors driving pollution and exclude any impacts by our surrounding neighbors on emission levels. However, we look to analyze CO2 emission at a state level by considering the concepts of spatial dependance. We segregate states being either a "high polluter" or a "low polluter" based on if the state is above or below the average US CO2 emissions. The latitudes and longitude values of the internal points of states are considered for our spatial analysis.



Source: IPCC (2014); Exit based on global emissions from 2010. Details about the sources included in these estimates can be found in the *Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.*

The motivation behind selecting the economic indicators influencing carbon dioxide emissions comes from trying to answer the question whether the level of CO2 emissions/pollution in a state/country as a result of factors within its own boundaries or is there a significant effect coming neighboring states. Information from the Environmental Protection Agency (EPA) confirms that the largest contributor to greenhouse gas emissions is Carbon Dioxide (CO2), hence we have focused our analysis on the CO2 emission levels. We can observe that the leading contributors to CO2 emissions are electricity/heat generation, agriculture, industry, and transportation. We choose variables which will help us to capture the effects of these 4 contributors on CO2 emission levels.

Dependent variable

- CO2 Emissions (2017) – Million Metric Tons

Independent variables

- Population growth (2016-2017)
- Number of firms per capita
- Log (2017 GDP)

Senal Peiris, Shanel Peiris - November 2022

Spatial dependance in carbon dioxide emissions by a state

Spatial dependance is a phenomenon which helps to capture the impact of neighbors and their spillover effect on variables of interest in an economic analysis. We hope to capture the spatial dependance using a Spatial Autoregressive (SAR) Model, where we try to model the impact of our immediate neighboring states and our higher order neighboring states on our own CO2 emissions to determine if it is classified as a high or low polluting state.

Expectation - We expect to see spatial dependance in CO2 emissions. We suspect that the economic variables/independent variables of one state to have a causal effect on CO2 emission levels of that particular state (marginal effect). We also expect there to be a feedback effect, the CO2 emissions of a neighbor increases because of the economic variables of this state, and as a result there will be a rebound effect which affect the CO2 emission level of the current focus state as well. Similarly, we expect there to be significant spillover (indirect) effects whereby the CO2 emissions levels of neighbors are likely to impact the emissions levels of the focus state of study.

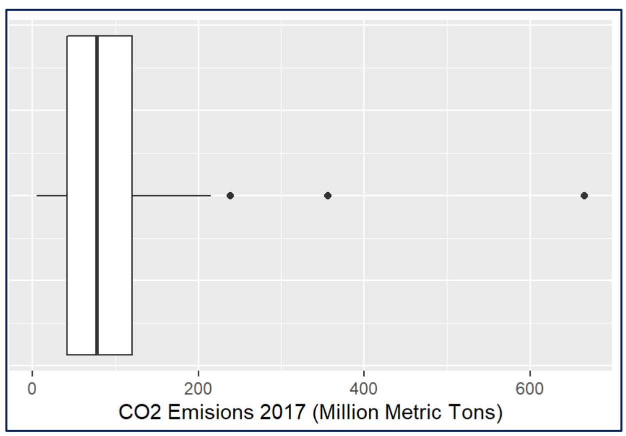
<u>Methodology – The SAR Probit Model</u>

$$Y^* = (In - \rho W)^{-1} X\beta + (In - \rho W)^{-1}\varepsilon$$

Where $Yi = 1 \text{ if } Y^* \geq 0$

$$Yi = 0 \text{ if } Y^* < 0$$

In our Probit model, we set up a new binary variable based on the CO2 emissions by a state, we classify a state as being a “high polluter” vs a “low polluter” based on whether the state’s annual emissions are above the US average of **103 million Metric tons**.



A latent variable is used to analyze our dependent variable of CO2 emissions by a state using Bayesian estimation techniques. Unlike traditional Probit models which does not look at the impact of Y^*, Bayesian estimation techniques uses “Data Augmentation” to predict the value of Y^* which is one of the sole focuses of interest often ignored in frequentist methods.

The SAR Probit model helps capture both the Direct and Indirect effects considering spatial dependance.

The direct effect captures both the **marginal effect** of the change in our independent variables as well as the **feedback effect** which comes from our neighboring states. The indirect effect captures the **spillover** coming from our neighboring states. The Regression results are shown in the next section.

<u>Regression results – Spatial Probit Model</u>

```
--------MCMC spatial autoregressive probit--------
Execution time  = 24.819 secs

N draws         =  15000, N omit (burn-in)=    500
N observations  =     50, K covariates    =      4
# of 0 Y values =     34, # of 1 Y values =     16
Min rho         = -1.000, Max rho         =  1.000
--------------------------------------------------

                       Estimate  Std. Dev   p-level t-value Pr(>|z|)
(Intercept)            -45.6002   14.4631    0.0000   -3.15   0.0027 **
population_growth     -189.9276  100.6608    0.0145   -1.89   0.0650 .
no_of_firms_per_capita -329.6806 134.7260    0.0018   -2.45   0.0180 *
log_gdp                  4.2962    1.3413    0.0000    3.20   0.0024 **
rho                      0.3638    0.1809    0.0353    2.01   0.0497 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

We can observe that our independent variables are statistically significant at the 5% significance level, except for population growth which is significant only at 10% significance level. As the Spatial Probit model is based on a latent variable model (Y*), we can only interpret the direction of the estimates, and hence are unable to directly interpret the estimations marginal effects without further calculations.

We observe that population growth and number of firms per capita have a negative relationship. The states which are classified as high polluters have high concentrations of populations and number of firms accounting for over 65% of both these metrics in the Unites States. This can be explained by the increasing trend of firms being more carbon neutral, especially in the last decade. We see firms taking more action to reduce their carbon footprint; hence this negative relationship/trend can be attributed to it, whereby we see lower CO2 emission levels correlating with higher number of firms.

Alternatively, we see that log GDP is positive indicating that as GDP within a state increase, we expect it to have a positive relationship whereby CO2 emissions is also expected to increase due to increased economic activity and hence it has a positive impact on the probability of success in being classified as a high polluter.

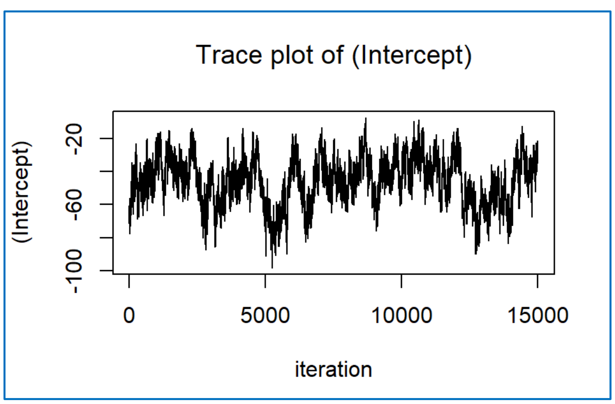
Importantly we also note that Rho is statistically significant, indicating that there is spatial dependance captured in this model. However, the value of our Spatial dependance is only 0.36 thus indicating that the strength of the positive spatial dependance is not too strong.

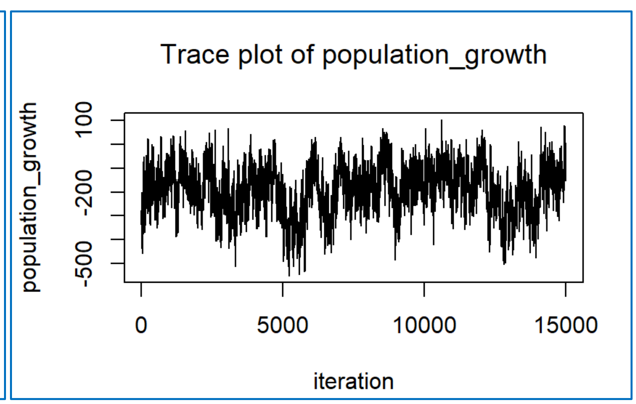
Direct, indirect and total effects – Spatial Probit Model

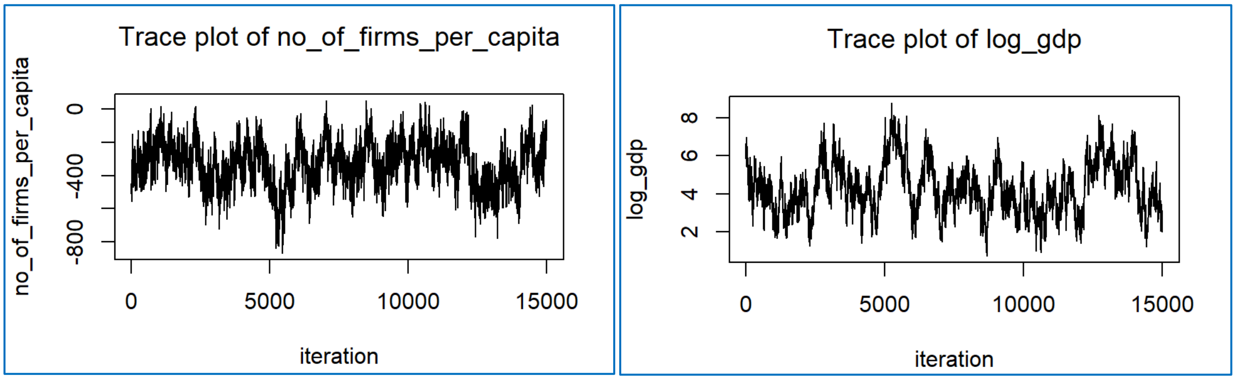
```
--------Marginal Effects--------

(a) Direct effects
                        lower_005 posterior_mean upper_095
population_growth         -30.549        -17.072     -4.36
no_of_firms_per_capita    -46.529        -29.705    -13.64
log_gdp                     0.296          0.387      0.51

(b) Indirect effects
                        lower_005 posterior_mean upper_095
population_growth        -25.3623        -9.0833      0.78
no_of_firms_per_capita   -32.5960       -14.5520      0.84
log_gdp                   -0.0102         0.2056      0.47

(c) Total effects
                        lower_005 posterior_mean upper_095
population_growth         -51.597        -26.155     -6.62
no_of_firms_per_capita    -65.819        -44.257    -21.30
log_gdp                     0.415          0.592      0.88
```

To interpret the effect our independent variables have on the probability of success in being classified as a “high polluter” we can observe the direct, indirect and total effects. We can observe from above that the direct effect is significant for all variables, and we can interpret the posterior mean to determine the impact each independent variable will have on the probability of being classified as a high polluter because of a marginal change in population growth, log GDP or Number of firms per capita. However, we note that the indirect effects are not significant, indicating that there is no spillover effect from neighbor states which would affect the probability of success of the focus state. However, as we noted in the regression results, Rho is significant, hence we deduce that the spatial dependance that is present, arises from the feedback effect which is reflected within our Direct effect table above.

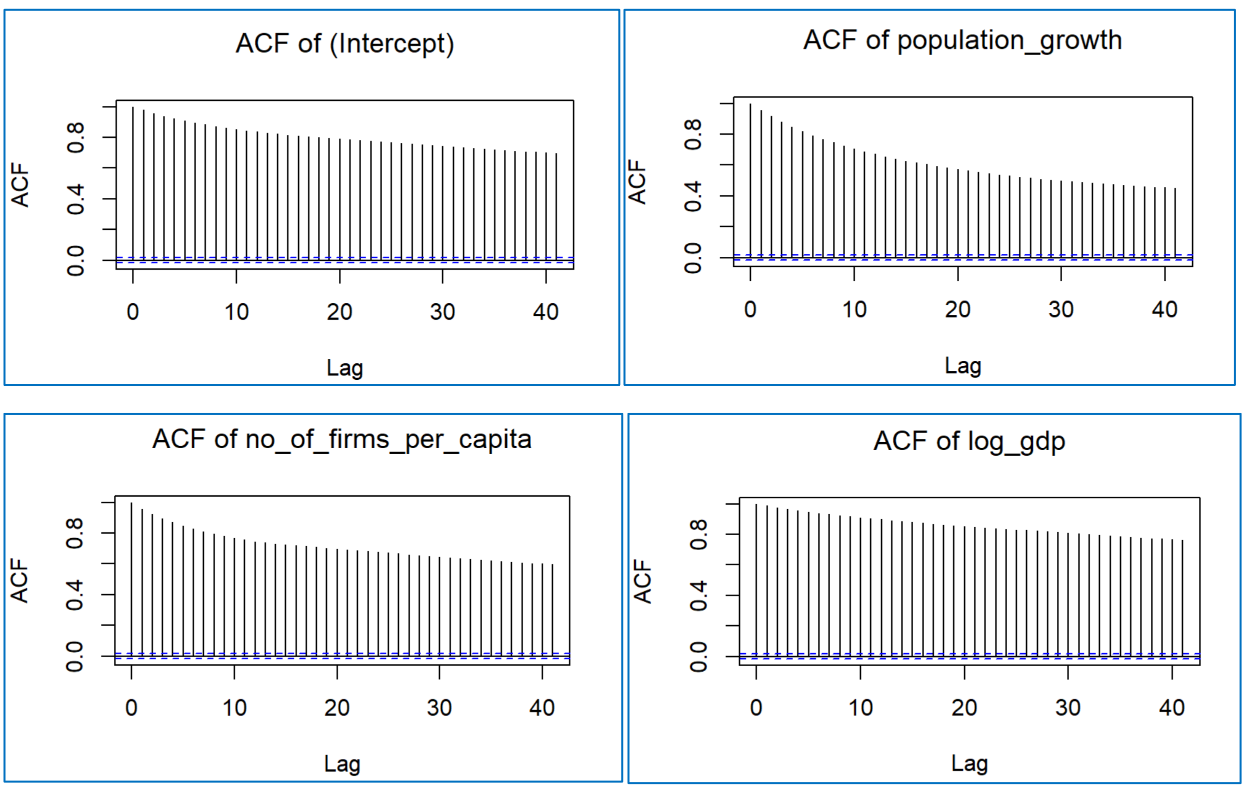
Convergence



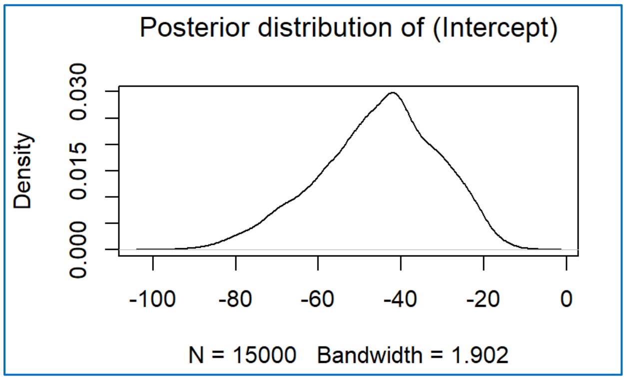


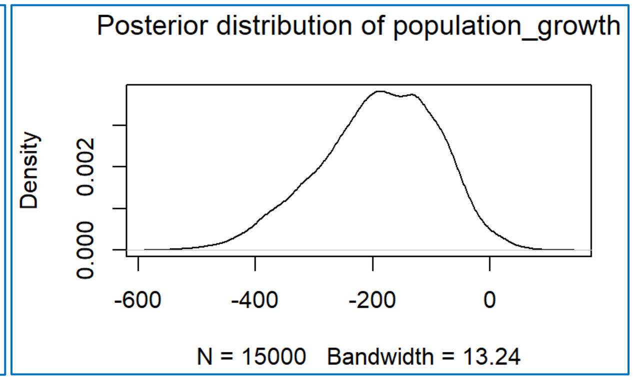


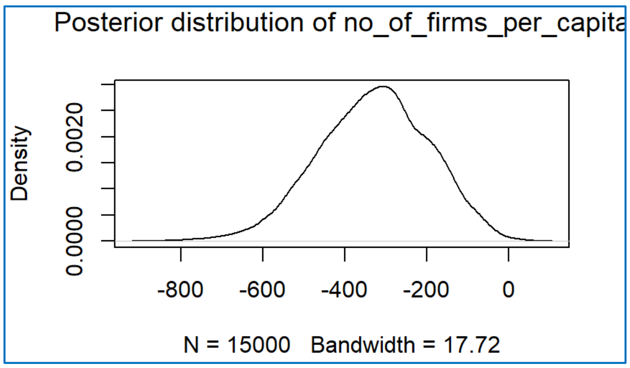
We have conducted our Spatial Probit analysis with 15,000 draws due to the limited data size to help provide more stable results and setting a burn in period of 500. We can observe that, the distribution converges as indicated in the graphs above, some variables show significant fluctuations between each individual draw while others indicate fewer volatile changes.

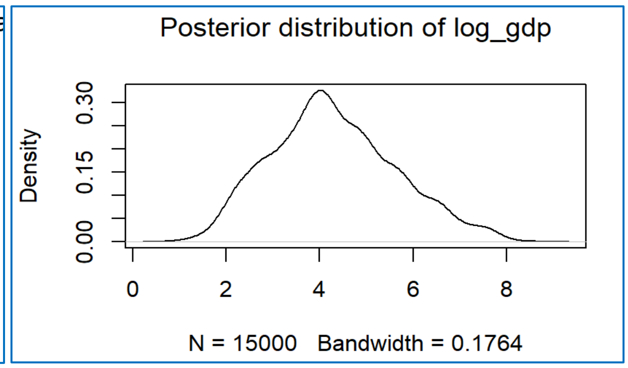


We can observe the posterior distributions of the independent variables below. Most variables indicate they follow a smooth distribution with centered means following a bell shaped curved close to a normal distribution

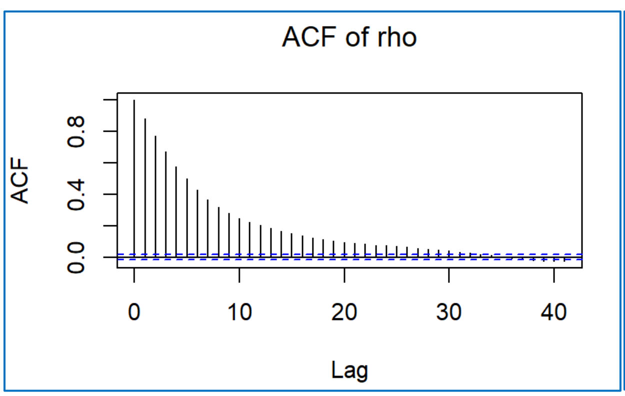


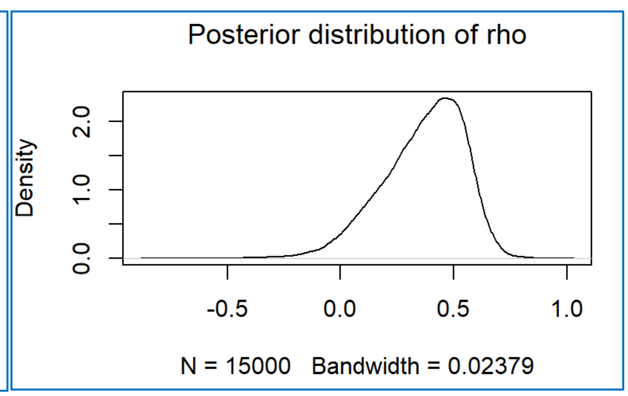






Rho – Spatial Dependance





As we noted from the regression output, Rho was statistically significant. This indicates that there is spatial dependence in the model and that CO2 level emissions of one state are affected by that of neighboring states and thus has an impact on being classified as being a high polluter vs a low polluter. However, a Rho value of 0.36 does not indicate a very strong spatial dependence effect. This is likely due to the analysis being done at a state level, and hence covers a wider geographical area. We might expect a higher Rho, as we focus on a smaller segment in geography, such as at a county level.

Non-spatial Probit Model

```
Call:
glm(formula = reg, family = binomial(link = "probit"), data = project_dataset)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0202  -0.1913   0.0000   0.0136   1.6118

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)             -39.61      14.13   -2.80   0.0050 **
population_growth      -173.64      85.93   -2.02   0.0433 *
no_of_firms_per_capita -410.53     160.49   -2.56   0.0105 *
log_gdp                   3.95       1.38    2.87   0.0041 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 62.687  on 49  degrees of freedom
Residual deviance: 21.110  on 46  degrees of freedom
AIC: 29.11

Number of Fisher Scoring iterations: 10
```

Looking at the basic Probit model, we can observe that the results are consistent with the analysis from our spatial model. We note that the relationship remains the same with population growth and number of firms per capita being negative on the dependent binary variable (Y) while log GDP is positive, and all variables are statistically significant. As we cannot read the table directly to interpret the effect the independent variables have on the probability of success in being classified as a high polluter, the marginal effect on the probability of success will be calculated using the Average Marginal Effect (AME) or the Partial Effect at the Average (PEA).

We compute the Partial Effect at the average (PEA) in the below table. As shown below, the marginal effects are statistically insignificant and hence, we are unable to comment further.

Partial effect at the Average (PEA)

```
Marginal Effects:
                          dF/dx Std. Err.     z P>|z|
population_growth       -0.8489    2.2912 -0.37  0.71
no_of_firms_per_capita -2.0069    5.3402 -0.38  0.71
log_gdp                  0.0193    0.0518  0.37  0.71
```

Conclusions

We used SAR to test for spatial dependence in our econometric model. We test and identify that our independent variables are significant and that there is a low spatial dependance arising primarily as a result of the feedback effect but there is no direct spillover effect. We understand a significant limitation in our study, which we intend to address through further detailed study, and this limitation is that the analysis is conducted at a state level, with the coordinates of the internal points of states taken for conducting spatial analysis. Due to the significant land mass size of a single state, we feel that a more focused study, possibly at a county level will yield better results on the causal effect the variables have on CO_2 emission levels, as well as the spatial dependencies.

Areas for further improvement

In future, we could look at improving upon this analysis by:

- Running a SEM model to capture spatial dependance in unobserved factors.
- Possible inclusion of prior information in Bayesian estimation to see the impact on the conditional posterior distributions.
- Conduct the analysis considering a timeseries data as well instead of a simpler cross-sectional data.
- Look at possible omitted variables which could help better explain what drives pollution.
- Delve deeper and expand the analysis to a county level. At present the state level analysis may not be able to represent the best results due to the small size of the data.

Appendix: Data Sources

- Summary statistics - https://www.epa.gov/ghgemissions/global-greenhouse-gas-emissions-data
- FRED Summary of US CO_2 emissions - https://fred.stlouisfed.org/series/EMISSCO2TOTVTTTOUSA
- Population data - https://www.census.gov/data/tables/time-series/dec/popchange-data-text.html
- Population by state by year - https://data.census.gov/table?q=population+by+state&tid=PEPPOP2019.PEPANNRES
- GDP Data - https://apps.bea.gov/regional/histdata/releases/0321gdpstate/index.cfm
- Agriculture data – https://quickstats.nass.usda.gov/#33338167-3CAB-3D8E-91B7-DE216C0A9529
- Energy generation data - https://www.eia.gov/electricity/state/archive/2017/
- All Sectors: Summary Statistics - https://data.census.gov/table?g=0100000US$0400000&tid=ECNBASIC2017.EC1700BASIC
- US States coordinates - https://www.census.gov/geographies/reference-files/2010/geo/state-area.html