

# Final Project

David Robinson, Aditi Jackson, and Jonathan Joyner

December 6th, 2023

# Contents

Rationale and Research Questions	4
Dataset Information	4
Wrangling eGRID	6
Wrangling Gov Data	8
Exploratory Analysis	10
Analysis	15
Geospatial Analysis	22
Summary and Conclusions	25

## List of Tables

1	Data Structure Summary - eGRID . . . . .	4
2	Data Structure Summary - Governors . . . . .	5

```
# setting Global options for code chunks
knitr::opts_chunk$set
```

```
## function (...)
## {
##   set2(resolve(...))
## }
## <bytecode: 0x00000250fa8a2040>
## <environment: 0x00000250fa8a8d90>
```

## Rationale and Research Questions

To conduct a preliminary exploration of the ways in which the political systems and the electricity grid coexist, our team has formulated the following research questions:

- 1) What is the relationship between the percent penetration of renewable technology (i.e., solar, wind) and state governor (using the political affiliation of the state governor as a proxy for a state's political leanings) in a given year / over a period of time?
- 2) Has renewable penetration grown in any states that have had a single-party governor over a longer period of time?

## Dataset Information

This dataset was prepared in Fall 2023 as a final project for Environmental Data Analytics (ENV 872) at Duke University's Nicholas School of the Environment.

The repository contains 10 years of data (2010-2020) detailing the United State's renewable energy generation capacity by governing state political party. There are three main categories for our dataset. They are: 1) eGRID plant data with coordinates 2) US state governor data by party and year 3) State geographic boundaries shapefile

Table 1: Data Structure Summary - eGRID

Detail	Description
Data Source	EPA
Retrieved from	<a href="https://www.epa.gov/egrid/download-data">https://www.epa.gov/egrid/download-data</a>
Variables Used	PNAME, PSTATABB, NAMEPCAP, PLFUELCT, LAT, LON
Data Range	2010 - 2020

Table 2: Data Structure Summary - Governors

Detail	Description
Data Source	University of Pennsylvania
Retrieved from	<a href="https://www.openicpsr.org/openicpsr/project/102000/version/V3/view?path=/openicpsr/102000/fcr:versions/V3/united_states_governors_1775_2020.csv&amp;type=file">https://www.openicpsr.org/openicpsr/project/102000/version/V3/view?path=/openicpsr/102000/fcr:versions/V3/united_states_governors_1775_2020.csv&amp;type=file</a>
Variables Used	governor, state, time_in_office, party, year
Data Range	1775 - 2020

## Wrangling eGRID

```
# Selecting data for desired columns from imported eGRID datasets

# Note column "PLPFGNCT" for 2010 -- note that "PLFUELCT" was dropped and replaced
eGRID2010_sub <- eGRID2010 %>%
  mutate(YEAR="2010") %>%
  select(YEAR,PSTATABB,ORISPL,PNAME,PLPFGNCT,NAMEPCAP,LAT,LON)

eGRID2010_sub$PLFUELCT <- eGRID2010_sub$PLPFGNCT
eGRID2010_sub$PLPFGNCT <- NULL

# Note column "PLPFGNCT" for 2012 -- note that "PLFUELCT" was dropped and replaced
eGRID2012_sub <- eGRID2012 %>%
  mutate(YEAR="2012") %>%
  select(YEAR,PSTATABB,ORISPL,PNAME,PLPFGNCT,NAMEPCAP,LAT,LON)

eGRID2012_sub$PLFUELCT <- eGRID2012_sub$PLPFGNCT
eGRID2012_sub$PLPFGNCT <- NULL

eGRID2014_sub <- eGRID2014 %>%
  mutate(YEAR="2014") %>%
  select(YEAR,PSTATABB,ORISPL,PNAME,PLFUELCT,NAMEPCAP,LAT,LON)

eGRID2016_sub <- eGRID2016 %>%
  mutate(YEAR="2016") %>%
  select(YEAR,PSTATABB,ORISPL,PNAME,PLFUELCT,NAMEPCAP,LAT,LON)

eGRID2018_sub <- eGRID2018 %>%
  select(YEAR,PSTATABB,ORISPL,PNAME,PLFUELCT,NAMEPCAP,LAT,LON)

eGRID2020_sub <- eGRID2020 %>%
  select(YEAR,PSTATABB,ORISPL,PNAME,PLFUELCT,NAMEPCAP,LAT,LON)

# Merging eGRID data, filtering for renewable fuels
eGRID_2010_2020 <- rbind(eGRID2010_sub,eGRID2012_sub,eGRID2014_sub,eGRID2016_sub,
  eGRID2018_sub,eGRID2020_sub)

# Checking fuel types
unique(eGRID_2010_2020$PLFUELCT)
```

```
## [1] "OIL"      "GAS"      "HYDRO"    "COAL"     NA
## [6] "WIND"     "OTHRFOSL" "BIOMASS"  "NUCLEAR"  "SOLAR"
## [11] "GEOTHERMAL" "WSTHTOTPUR" "OTHF"    "OFSL"
```

```
# Filtering for desired fuel types
eGRID_2010_2020_RENEW <- filter(eGRID_2010_2020,PLFUELCT=="SOLAR" |
  PLFUELCT=="WIND" |
  PLFUELCT=="GEOTHERMAL" |
  PLFUELCT=="HYDRO" |
  PLFUELCT=="BIOMASS" |
  PLFUELCT=="NUCLEAR")
```

```
# Change year column to character  
eGRID_2010_2020_RENEW$YEAR <- as.character(eGRID_2010_2020_RENEW$YEAR)
```

```
# Checking fuel types were filtered correctly  
unique(eGRID_2010_2020_RENEW$PLFUELCT)
```

```
## [1] "HYDRO"      "WIND"        "BIOMASS"     "NUCLEAR"     "SOLAR"  
## [6] "GEOTHERMAL"
```

```
# Checking to ensure all years of interest are present  
unique(eGRID_2010_2020_RENEW$YEAR)
```

```
## [1] "2010" "2012" "2014" "2016" "2018" "2020"
```

```
# Saving process data as CSV  
write_excel_csv(eGRID_2010_2020_RENEW,  
                path = "Data/Processed/eGRID_2010_2020_RENEW.csv")
```

```
## Warning: The 'path' argument of 'write_excel_csv()' is deprecated as of readr 1.4.0.  
## i Please use the 'file' argument instead.  
## This warning is displayed once every 8 hours.  
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was  
## generated.
```

## Wrangling Gov Data

```
# Filtering governor data for 2010 to 2020
Gov_Data_10_20 <- filter(Gov_Data, year >= 2010)

# Creating State abbreviation column for gov data
Gov_Data_State <- Gov_Data_10_20 %>%
  mutate(StateAbbreviation = state.abb[match(state, state.name)])

# Changing year to character
Gov_Data_State <- Gov_Data_State %>%
  mutate(year = as.character(year))

# Separating time in office column
# Key assumption is that end year is the year to be used for analysis of
# political party
# The rationale for this is that the incumbent governor is in office for the majority of the end year
Gov_Data_State <- Gov_Data_State %>%
  separate(time_in_office, into = c("Start_Year", "End_Year"), sep = " - ")

# Filtering for columns of interest
States_shapefile_shp_filtered <- States_shapefile_shp %>%
  select(basename, stusab, geometry)

# Merging governor data with state shape file
GOV_States <- left_join(Gov_Data_State, States_shapefile_shp_filtered, by =
  c("StateAbbreviation" = "stusab"))

# Joining gov data with eGRID processed data
GOV_eGRID_10_20 <- left_join(eGRID_2010_2020_RENEW, Gov_Data_State,
  by = c("PSTATABB" = "StateAbbreviation",
    "YEAR" = "year"))

## Warning in left_join(eGRID_2010_2020_RENEW, Gov_Data_State, by = c(PSTATABB = "StateAbbreviation", :
## i Row 774 of 'x' matches multiple rows in 'y'.
## i Row 305 of 'y' matches multiple rows in 'x'.
## i If a many-to-many relationship is expected, set 'relationship =
##   "many-to-many"' to silence this warning.

# Dropping N/A's
GOV_eGRID_10_20 <- GOV_eGRID_10_20 %>%
  drop_na(LAT)

# Checking years of final dataset
unique(GOV_eGRID_10_20$YEAR)

## [1] "2010" "2012" "2014" "2016" "2018" "2020"

# Checking fuel types of final dataset
unique(GOV_eGRID_10_20$PLFUELCT)
```



```
## [1] "HYDRO"      "WIND"        "BIOMASS"     "NUCLEAR"     "SOLAR"
## [6] "GEOTHERMAL"
```

```
# Dataset information -- checking dimensions and column names
dim(GOV_eGRID_10_20)
```

```
## [1] 33477    13
```

```
colnames(GOV_eGRID_10_20)
```

```
## [1] "YEAR"      "PSTATABB"   "ORISPL"     "PNAME"      "NAMEPCAP"
## [6] "LAT"       "LON"        "PLFUELCT"   "governor"   "state"
## [11] "Start_Year" "End_Year"   "party"
```

```
# Adding Final Processed file to processed folder
write_csv(GOV_eGRID_10_20,
          path = "Data/Processed/GOV_eGRID_10_20.csv")
```

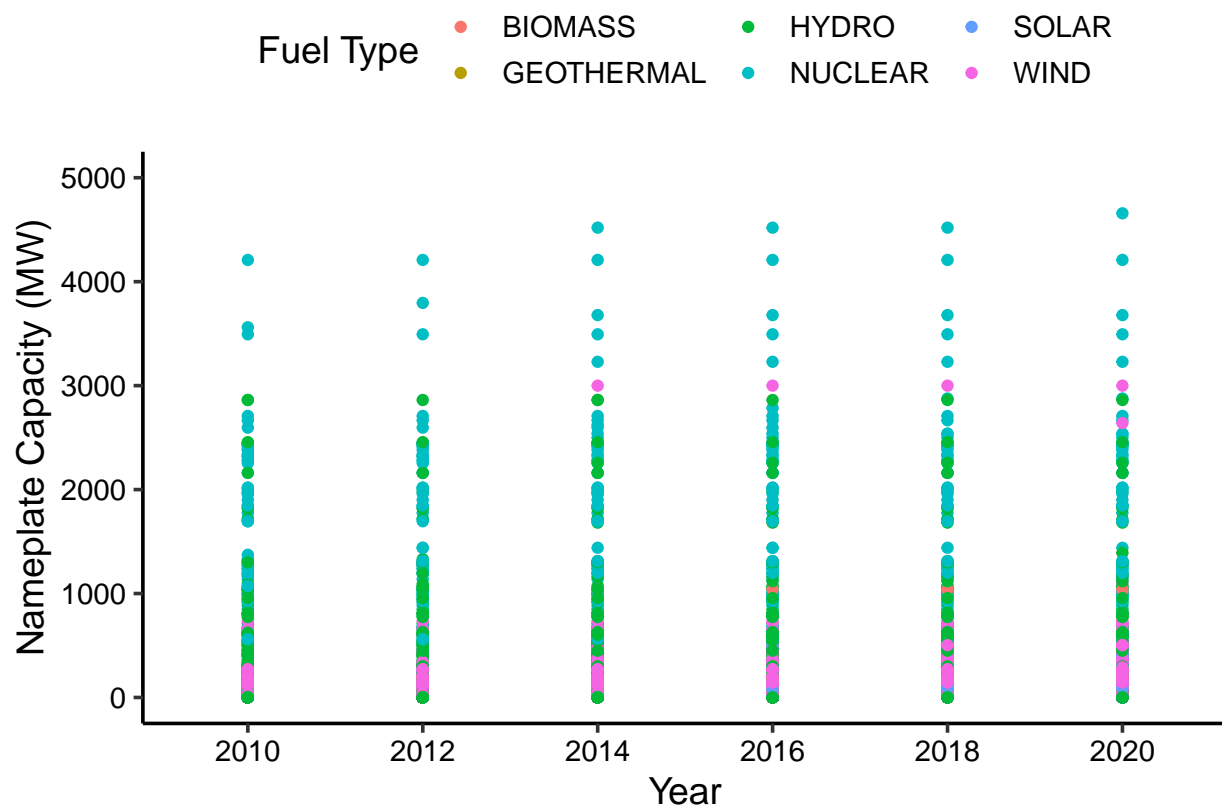
```
## Warning: The 'path' argument of 'write_csv()' is deprecated as of readr 1.4.0.
## i Please use the 'file' argument instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

## Exploratory Analysis

```
# Exploratory distribution analysis using box, point, and violin plot

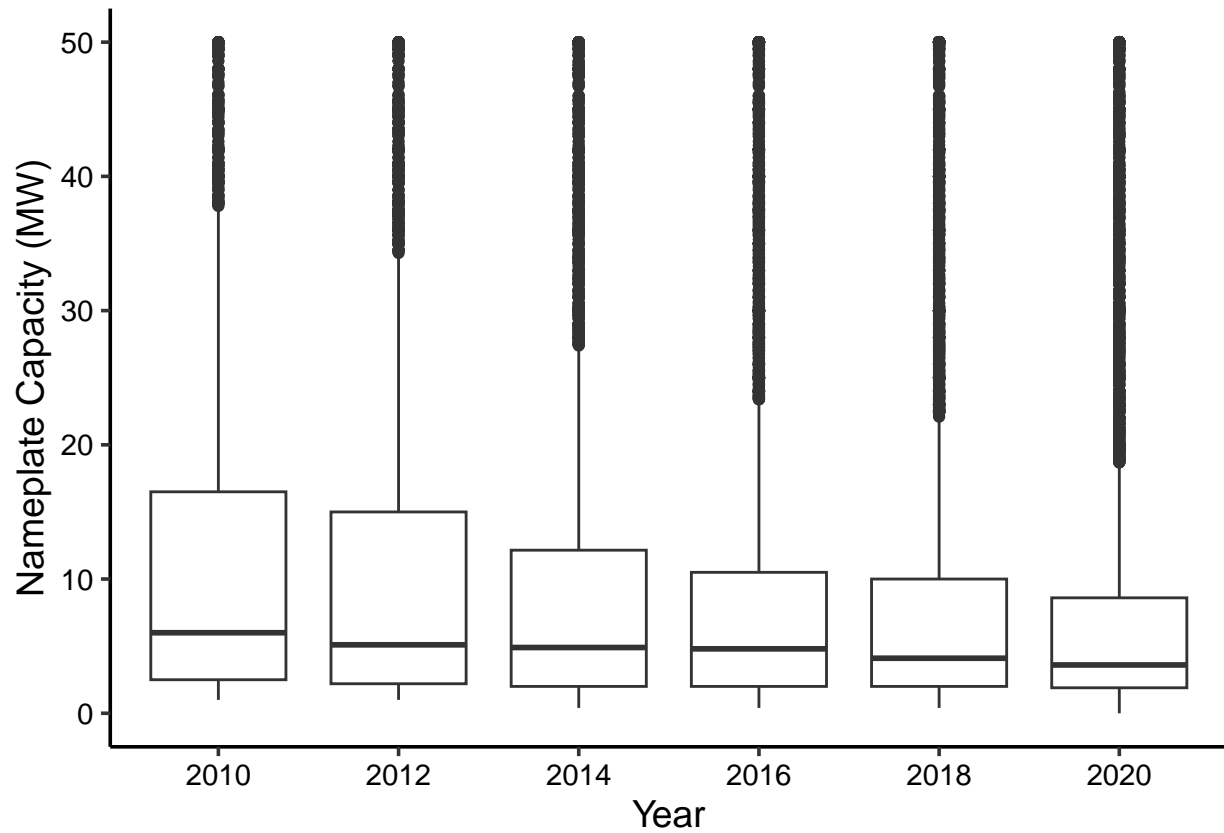
ggplot(GOV_eGRID_10_20, aes(x = YEAR, y = NAMEPCAP, color= PLFUELCT)) +
  geom_point() +
  ylim(0, 5000) +
  labs(x = "Year", y = "Nameplate Capacity (MW)", color = "Fuel Type")
```

## Warning: Removed 7 rows containing missing values ('geom\_point()').



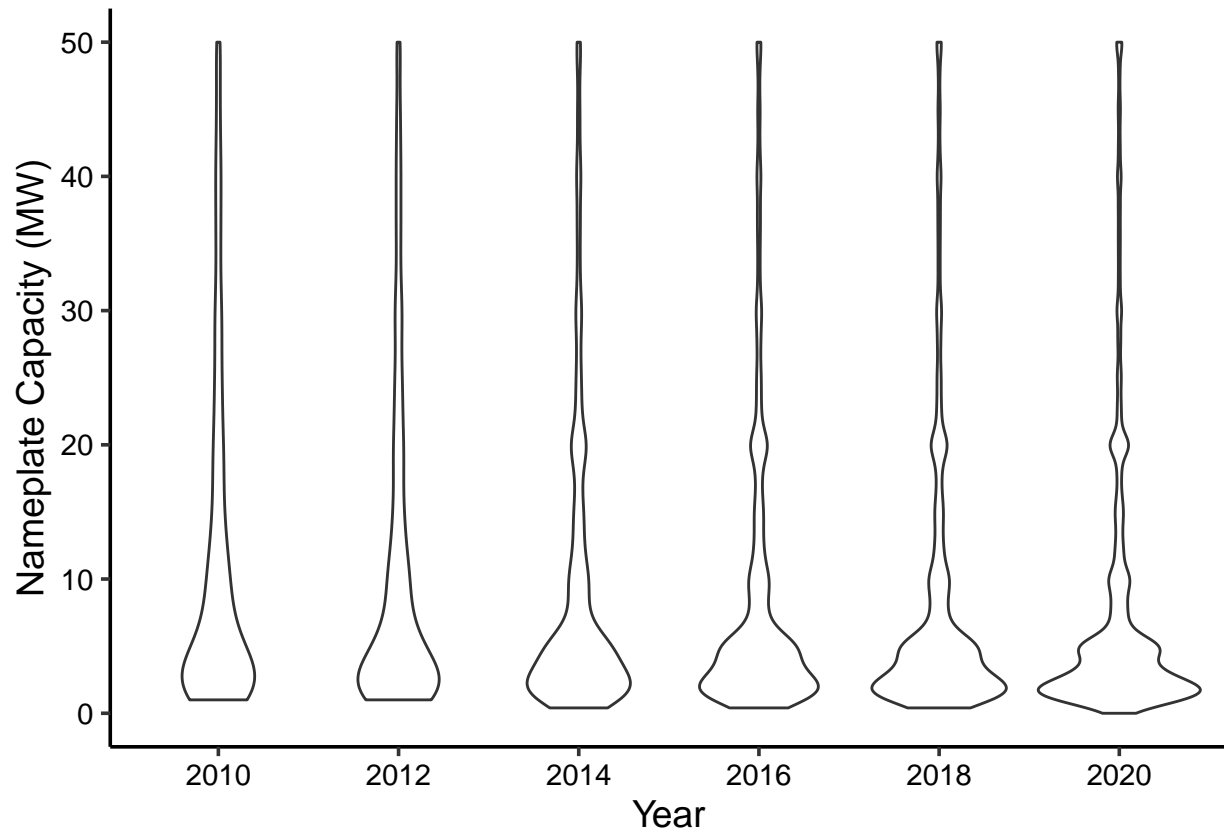
```
ggplot(GOV_eGRID_10_20, aes(x = YEAR, y = NAMEPCAP)) +
  geom_boxplot() +
  ylim(0, 50) +
  labs(x = "Year", y = "Nameplate Capacity (MW)")
```

## Warning: Removed 7713 rows containing non-finite values ('stat\_boxplot()').

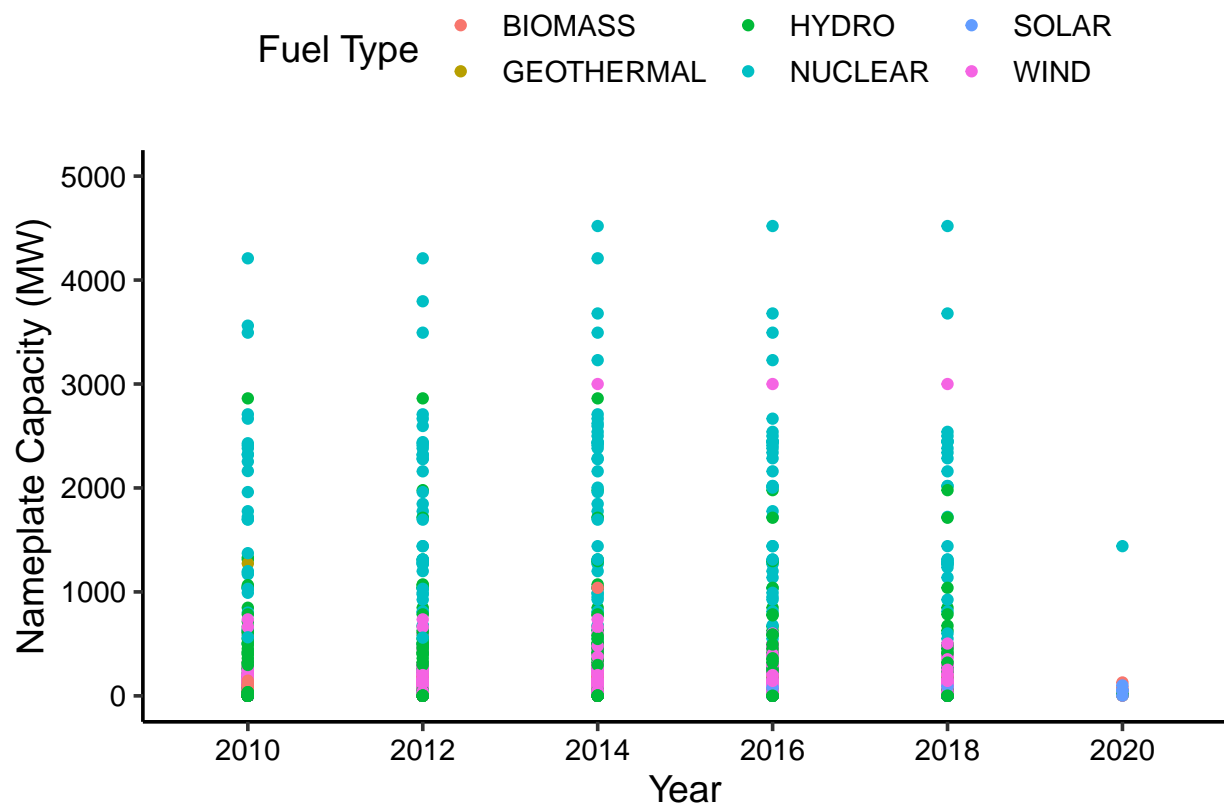


```
ggplot(GOV_eGRID_10_20, aes(x = YEAR, y = NAMEPCAP)) +  
  geom_violin() +  
  ylim(0, 50) +  
  labs(x = "Year", y = "Nameplate Capacity (MW)")
```

```
## Warning: Removed 7713 rows containing non-finite values ('stat_ydensity()').
```

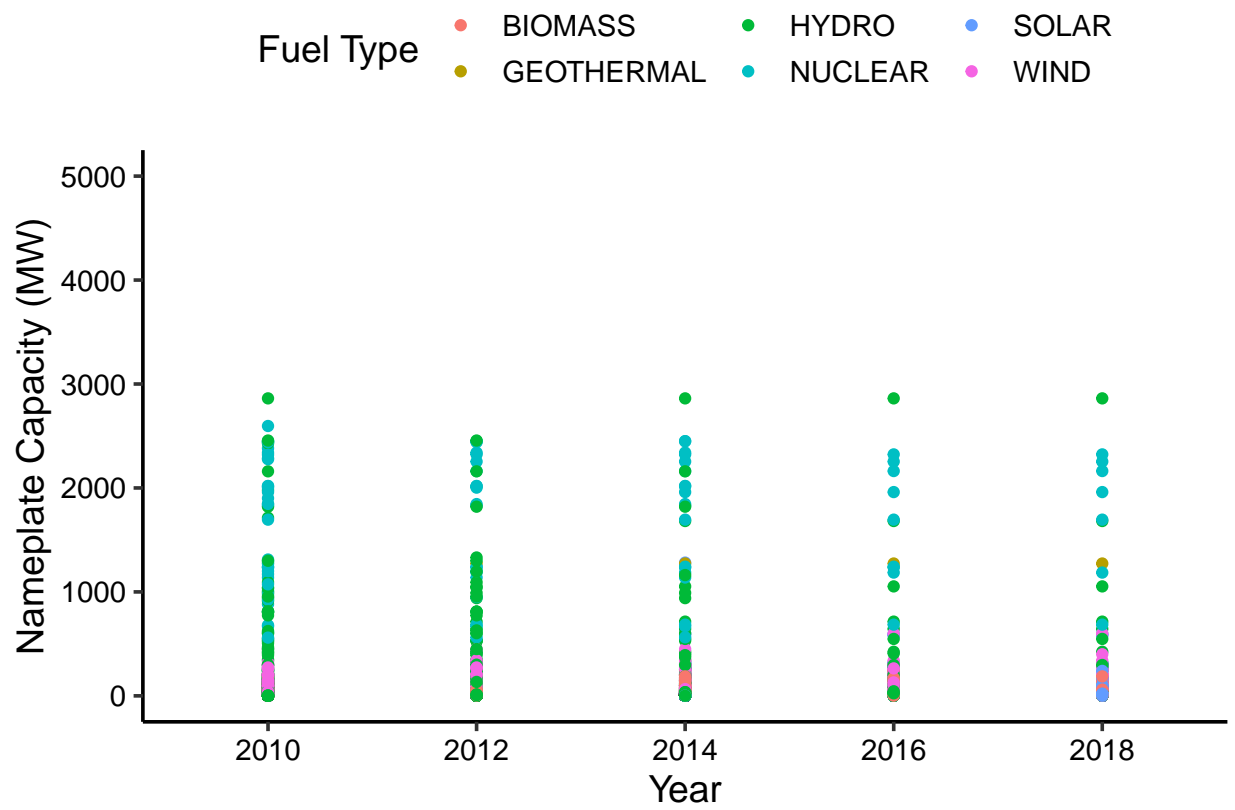


```
Republican_Renewables <- GOV_eGRID_10_20 %>%  
  filter(party == "Republican") %>%  
  ggplot(aes(x = YEAR, y = NAMEPCAP, color = PLFUELCT)) +  
  geom_point() +  
  ylim(0, 5000) +  
  labs(x = "Year", y = "Nameplate Capacity (MW)", color = "Fuel Type")  
plot(Republican_Renewables)
```



```
Democrat_Renewables <- GOV_eGRID_10_20 %>%
  filter(party == "Democrat") %>%
  ggplot(aes(x = YEAR, y = NAMEPCAP, color = PLFUELCT)) +
  geom_point() +
  ylim(0, 5000) +
  labs(x = "Year", y = "Nameplate Capacity (MW)", color = "Fuel Type")
plot(Democrat_Renewables)
```

## Warning: Removed 2 rows containing missing values ('geom\_point()').



## Analysis

```
# WRANGLING
# Creating subset of governor data to sum no. of democrats in office by state
Years_in_Office <- Gov_Data_State %>%
  mutate(dem_count = ifelse(party=="Democrat",1,0),
         rep_count=ifelse(party=="Republican",1,0))%>%
  group_by(StateAbbreviation)%>%
  summarise(total_dem_count = sum(dem_count),total_rep_count=sum(rep_count))

# Aggregating nameplate capacity by state
GOV_eGRID_10_20_TOTAL_CAPACITY <- GOV_eGRID_10_20 %>%
  group_by(PSTATABB) %>%
  summarize(TOTAL_NAMEPCAP = sum(NAMEPCAP))

# Joining datasets to create data set for regression analysis
#Note that all subsequent analyses depend on the below dataframe, which
#aggregates nameplate capacity for each US State and the number of years across
#the time period of interest during which a Democrat or a Republican governor
#was in office. Structuring the data in this manner enabled statistical analysis
#assuming a linear relationship and was per the suggestion of the course
#instructor.
Capacity_by_Party <-left_join(Years_in_Office,GOV_eGRID_10_20_TOTAL_CAPACITY,
                             by = c("StateAbbreviation"="PSTATABB"))

# SINGLE REGRESSION
# Running simple linear regression of Democratic party on nameplate capacity
Regression_DemParty_by_Capacity <-
  lm(Capacity_by_Party$TOTAL_NAMEPCAP~Capacity_by_Party$total_dem_count)
summary(Regression_DemParty_by_Capacity)

##
## Call:
## lm(formula = Capacity_by_Party$TOTAL_NAMEPCAP ~ Capacity_by_Party$total_dem_count)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -42707 -26809 -19512  12994 194576
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      44694       9915   4.508 4.49e-05 ***
## Capacity_by_Party$total_dem_count      -201       2094  -0.096   0.924
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 48170 on 46 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.0002003, Adjusted R-squared:  -0.02153
## F-statistic: 0.009217 on 1 and 46 DF, p-value: 0.9239
```

```

Regression_RepParty_by_Capacity <-
  lm(Capacity_by_Party$TOTAL_NAMEPCAP~Capacity_by_Party$total_rep_count)
summary(Regression_RepParty_by_Capacity)

##
## Call:
## lm(formula = Capacity_by_Party$TOTAL_NAMEPCAP ~ Capacity_by_Party$total_rep_count)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -41839 -27408 -19556  12734 194640
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      42138.1     11597.6   3.633 0.000702 ***
## Capacity_by_Party$total_rep_count    341.4      1687.9   0.202 0.840611
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 48160 on 46 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.0008885, Adjusted R-squared:  -0.02083
## F-statistic: 0.04091 on 1 and 46 DF, p-value: 0.8406

# Plotting single regressions - capacity by Demomcratic years in office
ggplot(Capacity_by_Party, aes(x=total_dem_count, y=TOTAL_NAMEPCAP)) +
  geom_point() +
  geom_smooth(method = lm) +
  labs(x = "Number of Years Democrat in Office", y = "Nameplate Capacity (MW)")

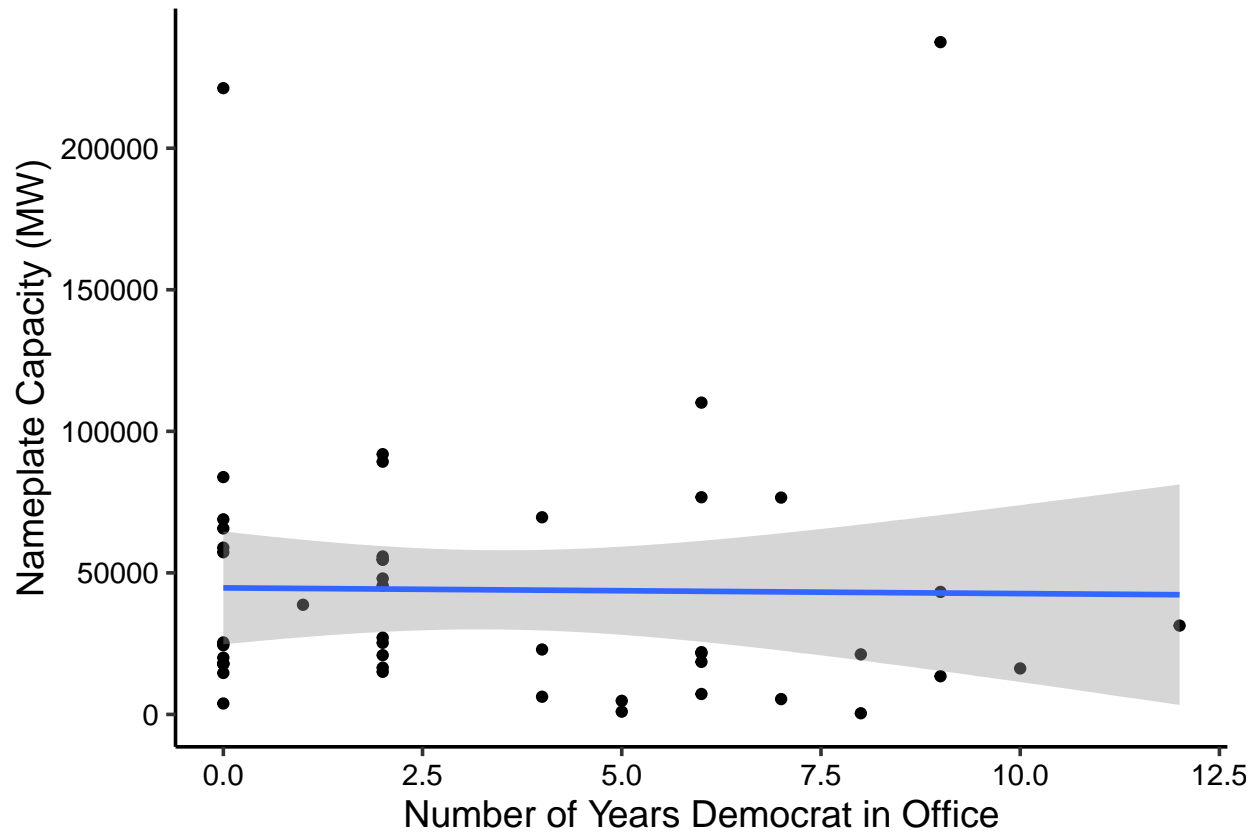
## 'geom_smooth()' using formula = 'y ~ x'

## Warning: Removed 2 rows containing non-finite values ('stat_smooth()').

## Warning: Removed 2 rows containing missing values ('geom_point()').

```



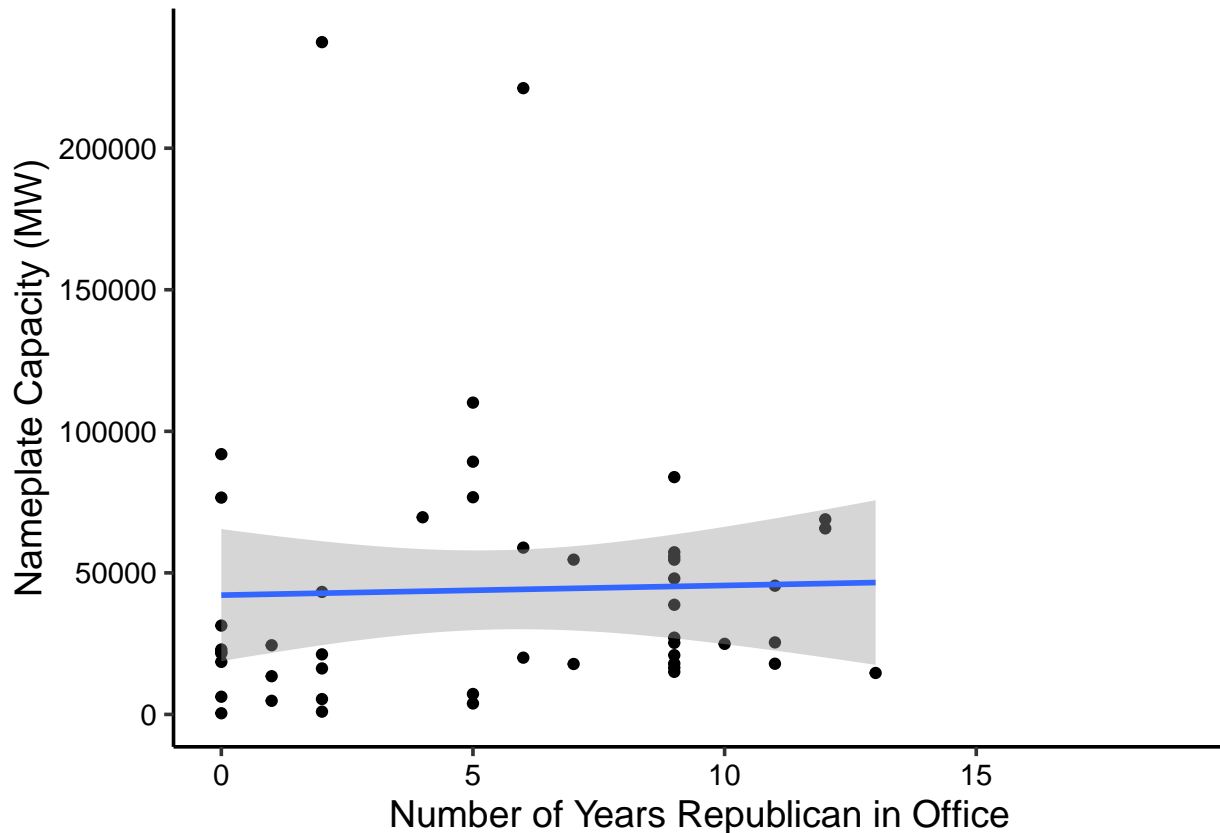


```
# Plotting single regressions - capacity by Republican years in office
ggplot(Capacity_by_Party, aes(x=total_rep_count, y=TOTAL_NAMEPCAP)) +
  geom_point() +
  geom_smooth(method = lm) +
  labs(x = "Number of Years Republican in Office", y = "Nameplate Capacity (MW)")
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

```
## Warning: Removed 2 rows containing non-finite values ('stat_smooth()').
```

```
## Removed 2 rows containing missing values ('geom_point()').
```



```
# Results
#The x variable is "number of years a Democratic governor was in office" and the
#y variable is "Total nameplate capacity from renewable energy sources".
#Ho: There is no relationship between renewable capacity and the number of
#years a Democratic governor was in office.
#Given the R-squared value, virtually none of the variability in total
#nameplate capacity from renewable energy sources is explained by the number of
#years a Democratic governor was in office. Given the high p-value of 0.9239,
#we fail to reject H0. Thus, there is no relationship between renewable
#capacity and the number of years a Democratic governor was in office.
#(Linear regression; df = 46, R-squared = 0.0002003, F-statistic = 0.009217,
#p-value = 0.9239)
#Note that this single regression can also be applied to "# of years a
#Republican governor was in office" -- assuming the party is binary between
#Democrat and Republican, this should drive a similar regression result.
#Indeed, that is the case as evidenced by the analogous output for Republican
#governors.
#See the below plots of these regressions which help to visualize
#the numerical outputs referenced in the write-ups above.

# MULTIPLE REGRESSION
# Running multiple linear regression
MP_Reg_Capacity_by_Party <-
  lm(data=Capacity_by_Party, TOTAL_NAMEPCAP~total_dem_count+total_rep_count)

summary(MP_Reg_Capacity_by_Party)
```

```
##
## Call:
## lm(formula = TOTAL_NAMEPCAP ~ total_dem_count + total_rep_count,
##     data = Capacity_by_Party)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -42044 -27687 -19314  13129 193840
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   40664.3    23407.9   1.737   0.0892 .
## total_dem_count    223.6     3073.7   0.073   0.9423
## total_rep_count    472.1     2478.5   0.190   0.8498
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 48690 on 45 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.001006, Adjusted R-squared: -0.04339
## F-statistic: 0.02266 on 2 and 45 DF, p-value: 0.9776
```

```
# Results
#The predictor variables are "number of years a Democratic governor was in
#office" and "# of years a Republican governor was in office" and the y
#variable is "Total nameplate capacity from renewable energy sources".
#Given the R-squared value, virtually none of the variability in total
#nameplate capacity from renewable energy sources is explained by the number of
#years a Democratic governor was in office and the number of years that a
#Republican governor was in office. Given the high p-value of 0.9776, we fail
#to reject H0. Thus, there is no relationship between renewable capacity and
#the number of years a Democratic governor was in office and the number of
#years a Republican governor was in office.
 #(Multiple regression; df = 45, R-squared = 0.001006, F-statistic = 0.02266,
#p-value = 0.9776)

# PEARSON CORRELATION
# Running a correlation test
cor.test(Capacity_by_Party$TOTAL_NAMEPCAP, Capacity_by_Party$total_dem_count)
```

```
##
## Pearson's product-moment correlation
##
## data: Capacity_by_Party$TOTAL_NAMEPCAP and Capacity_by_Party$total_dem_count
## t = -0.096005, df = 46, p-value = 0.9239
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.2970936 0.2710711
## sample estimates:
##      cor
## -0.01415372
```

```

# Results
#The x variable is "number of years a Democratic governor was in office" and
#the y variable is "Total nameplate capacity from renewable energy sources".
#The calculated correlation coefficient between the two variables is
#approximately -0.014, which suggests a very weak relationship between the
#variables x and y. Moreover, the high p-value means that we fail to reject
#the null hypothesis (Ho: There is no relationship between renewable capacity
#and the number of years a Democratic governor was in office). Thus, it appears
#that there is no significant correlation between the variables.
#(Pearson correlation; df = 46, Correlation coefficient = -0.01415372,
#p-value = 0.9239, 95% confidence interval ranges from -0.2970936 to 0.2710711)

# ANOVA
#ANOVA Test for party correlative effect on nameplate capacity
ANOVA_Capacity <- aov(data = Capacity_by_Party, TOTAL_NAMEPCAP ~ total_dem_count)
summary(ANOVA_Capacity)

```

```

##              Df      Sum Sq   Mean Sq F value Pr(>F)
## total_dem_count  1 2.139e+07 2.139e+07   0.009  0.924
## Residuals       46 1.068e+11 2.321e+09
## 2 observations deleted due to missingness

```

```

# Results
#Similar to the results above, the high p-value of 0.924 indicates that the
#relationship between "number of years a Democratic governor was in office" and
#"total nameplate capacity between from renewable energy sources" is not
#statistically significant (ANOVA; df = 1, F-value = 0.009, p-value = 0.924,.).

```

```

# TIME SERIES PLOT

```

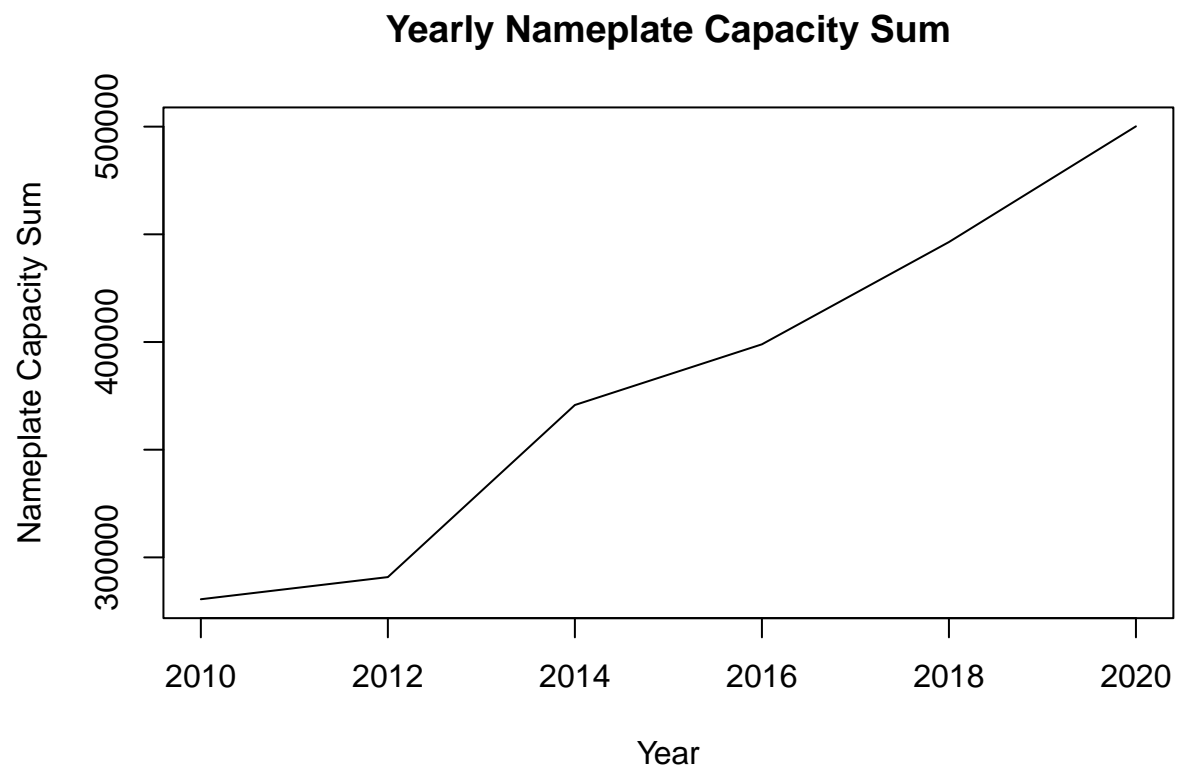
```

GOV_eGRID_10_20_aggregate <- GOV_eGRID_10_20 %>%
  group_by(YEAR) %>%
  summarize(NAMEPCAPSUM = sum(NAMEPCAP, na.rm = TRUE))

GOV_eGRID.yearly.ts <- ts(GOV_eGRID_10_20_aggregate$NAMEPCAPSUM,
  start=c(2010,1),
  end = c(2020,1),
  deltat = 2)

plot(GOV_eGRID.yearly.ts,
  main = "Yearly Nameplate Capacity Sum",
  xlab = "Year",
  ylab = "Nameplate Capacity Sum")

```



```
# MANN-KENDALL TEST
```

```
GOV_eGRID.yearly.ts.mk <- Kendall::MannKendall(GOV_eGRID.yearly.ts)  
summary(GOV_eGRID.yearly.ts.mk)
```

```
## Score = 15 , Var(Score) = 28.33333  
## denominator = 15  
## tau = 1, 2-sided pvalue =0.0085349
```

```
# Results
```

```
#Given a p-value of less than 0.05 at 0.0085349, we reject H0 and conclude  
#that there is a monotonic, yearly trend to the data. Because tau is positive,  
#we conclude that the trend is increasing.  
#(Mann-Kendall; tau = 1, p-value = 0.0085349)
```

# Geospatial Analysis

```
# EXPLORATORY
# Viewing State Boundaries
mapview(States_shapefile_shp)
```

```
# NC GEOSPATIAL ANALYSIS
# Viewing NC power plants in three different years
Gov_eGRID_NC_2012.sf <- GOV_eGRID_10_20 %>%
  st_as_sf(coords = c('LON', 'LAT'), crs=4326) %>%
  filter(PSTATABB == "NC") %>%
  filter(YEAR == 2010:2012)

mapview(Gov_eGRID_NC_2012.sf, zcol="PLFUELCT")
```

```
Gov_eGRID_NC_2016.sf <- GOV_eGRID_10_20 %>%
  st_as_sf(coords = c('LON', 'LAT'), crs=4326) %>%
  filter(PSTATABB == "NC") %>%
  filter(YEAR == 2010:2016)
```

```
## Warning: There was 1 warning in 'stopifnot()'.
## i In argument: 'YEAR == 2010:2016'.
## Caused by warning in 'YEAR == 2010:2016':
## ! longer object length is not a multiple of shorter object length
```

```
mapview(Gov_eGRID_NC_2016.sf, zcol="PLFUELCT")
```

```
Gov_eGRID_NC_2020.sf <- GOV_eGRID_10_20 %>%
  st_as_sf(coords = c('LON', 'LAT'), crs=4326) %>%
  filter(PSTATABB == "NC") %>%
  filter(YEAR == 2010:2020)
```

```
## Warning: There was 1 warning in 'stopifnot()'.
## i In argument: 'YEAR == 2010:2020'.
## Caused by warning in 'YEAR == 2010:2020':
## ! longer object length is not a multiple of shorter object length
```

```
mapview(Gov_eGRID_NC_2020.sf, zcol="PLFUELCT")
```

```
# Combine spatial layers for NC
# Filter NC State Boundary by party
GOV_States.sf <- st_as_sf(GOV_States) %>%
  filter(StateAbbreviation == "NC")

# Layering NC Boundaries and plants for 2012
mapview(GOV_States.sf, col.regions = "blue") +
  mapview(Gov_eGRID_NC_2012.sf, zcol="PLFUELCT")
```

```
# Layering NC Boundaries and plants for 2016
mapview(GOV_States.sf, col.regions = "red") +
  mapview(Gov_eGRID_NC_2016.sf, zcol="PLFUELCT")
```

```
# Layering NC Boundaries and plants for 2020
mapview(GOV_States.sf, col.regions = "blue") +
  mapview(Gov_eGRID_NC_2020.sf, zcol="PLFUELCT")
```

```
# TX GEOSPATIAL ANALYSIS
# Viewing TX power plants in three different years
Gov_eGRID_TX_2012.sf <- GOV_eGRID_10_20 %>%
  st_as_sf(coords = c('LON', 'LAT'), crs=4326) %>%
  filter(PSTATABB == "TX") %>%
  filter(YEAR == 2012)

mapview(Gov_eGRID_TX_2012.sf, zcol="PLFUELCT")
```

```
Gov_eGRID_TX_2020.sf <- GOV_eGRID_10_20 %>%
  st_as_sf(coords = c('LON', 'LAT'), crs=4326) %>%
  filter(PSTATABB == "TX") %>%
  filter(YEAR == 2012:2020)
```

```
## Warning: There was 1 warning in 'stopifnot()'.
## i In argument: 'YEAR == 2012:2020'.
## Caused by warning in 'YEAR == 2012:2020':
## ! longer object length is not a multiple of shorter object length
```

```
mapview(Gov_eGRID_TX_2020.sf, zcol="PLFUELCT")
```

```
# Combine spatial layers for TX
# Filter TX State Boundary by party
GOV_States.sf <- st_as_sf(GOV_States) %>%
  filter(StateAbbreviation == "TX")

# Layering TX Boundaries and plants for 2012
mapview(GOV_States.sf, col.regions = "red") +
  mapview(Gov_eGRID_TX_2012.sf, zcol="PLFUELCT")
```

```
# Layering TX Boundaries and plants for 2016
mapview(GOV_States.sf, col.regions = "red") +
  mapview(Gov_eGRID_TX_2020.sf, zcol="PLFUELCT")
```

```
# CA GEOSPATIAL ANALYSIS
# Viewing CA power plants in three different years
Gov_eGRID_CA_2012.sf <- GOV_eGRID_10_20 %>%
  st_as_sf(coords = c('LON', 'LAT'), crs=4326) %>%
  filter(PSTATABB == "CA") %>%
  filter(YEAR == 2010:2012)
```

```
## Warning: There was 1 warning in 'stopifnot()'.
## i In argument: 'YEAR == 2010:2012'.
## Caused by warning in 'YEAR == 2010:2012':
## ! longer object length is not a multiple of shorter object length
```

```
mapview(Gov_eGRID_CA_2012.sf, zcol="PLFUELCT")
```

```
Gov_eGRID_CA_2020.sf <- GOV_eGRID_10_20 %>%  
  st_as_sf(coords = c('LON','LAT'), crs=4326) %>%  
  filter(PSTATABB == "CA") %>%  
  filter(YEAR == 2012:2020)
```

```
## Warning: There was 1 warning in 'stopifnot()'.  
## i In argument: 'YEAR == 2012:2020'.  
## Caused by warning in 'YEAR == 2012:2020':  
## ! longer object length is not a multiple of shorter object length
```

```
mapview(Gov_eGRID_CA_2020.sf, zcol="PLFUELCT")
```

```
# Combine spatial layers for CA  
# Filter CA State Boundary by party  
GOV_States.sf <- st_as_sf(GOV_States) %>%  
  filter(StateAbbreviation == "CA")  
  
# Layering CA Boundaries and plants for 2012  
mapview(GOV_States.sf, col.regions = "blue") +  
  mapview(Gov_eGRID_CA_2012.sf, zcol="PLFUELCT")
```

```
# Layering CA Boundaries and plants for 2016  
mapview(GOV_States.sf, col.regions = "blue") +  
  mapview(Gov_eGRID_CA_2020.sf, zcol="PLFUELCT")
```



## Summary and Conclusions

As it relates to the first research question, there is no apparent relationship between the percent penetration of renewable energy technologies and state governor over the period from 2010 - 2020. There are a number of important caveats to make here – renewable energy development depends on many factors ranging from federal incentives to state-level renewable portfolio standards to local zoning ordinances and regulations. Moreover, the dynamics of power markets / grid operators (e.g., investor-owned / vertically integrated / non-competitive markets such as in the southeast versus competitive markets such as ERCOT in Texas) are critically influential, particularly for the process of interconnection. Regarding the second research question, spatial analysis was conducted across three example states in California (Democratic governor for the entire time period of interest), Texas (Republican governor for the entire time period of interest), and North Carolina (both Democratic and Republican governors for the time period of interest). In summary, the output here shows that all three states had noticeable increases in renewable generation capacity online despite their differences in political leanings. While this project led to some preliminary, high-level conclusions, it prompts a number of questions for further exploration: 1. What is the magnitude of changes in renewables in the three example states of California, Texas, and North Carolina? 2. How do local (e.g., country-level) politics influence renewable energy generation by sub-region of states?