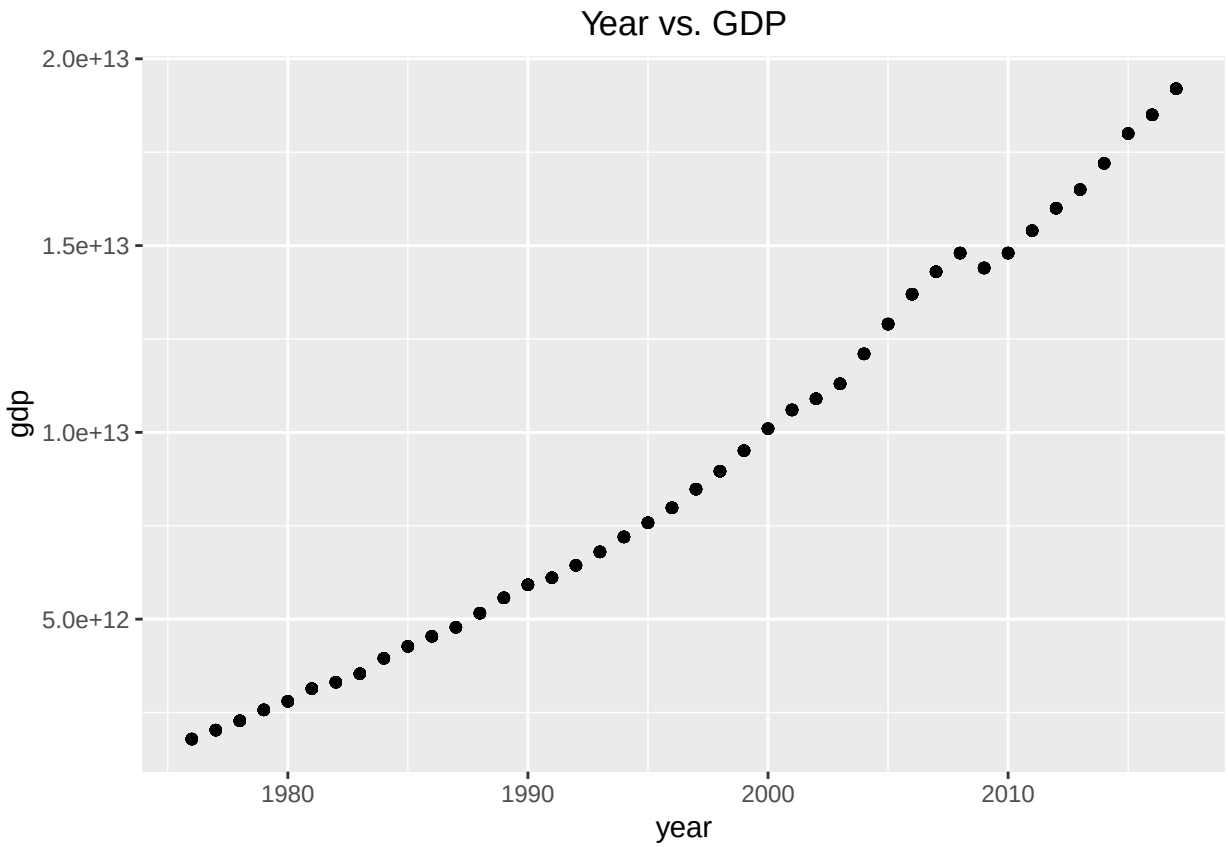
Homework #3

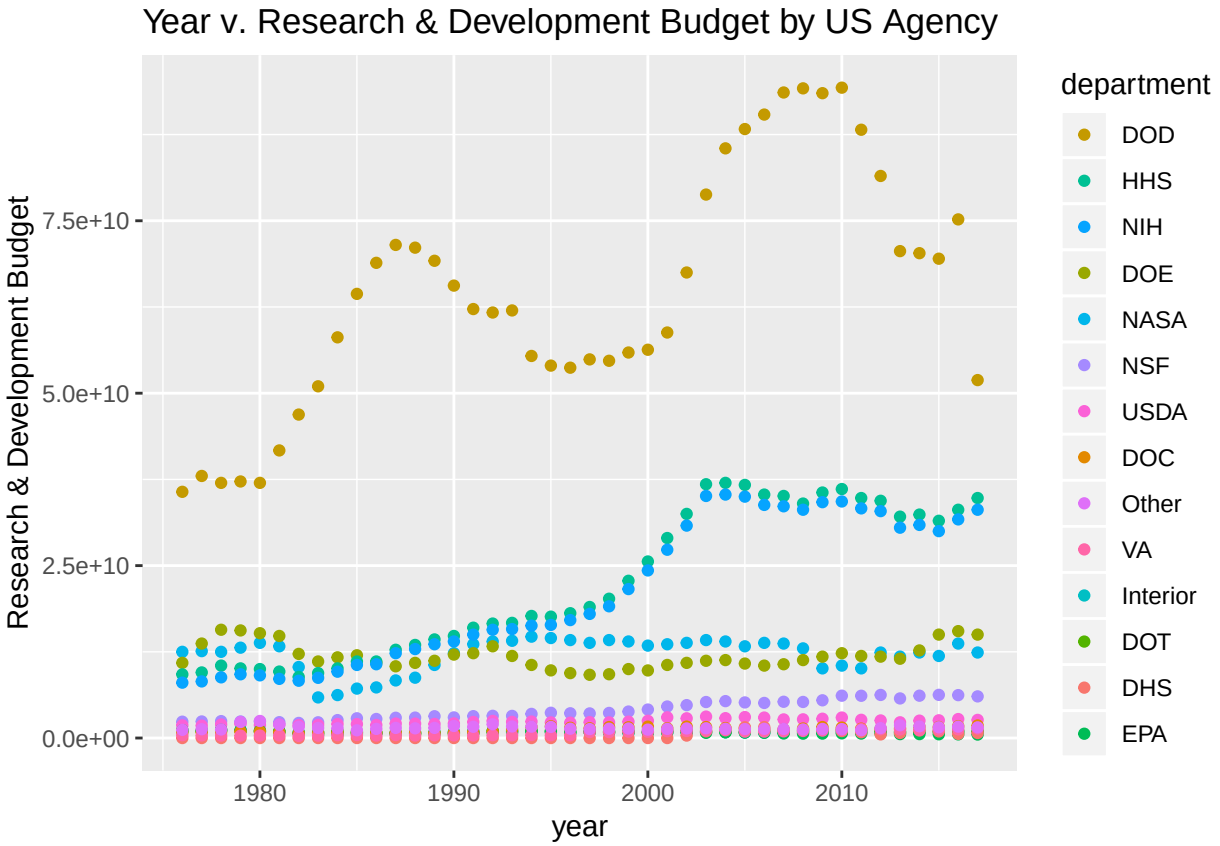
Julie Osborne

4/21/2019

Fed_r_d_spending Dataset

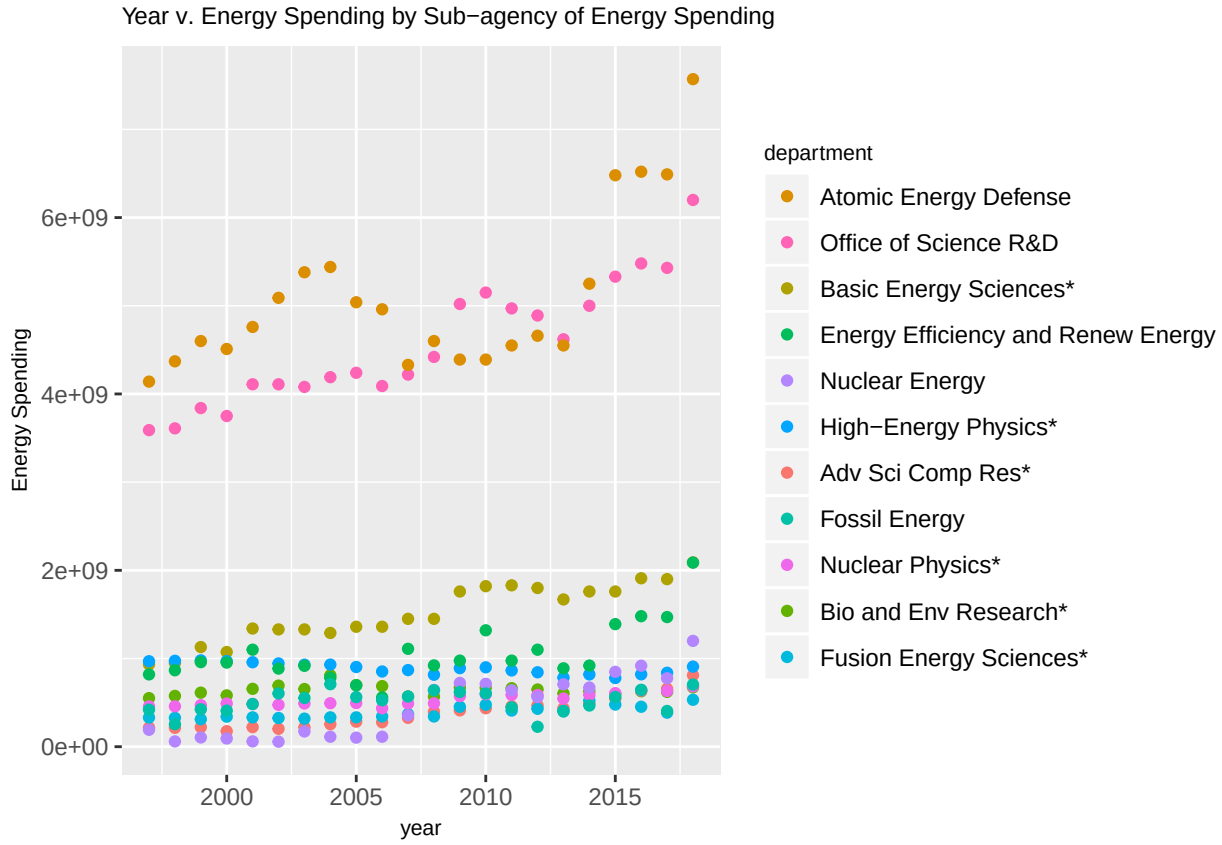


Over the course of 40+ years, there is a steady growth in GDP. There is a drop in the GDP between 2008 and 2009, which correlates to the 2008 financial crisis that occurred in the United States.



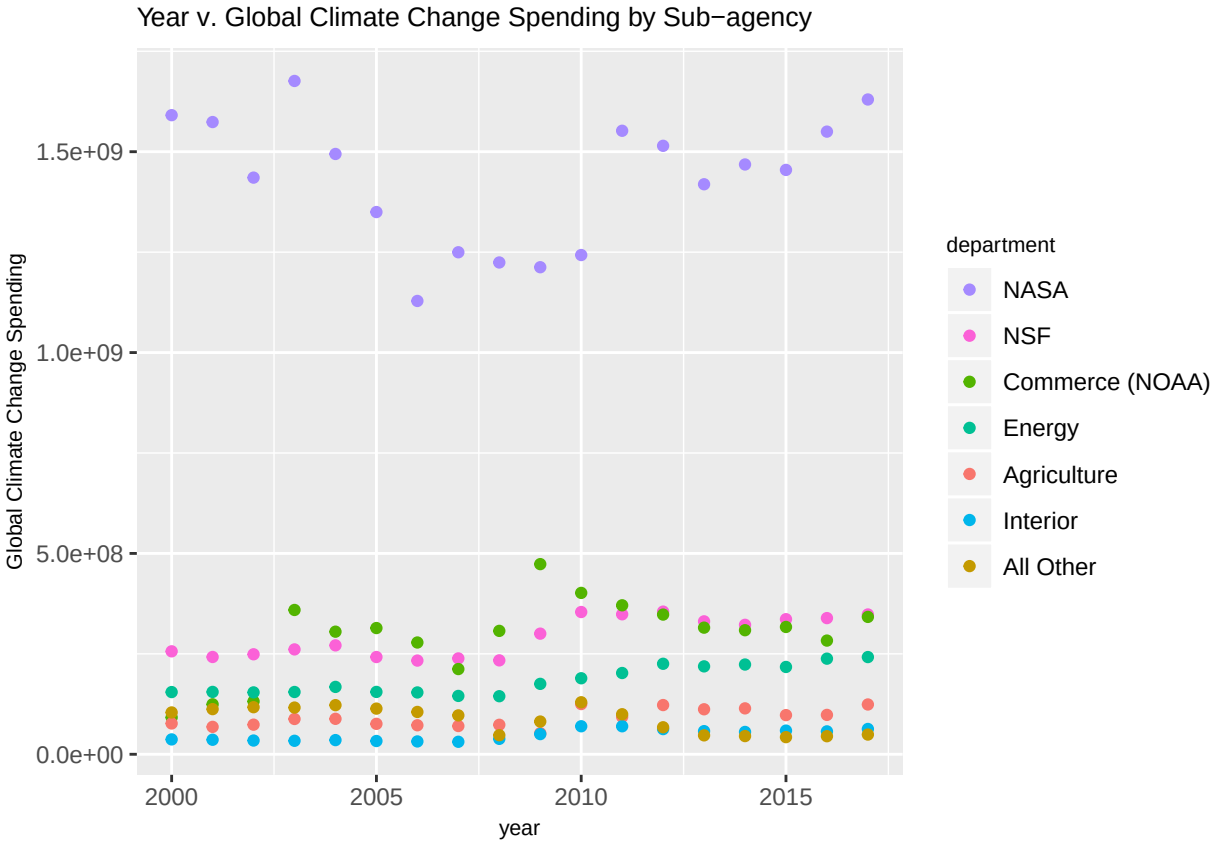
This graph shows how much research and development budget each US agency recieved over the course of 40+ years. Over this time period, the Department of Defense (DOD) recieved far more budget compared to all other departments. There was a large spike in budget that the DOD recieved between 2001 and 2005 that can correlate to 9/11, the War in Afghanistan, and the Iraq War, where the US was increase its defense. From approximately, 1990 onward, the Department of Health and Human Services and the National Institute of Health recieved the second and third highest amount of research and development budget. This could be related to the AIDS epidemic and increase in R&D being allocated to find a treatment/cure for it.

Energy_spending Dataset



This graph shows how much energy spending each US sub-agency of Energy Spending had over the course of 20 years. The two sub-agencies that have the the highest energy spending over the span from 1997-2018, is both Atomic Energy Defense and Office of Science R&D. There is a very large gap between these two sub-agencies in comparison to all the other sub-agencies.

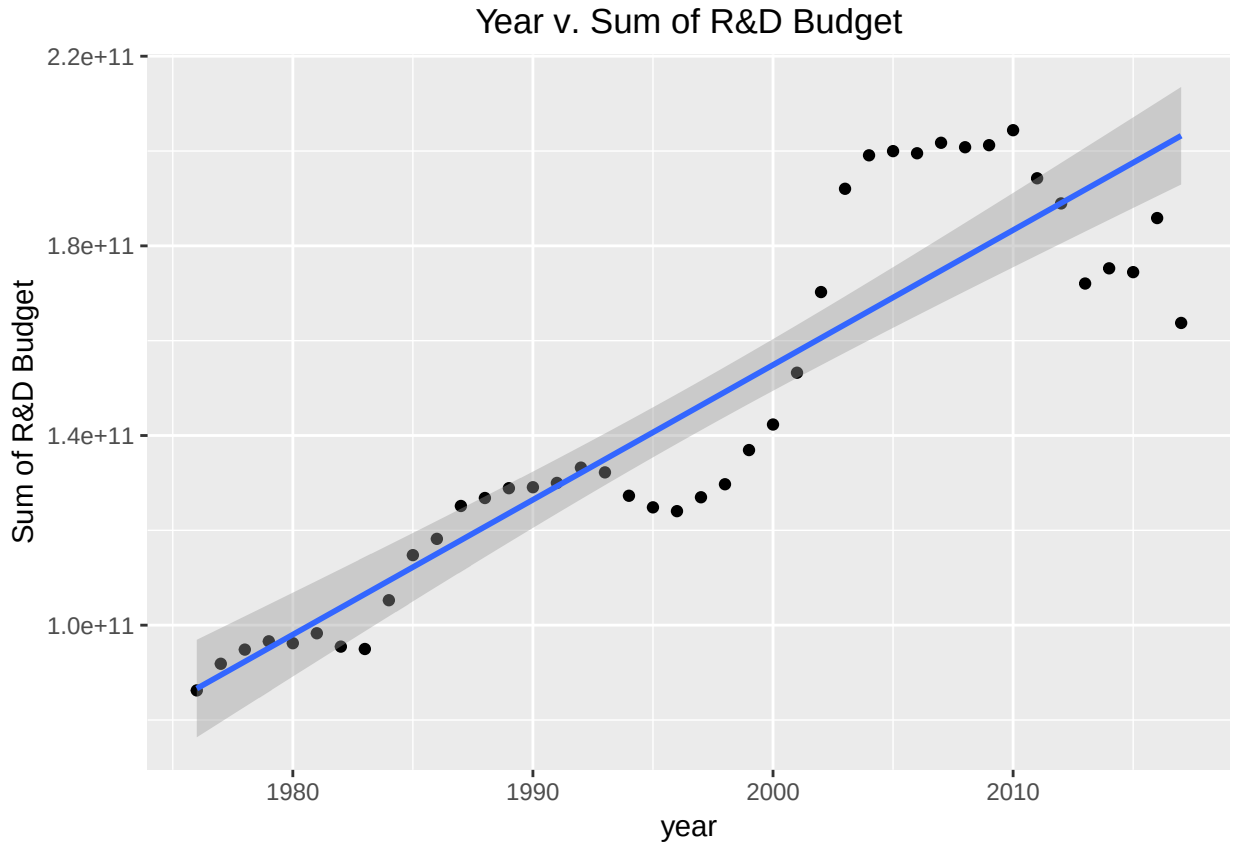
Climate_spending Dataset

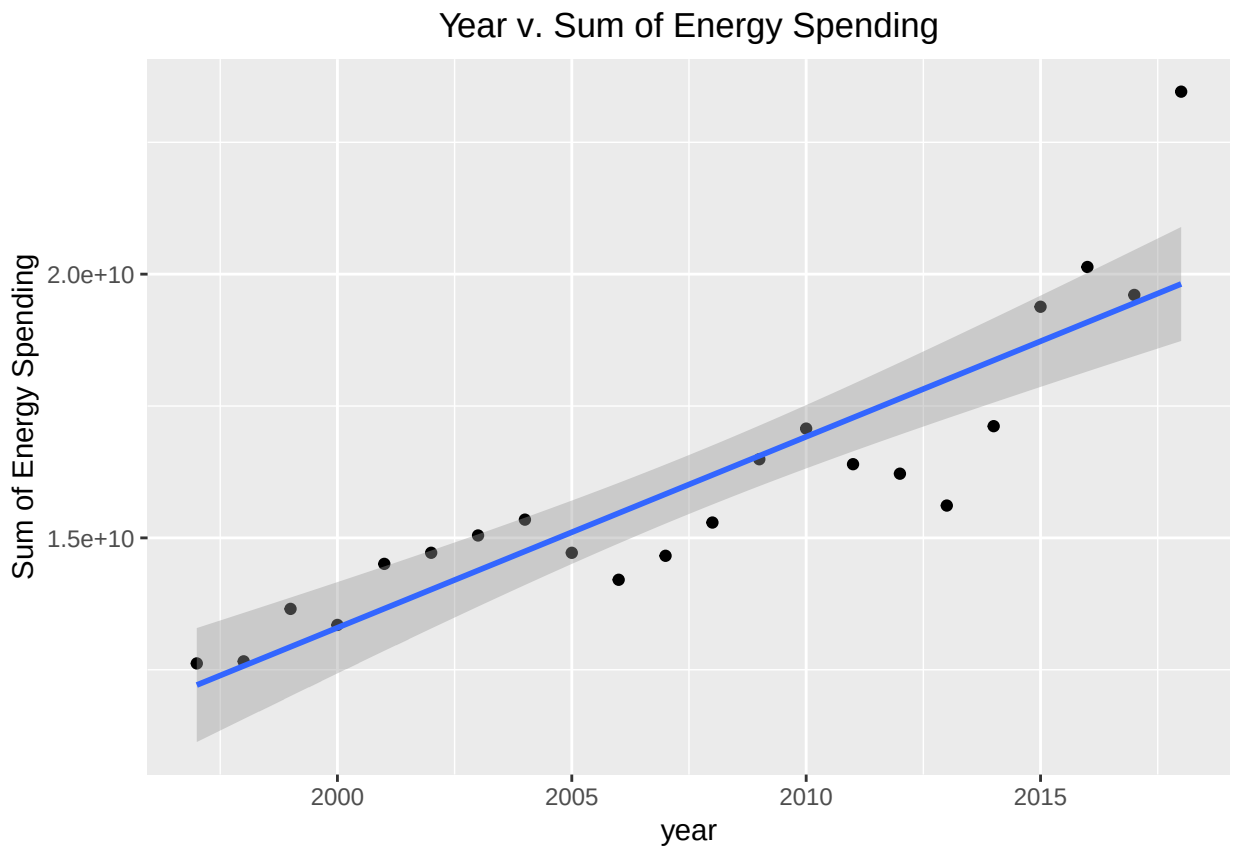


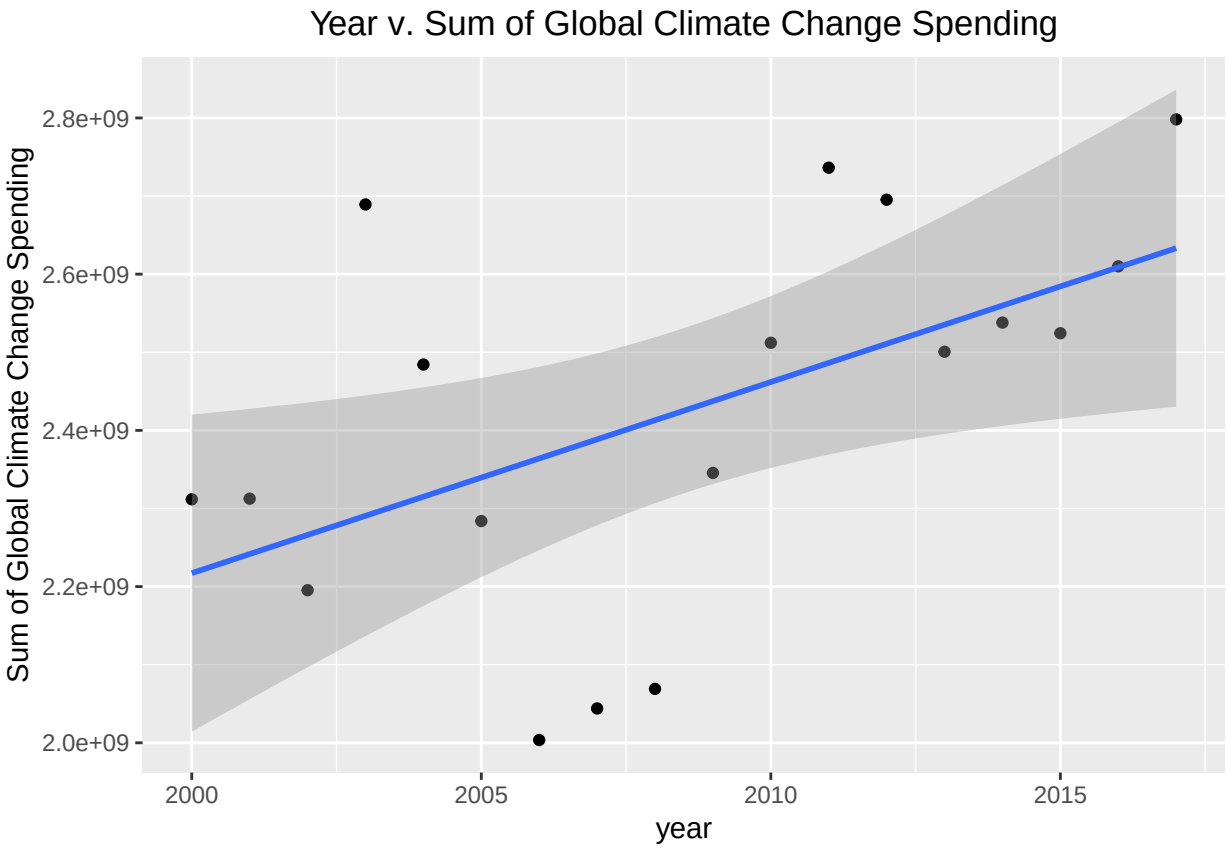
This graph shows how much global climate change spending each US sub-agency of Global Climate Change spending had since 2000. Over this time period, NASA had the highest amount of global climate change spending. A possible reason that NASA has the highest amount is because with the number of satellites it possess, it would allow them to be a leading force in monitoring and contributing to the field of global climate.

Statistical Analysis

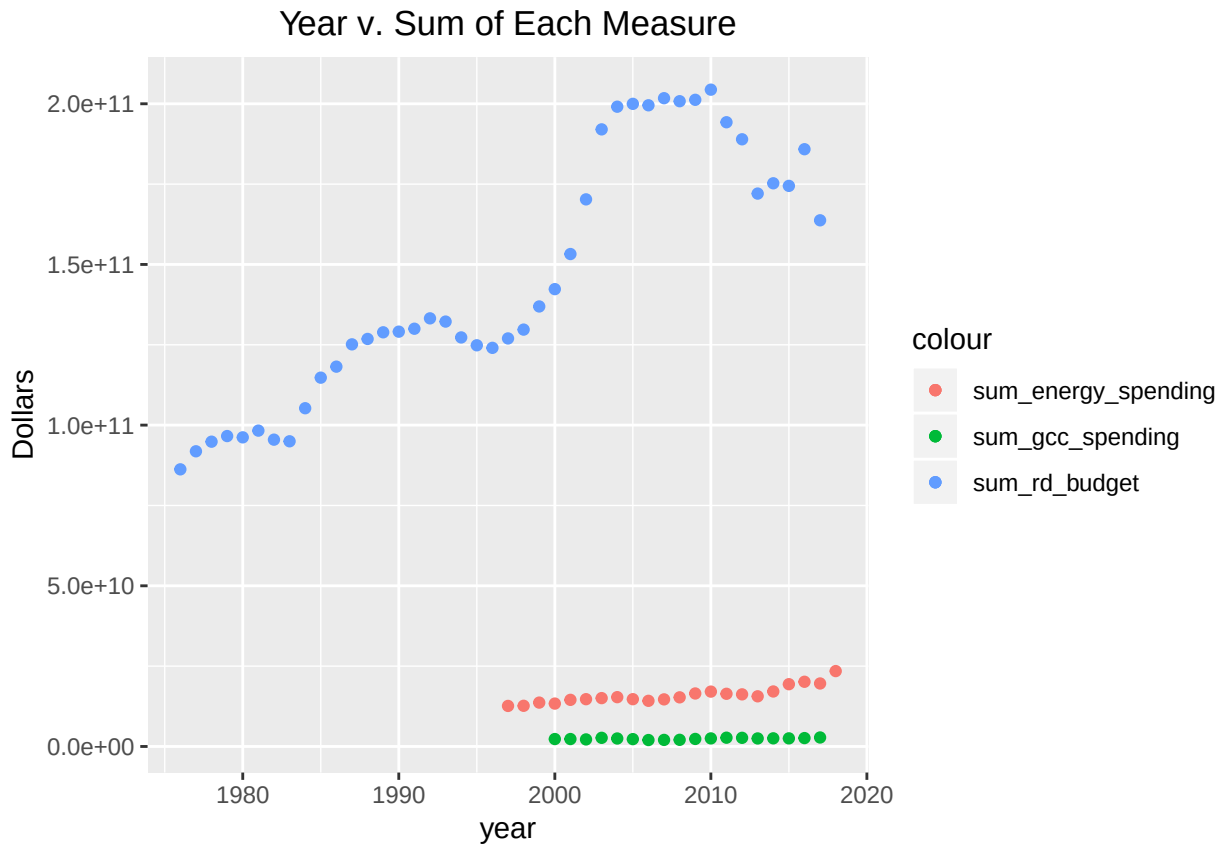
After looking at the how R&D budgets, energy spending, and global climate change spending varies over time for various departments, the overall sum of budget/spending across all departments will be analyzed to see how it changes over time.




Year v. Sum of Energy Spending
Sum of Energy Spending
2.0e+10
1.5e+10
2000
2005
2010
2015
year



For all three measures, R&D budget, energy spending, and global climate change spending, the trendlines indicate an increase in budget/spending over time.



This graph plots the sum of R&D budget, energy spending, and global climate change spending over time to see how they compare to each other. All three lines start at a different year, but this correlates to the start year of the corresponding dataset. Of the three measures, the sum of R&D budget far outweighs both the sum of energy spending and the sum of global climate change spending.

The sum of all three measures, will be used in a linear model to see if they are good predictors in measuring GDP.

```
##
## Call:
## lm(formula = mean_gdp ~ sum_rd_budget + sum_energy_spending +
##     sum_gcc_spending, data = cc_rd_energy_gcc_gdp)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -1.772e+12 -1.159e+12 -1.893e+11  9.988e+11  2.923e+12
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -7.101e+12  5.654e+12  -1.256    0.230
## sum_rd_budget        1.533e+01  2.022e+01   0.758    0.461
## sum_energy_spending  1.311e+03  2.426e+02   5.404  9.3e-05 ***
## sum_gcc_spending    -9.703e+02  1.956e+03  -0.496    0.628
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.508e+12 on 14 degrees of freedom
## Multiple R-squared:  0.7581, Adjusted R-squared:  0.7063
## F-statistic: 14.63 on 3 and 14 DF,  p-value: 0.0001351
```

In this linear model, the only predictor that is statistically significant (at the 5% level) in predicting GDP is sum_energy_spending which has a p-value of 9.296e-05 which is less than 0.05. Also, the R^2 for this model is 0.7063086 which means 70.63% of variation is explained by the model.

A stepwise regression was used to ensure that using these predictors that the model with only eneregy_spending as the predictor was the best model.

```
##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select

##
## Call:
## lm(formula = mean_gdp ~ sum_energy_spending, data = cc_rd_energy_gcc_gdp)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -1.909e+12 -9.752e+11 -2.025e+11  1.363e+12  2.624e+12
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -5.431e+12  2.968e+12  -1.830   0.0859 .
## sum_energy_spending  1.237e+03  1.830e+02   6.756 4.61e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.461e+12 on 16 degrees of freedom
## Multiple R-squared:  0.7404, Adjusted R-squared:  0.7242
## F-statistic: 45.64 on 1 and 16 DF,  p-value: 4.614e-06
```

The stepwise regression confirms that using these predictors, the model with only energy_spending is the best model. Energy_spending is statistically significant with a p-value of 4.6139e-06 and an $R^2 = 0.724223$.