# Data Analysis Report

*Boren Zheng*

*4/20/2019*

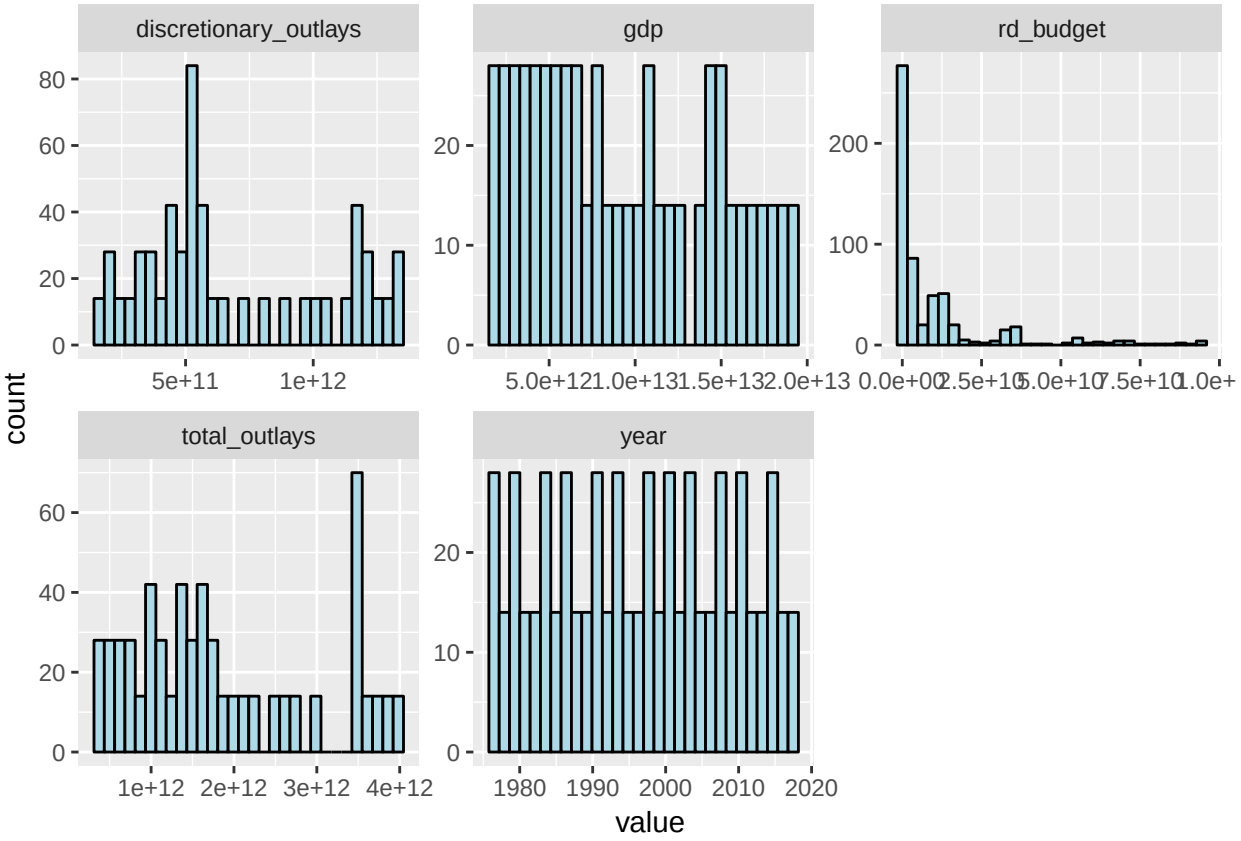
## Fed RD Spending Analysis

The "fed_r_d_spending" dataset contains 588 objects of 6 variables. The first 6 rows of the dataset are shown as following:

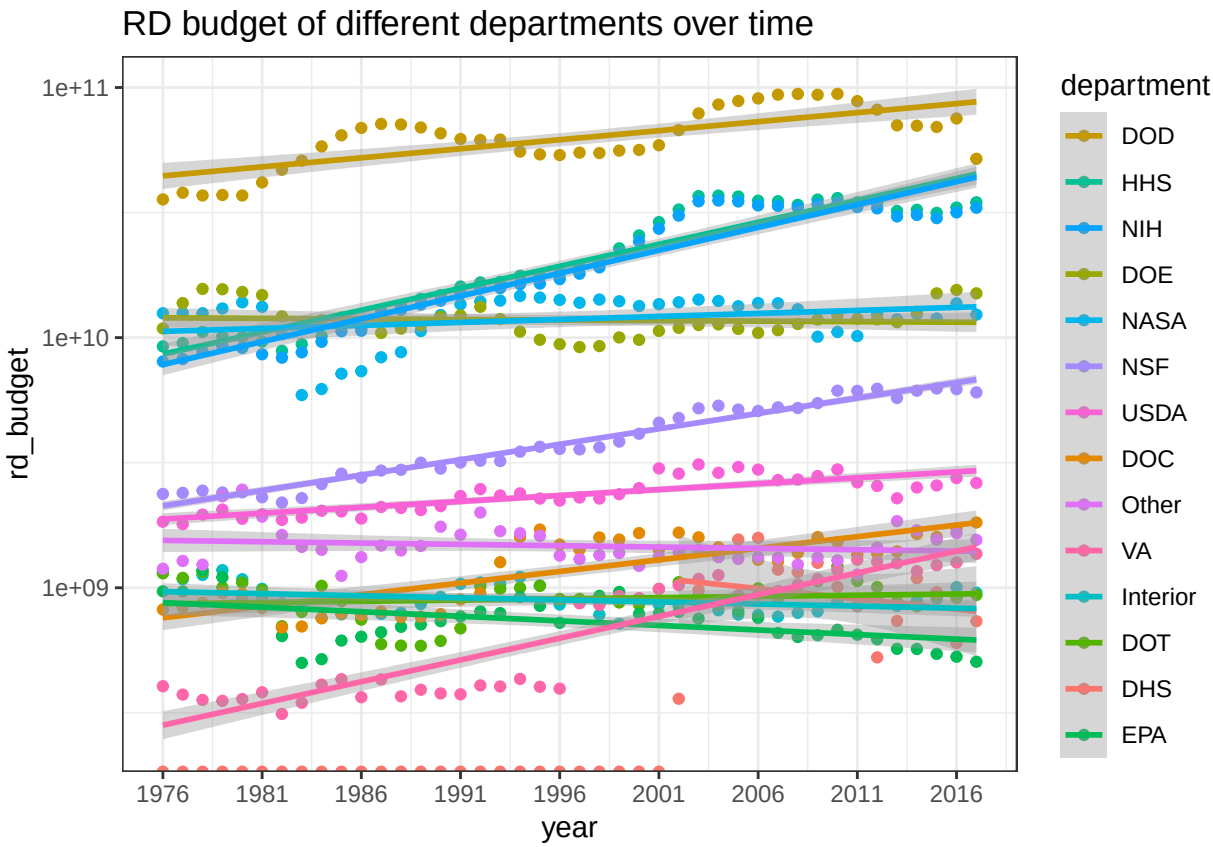
```
##   department year  rd_budget total_outlays discretionary_outlays      gdp
## 1        DOD 1976 3.5696e+10     3.718e+11             1.756e+11 1.79e+12
## 2       NASA 1976 1.2513e+10     3.718e+11             1.756e+11 1.79e+12
## 3        DOE 1976 1.0882e+10     3.718e+11             1.756e+11 1.79e+12
## 4        HHS 1976 9.2260e+09     3.718e+11             1.756e+11 1.79e+12
## 5        NIH 1976 8.0250e+09     3.718e+11             1.756e+11 1.79e+12
## 6        NSF 1976 2.3720e+09     3.718e+11             1.756e+11 1.79e+12
```

### Univariate analysis with plots

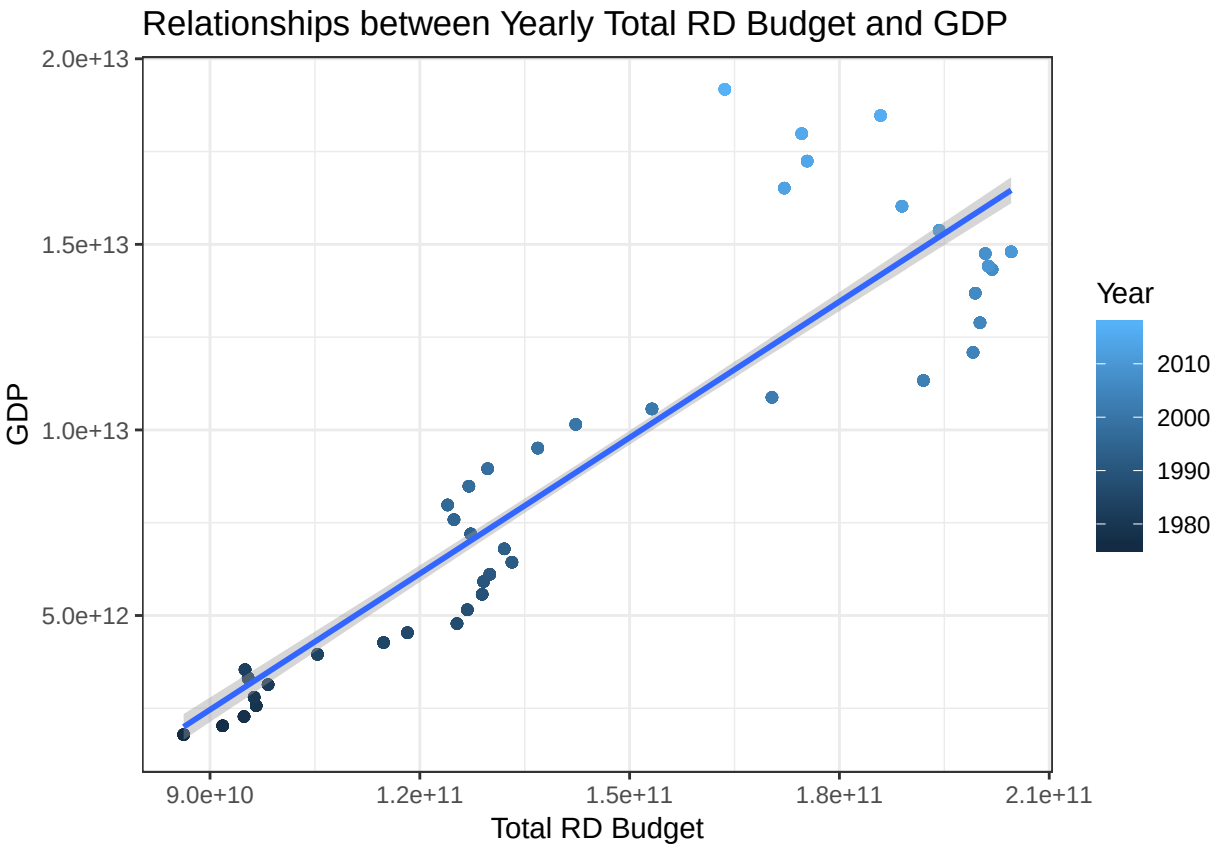
The histograms of each numeric variable are shown as following:



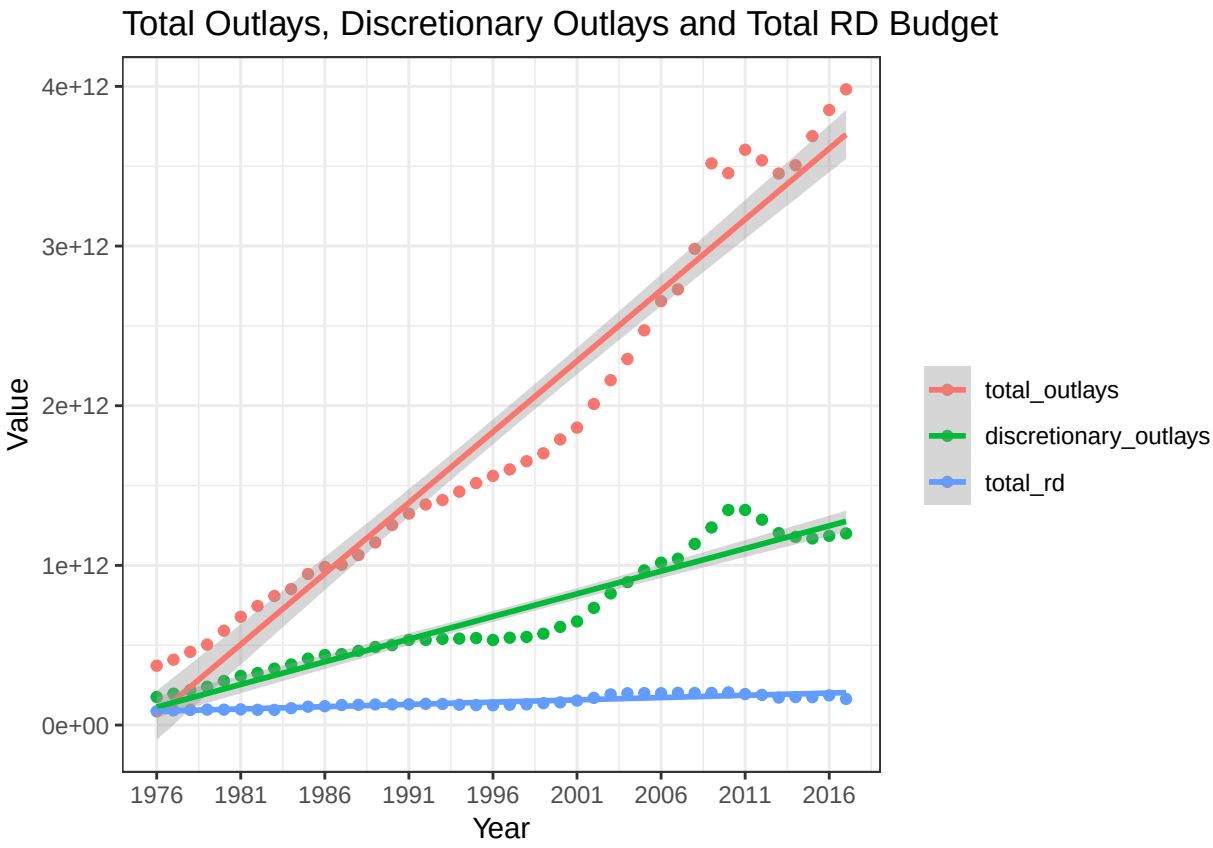
**Multivariate analysis with plots**




The plot above shows the relationships between RD budget and time based on different departments.We can see that RD budgets is independent of time because not all departments have increasing RD budgets over time.The RD budget of the Department of Defense are the highest all the time.




The polt above shows the relationships between yearly total RD budget and GDP. It shows that GDP has a positive correlation with total RD budget. And we could see that both GDP and total RD budget are increasing over time.



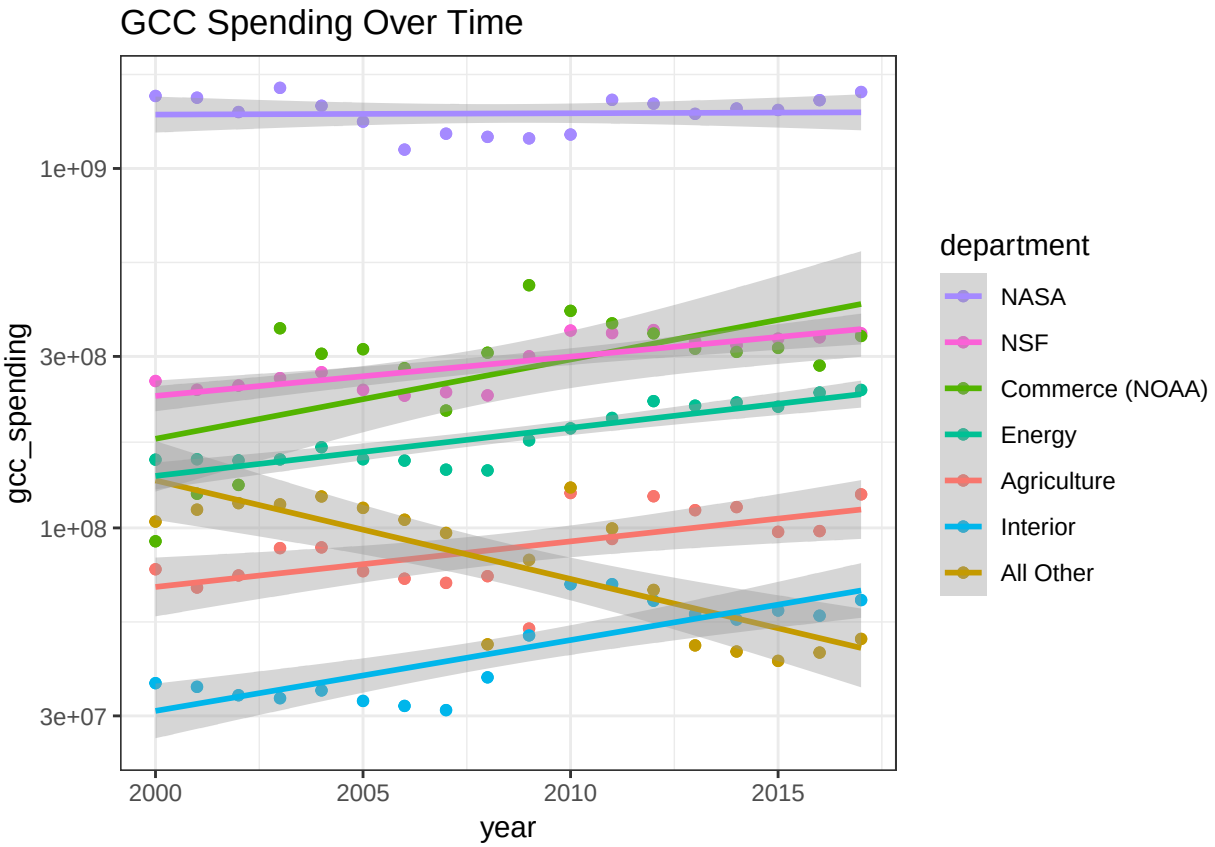

The plot above shows that both total outlays and discretionary outlays are increasing over time but total RD budget doesn't have significant change.

## Model: least square regression

Set gdp as the response variable, all predictors are in the model. The R-squared of this model is 0.9921604, p-value is 2.2e-16. ANOVA is used to test this models.

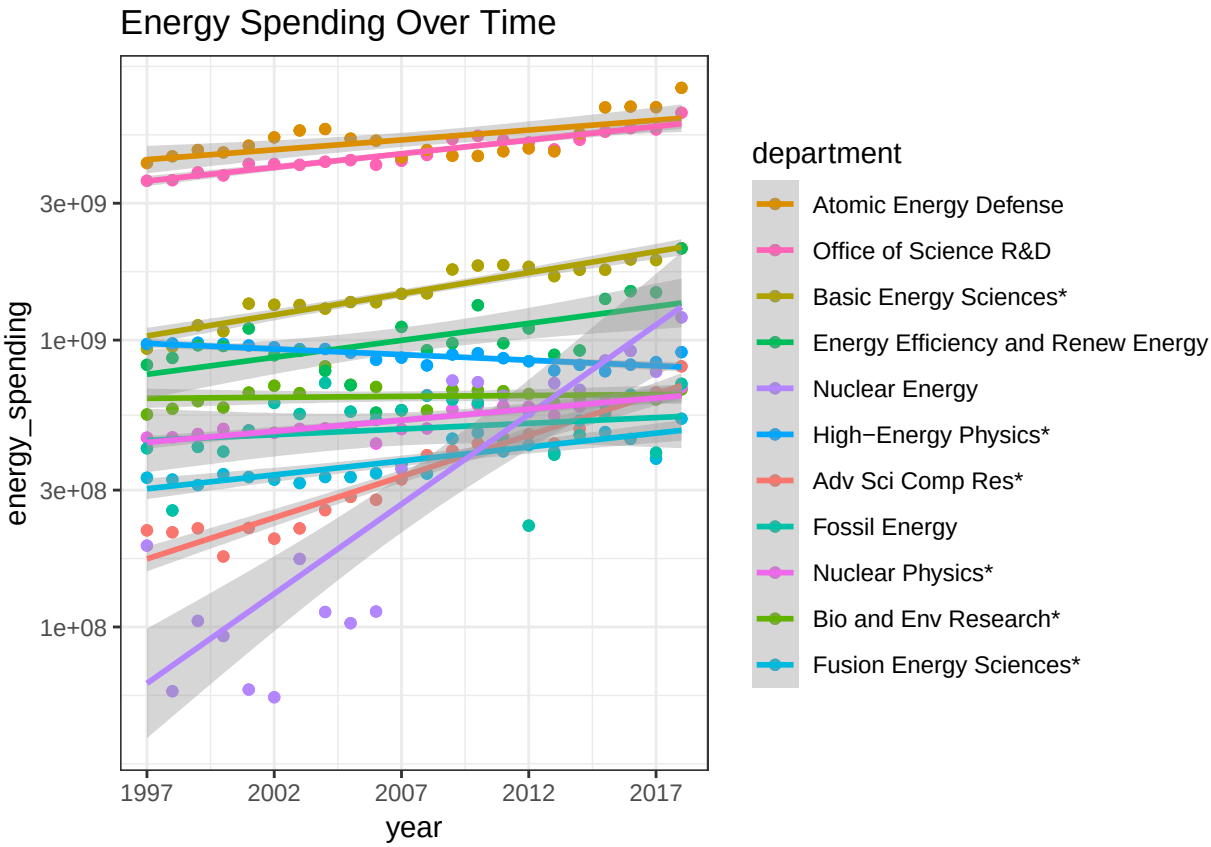
```
## Analysis of Variance Table
##
## Response: gdp
##                          Df     Sum Sq    Mean Sq    F value    Pr(>F)
## department               13 0.0000e+00 0.0000e+00     0.0000     1.000
## year                      1 1.5701e+28 1.5701e+28 71447.8735 < 2.2e-16 ***
## rd_budget                 1 2.8976e+20 2.8976e+20     0.0013     0.971
## total_outlays             1 1.3794e+26 1.3794e+26   627.7187 < 2.2e-16 ***
## discretionary_outlays     1 1.3686e+25 1.3686e+25    62.2803 1.533e-14 ***
## Residuals               570 1.2526e+26 2.1975e+23
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Climate Spending Analysis

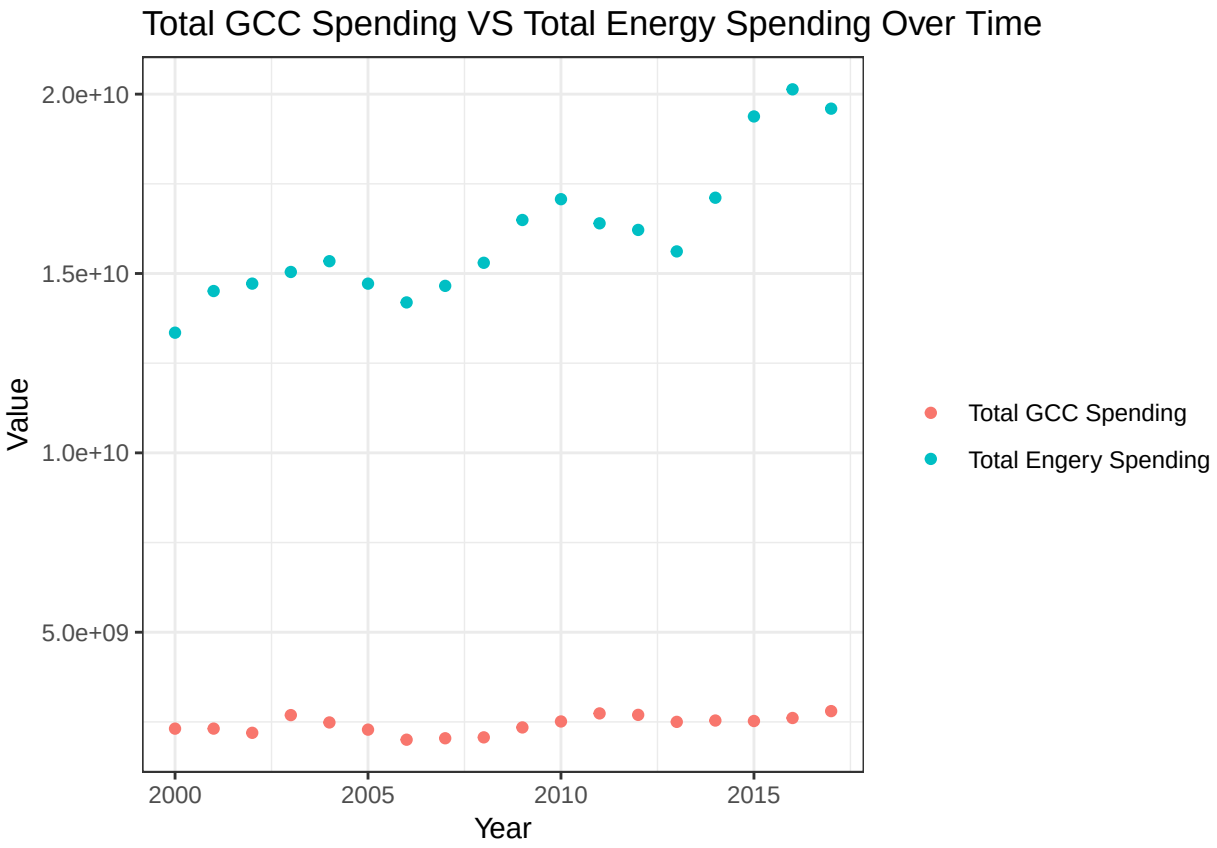



The plot above shows the relationships between gcc spending and time based on different departments. Most of the departments have increasing gcc spending over time.

## Energy Spending Analysis




The plot above shows the relationships between energy spending and time based on different departments. The department of Nuclear Energy has significant increasing energy spending over time.




The plot above shows that the total energy spending is much higher than total GCC spending. Also, the total energy spending is increasing over time.