# Energy Generation Analysis Report

The Energy Generators

2022-12-07

## Group Members


- Richard Abegg
- Matthew McCullough
- Shivani Sharma
- Kevin Tam


## Abstract

The United States uses various energy sources to create electricity. Over time, the energy sources change, resulting in some becoming used more than others. Our data set looks at the monthly United States energy generation per state, type of producer, and type of source from 2001 to 2022. The set is comprised of data collected by the EIA (US Energy Information Administration) which shows the types of energies produced by each state. The following columns of information are collected: year, month, state, type of producer, energy source, and generation. The three major categories of energy sources found from the data set are fossil fuels, nuclear and renewable energy. These energy sources combined with the producer types create energy for the entire country. From this data, we hope the answer the following questions below.

## Introduction

**Questions within the data - Clean up when creating report**

- Does the cost of fuel influence the energy production rates? Kevin
- How has the reduction in renewables cost affected generation numbers? Matt
- Which energy source produces the most energy? Rich
- What is the average yearly production of each energy source by state? Rich mess with this to show
- Which does the future of renewable energy look like? Shivani - reword how every you need to
- When would you expect renewable energy to surpass fossil fuels? Matt

## Data Collection and Arrangement

```
# Read in Data
energy = read_csv("organised_Gen.csv") # Data in organised_gen is by month from 2001 to 2022
```

```
## New names:
## Rows: 496774 Columns: 7
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (3): STATE, TYPE OF PRODUCER, ENERGY SOURCE dbl (4): ...1, YEAR, MONTH,
## GENERATION (Megawatthours)
## i Use `spec()` to retrieve the full column specification for this data. i
```

```
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `` -> `...1`
```

```
states = read_csv("states.csv") # 51 states inclusive of DC with coal region
```

```
## Rows: 51 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (4): State, Abbrev, Code, Coal_Region
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
og = read_csv("OilandGas.csv") # Daily oil and natural gas prices 01/2000 to 02/2022
```

```
## Rows: 5785 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl  (2): Price_oil, Price_natural_gas
## date (1): Date
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
uranium = read_csv("uranium.csv") # Monthly Uranium prices from 01/2000 to 11/2021
```

```
## Rows: 265 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Month
## dbl (2): Year, Price_uran
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
coal = read_csv("coal.csv") # Weekly coal prices from 05/2008 to 10/2021
```

```
## Rows: 745 Columns: 6
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl  (5): CAP, NAP, ILB, PRB, UIB
## date (1): Week_Ended
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# generate a whole data set
dateStart = make_date(2008,5,1)
dateEnd = make_date(2021,10,30)

print(dateStart)
```

```
## [1] "2008-05-01"
```

```
print(dateEnd)
```

```
## [1] "2021-10-30"
```

```r
# Energy
energyWhole = energy %>%
  mutate(date_col = make_date(year=YEAR,month=MONTH)) %>%
  select(date_col, everything()) %>%
  filter(date_col >= dateStart & date_col <= dateEnd) %>%
  arrange(date_col)
print(energyWhole)
```

```
## # A tibble: 325,431 x 8
##    date_col    ...1  YEAR MONTH STATE `TYPE OF PRODUCER`         ENERG~1 GENER~2
##    <date>     <dbl> <dbl> <dbl> <chr> <chr>                      <chr>     <dbl>
##  1 2008-05-01  7446  2008     5 AK    Total Electric Power Indu~ Coal      52608
##  2 2008-05-01  7447  2008     5 AK    Total Electric Power Indu~ Petrol~   80574
##  3 2008-05-01  7448  2008     5 AK    Total Electric Power Indu~ Natura~  296329
##  4 2008-05-01  7449  2008     5 AK    Total Electric Power Indu~ Hydroe~   77437
##  5 2008-05-01  7450  2008     5 AK    Total Electric Power Indu~ Wind          6
##  6 2008-05-01  7451  2008     5 AK    Total Electric Power Indu~ Other ~     488
##  7 2008-05-01  7452  2008     5 AK    Total Electric Power Indu~ Other         0
##  8 2008-05-01  7453  2008     5 AK    Total Electric Power Indu~ Total    507442
##  9 2008-05-01  7454  2008     5 AK    Electric Generators, Elec~ Coal      19071
## 10 2008-05-01  7455  2008     5 AK    Electric Generators, Elec~ Petrol~   76098
## # ... with 325,421 more rows, and abbreviated variable names
## #   1: `ENERGY SOURCE`, 2: `GENERATION (Megawatthours)`
```

```r
# States
print(states)
```

```
## # A tibble: 51 x 4
##    State                Abbrev Code  Coal_Region
##    <chr>                <chr>  <chr> <chr>
##  1 Alabama              Ala.   AL    ILB_CAP
##  2 Alaska               Alaska AK    ALL
##  3 Arizona              Ariz.  AZ    UIB
##  4 Arkansas             Ark.   AR    ILB
##  5 California           Calif. CA    UIB_PRB
##  6 Colorado             Colo.  CO    UIB
##  7 Connecticut          Conn.  CT    NAP
##  8 Delaware             Del.   DE    NAP
##  9 District of Columbia D.C.   DC    NAP
## 10 Florida              Fla.   FL    ILB_CAP
## # ... with 41 more rows
```

```r
# OG
ogWhole = og %>%
  filter(Date >= dateStart & Date <= dateEnd) %>%
  mutate(MONTH = month(Date), YEAR = year(Date)) %>%
  group_by(MONTH, YEAR) %>%
  summarize(
    Price_natural_gas_m = mean(Price_natural_gas, na.rm = T),
    Price_oil_m = mean(Price_oil, na.rm = T)) %>%
  mutate(date_col = make_date(year=YEAR,month=MONTH)) %>%
  ungroup() %>%
  arrange(date_col) %>%
  select(date_col, Price_natural_gas_m, Price_oil_m)
```

```
## `summarise()` has grouped output by 'MONTH'. You can override using the
## `.groups` argument.
```

```
print(ogWhole)
```

```
## # A tibble: 162 x 3
##    date_col   Price_natural_gas_m Price_oil_m
##    <date>                   <dbl>       <dbl>
##  1 2008-05-01               11.4        125.
##  2 2008-06-01               12.8        134.
##  3 2008-07-01               11.1        135.
##  4 2008-08-01                8.30       115.
##  5 2008-09-01                7.49       101.
##  6 2008-10-01                6.73        73.7
##  7 2008-11-01                6.70        54.7
##  8 2008-12-01                5.79        43.1
##  9 2009-01-01                5.07        45.7
## 10 2009-02-01                4.37        43.9
## # ... with 152 more rows
```

```
# Uranium
uraniumWhole = uranium %>%
  mutate(date_col = make_date(year = Year, month = match(Month,month.abb))) %>%
  filter(date_col >= dateStart & date_col <= dateEnd) %>%
  arrange(date_col) %>%
  select(date_col, Price_uran)
print(uraniumWhole)
```

```
## # A tibble: 162 x 2
##    date_col   Price_uran
##    <date>          <dbl>
##  1 2008-05-01       61.7
##  2 2008-06-01       59
##  3 2008-07-01       61.8
##  4 2008-08-01       64.5
##  5 2008-09-01       63
##  6 2008-10-01       48.6
##  7 2008-11-01       50.5
##  8 2008-12-01       54.3
##  9 2009-01-01       51.4
## 10 2009-02-01       47
## # ... with 152 more rows
```

```
# Coal
coalWhole = coal %>%
  filter(Week_Ended >= dateStart & Week_Ended <= dateEnd) %>%
  mutate(MONTH = month(Week_Ended), YEAR = year(Week_Ended)) %>%
  group_by(MONTH, YEAR) %>%
  summarize(
    CAP = mean(CAP, na.rm = T),
    NAP = mean(NAP, na.rm = T),
    ILB = mean(ILB, na.rm = T),
    PRB = mean(PRB, na.rm = T),
    UIB = mean(UIB, na.rm = T)) %>%
  mutate(date_col = make_date(year=YEAR,month=MONTH)) %>%
  ungroup() %>%
```

```
  arrange(date_col) %>%
  mutate(
    ALL = (CAP + NAP + ILB + PRB + UIB)/5,
    CAP_NAP = (CAP + NAP)/2,
    ILB_CAP = (ILB + CAP)/2,
    ILB_PRB = (ILB + PRB)/2,
    UIB_PRB = (UIB + PRB)/2,
    UIB_ILB = (UIB + ILB)/2
  ) %>%
  select(date_col, CAP, NAP, ILB, PRB, UIB, ALL, CAP_NAP, ILB_CAP, ILB_PRB, UIB_PRB, UIB_ILB)
```

```
## `summarise()` has grouped output by 'MONTH'. You can override using the
## `.groups` argument.
```

```
print(coalWhole)
```

```
## # A tibble: 162 x 12
##    date_col     CAP   NAP   ILB   PRB   UIB   ALL CAP_NAP ILB_CAP ILB_PRB
##    <date>     <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>   <dbl>   <dbl>   <dbl>
##  1 2008-05-01 120.  110.   63.2  13.7  47.5  70.9   115.     91.5    38.5
##  2 2008-06-01 133.  129.   67.8  13.6  53.8  79.4   131     100.     40.7
##  3 2008-07-01 138.  143.   82.6  12.4  61.4  87.3   140.    110.     47.5
##  4 2008-08-01 135   142.   84    11.1  61    86.7   139.    110.     47.6
##  5 2008-09-01 131   143    84    14.2  60.9  86.6   137     108.     49.1
##  6 2008-10-01 117   121    84    13.7  66.6  80.5   119     100.     48.8
##  7 2008-11-01 104.  102    81    13.0  72    74.4   103.     92.4    47.0
##  8 2008-12-01  79.9  82.2  69.2  13.4  73.5  63.7    81.1    74.6    41.3
##  9 2009-01-01  69.1  66    55.5  13.1  72.5  55.2    67.6    62.3    34.3
## 10 2009-02-01  68.3  61    53    13    71.5  53.4    64.6    60.6    33
## # ... with 152 more rows, and 2 more variables: UIB_PRB <dbl>, UIB_ILB <dbl>
```

```
# Join tables into super table
# Prep Coal
coalWholeLong = coalWhole %>%
  pivot_longer(c(CAP, NAP, ILB, PRB, UIB, ALL, CAP_NAP, ILB_CAP, ILB_PRB, UIB_PRB, UIB_ILB))

print(coalWholeLong)
```

```
## # A tibble: 1,782 x 3
##    date_col   name    value
##    <date>     <chr>   <dbl>
##  1 2008-05-01 CAP     120.
##  2 2008-05-01 NAP     110.
##  3 2008-05-01 ILB      63.2
##  4 2008-05-01 PRB      13.7
##  5 2008-05-01 UIB      47.5
##  6 2008-05-01 ALL      70.9
##  7 2008-05-01 CAP_NAP 115.
##  8 2008-05-01 ILB_CAP  91.5
##  9 2008-05-01 ILB_PRB  38.5
## 10 2008-05-01 UIB_PRB  30.6
## # ... with 1,772 more rows
```

```
# Build Table
allTableWhole = energyWhole %>%
  left_join(states, by = c("STATE"="Code")) %>%
```

```
  select(date_col, STATE, `TYPE OF PRODUCER`,`ENERGY SOURCE`, `GENERATION (Megawatthours)`,Coal_Region) %>%
  pivot_wider(names_from = `ENERGY SOURCE`, values_from = `GENERATION (Megawatthours)`) %>%
  group_by(STATE,date_col,Coal_Region) %>%
  summarize(
    Coal_Gen_Total = sum(Coal, na.rm = T),
    Petroleum_Gen_Total = sum(Petroleum, na.rm = T),
    Natural_Gas_Gen_Total = sum(`Natural Gas`, na.rm = T),
    Hydro_Gen_Total = sum(`Hydroelectric Conventional`, na.rm = T),
    Wind_Gen_Total = sum(Wind, na.rm = T),
    Biomass_Gen_Total = sum(`Other Biomass`, na.rm = T),
    Other_Gen_Total = sum(Other, na.rm = T),
    Other_Gas_Gen_Total = sum(`Other Gases`, na.rm = T),
    Nuclear_Gen_Total = sum(Nuclear, na.rm = T),
    Wood_Gen_Total = sum(`Wood and Wood Derived Fuels`, na.rm = T),
    Storage_Gen_Total = sum(`Pumped Storage`, na.rm = T),
    Solar_Gen_Total = sum(`Solar Thermal and Photovoltaic`, na.rm = T),
    Geothermal_Gen_Total = sum(Geothermal, na.rm = T)
    ) %>%
  left_join(ogWhole, by = c( "date_col"="date_col")) %>%
  left_join(uraniumWhole, by = c( "date_col"="date_col")) %>%
  left_join(coalWholeLong, by = c( "date_col"="date_col", "Coal_Region"="name")) %>%
  rename(Price_coal = value) %>%
  select(-Coal_Region) %>%
  ungroup() %>%
  mutate(Other_Gen_Total_new =
           Other_Gen_Total + Biomass_Gen_Total + Other_Gas_Gen_Total +
           Wood_Gen_Total + Petroleum_Gen_Total + Geothermal_Gen_Total +
           Storage_Gen_Total) %>%
  select(
    STATE, date_col, Coal_Gen_Total, Natural_Gas_Gen_Total,
    Hydro_Gen_Total, Wind_Gen_Total, Solar_Gen_Total,
    Other_Gen_Total_new) %>%
  arrange(date_col)
```

```
## `summarise()` has grouped output by 'STATE', 'date_col'. You can override using
## the `.groups` argument.
```

```
print(allTableWhole)
```

```
## # A tibble: 8,424 x 8
##    STATE date_col   Coal_Gen_Total Natural_Gas~1 Hydro~2 Wind_~3 Solar~4 Other~5
##    <chr> <date>              <dbl>         <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
##  1 AK    2008-05-01         105217        592659  154874      12       0  162124
##  2 AL    2008-05-01       13337496       1459306  829834       0       0  623695
##  3 AR    2008-05-01        3745017       1192742  872207       0       0  265714
##  4 AZ    2008-05-01        7065474       5228909 1444374       0    3016   36596
##  5 CA    2008-05-01         290992      15287108 6462880 1381362  151812 3612505
##  6 CO    2008-05-01        5515148       1736728  464060  591780    4270  -71166
##  7 CT    2008-05-01         661754       1458402   93087       0       0  316081
##  8 DC    2008-05-01              0             0       0       0       0       0
##  9 DE    2008-05-01         625445         67052       0       0       0   52859
## 10 FL    2008-05-01       11510032      18931769   40776       0       0 3417149
## # ... with 8,414 more rows, and abbreviated variable names
## #   1: Natural_Gas_Gen_Total, 2: Hydro_Gen_Total, 3: Wind_Gen_Total,
```

```
## #   4: Solar_Gen_Total, 5: Other_Gen_Total_new
```

The main data set is relatively clean, there are a few data points that probably need to be removed. For instance in the skim there is negative megawatt hour generation value in the P0 col which does not make sense. It seems the negative values are related to "Pumped Storage". This will need further investigation.

The start of formatting the price data, sourced from various other data sources then the original energy data set. Pulling in the various csv files in their own formats and making adjustments to a common format with just the data of interest.

One of the price data sets does not have the monthly fidelity that the overall energy generation data set has, additionally price may be more dependent on region for coal plants which is not feasible to transport in many cases. There is an unresolved issue of relating the price to the generation. Either the generation will have to be summarized to the year or the yearly price data extrapolated. Another possible data set for us to consider is https://data.nasdaq.com/data/EIA/COAL-us-coal-prices-by-region but will require come way to connect the location in the main file to the region in that potential dataset.

At one point with only limited number of rows the original coal price csv was easier to edit by removing the other regions(non-us) than through R parsing.

# Discussion

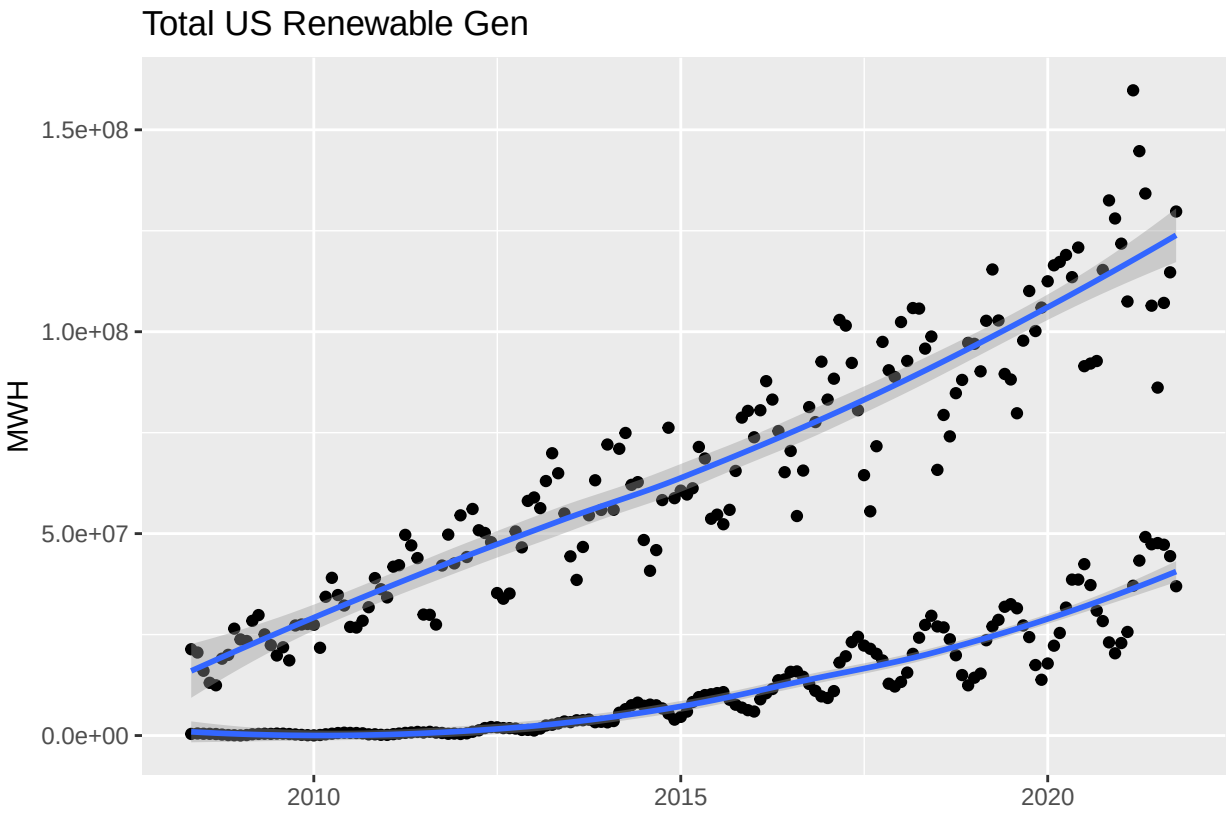
### Does the cost of fuel influence the energy production rates?

### How has the reduction in renewables cost affected generation numbers?

```
totalUSSrenewable = allTableWhole %>%
  select(STATE,date_col,Solar_Gen_Total, Wind_Gen_Total) %>%
  group_by(date_col) %>%
  summarise(
    US_Solar_Gen = sum(Solar_Gen_Total),
    US_Wind_Gen = sum(Wind_Gen_Total))

ggplot(data = totalUSSrenewable) +
  geom_point(mapping = aes(x = date_col, y = US_Solar_Gen)) +
  geom_smooth(mapping = aes(x = date_col, y = US_Solar_Gen)) +
  geom_point(mapping = aes(x = date_col, y = US_Wind_Gen)) +
  geom_smooth(mapping = aes(x = date_col, y = US_Wind_Gen)) +
  labs(title = "Total US Renewable Gen",
       x = "",
       y = "MWH")
```

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




```
renewableCost = read_csv("windSolarCost.csv")
```

```
## Rows: 19 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): gen
## dbl (2): year, kWcost
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
print(renewableCost)
```

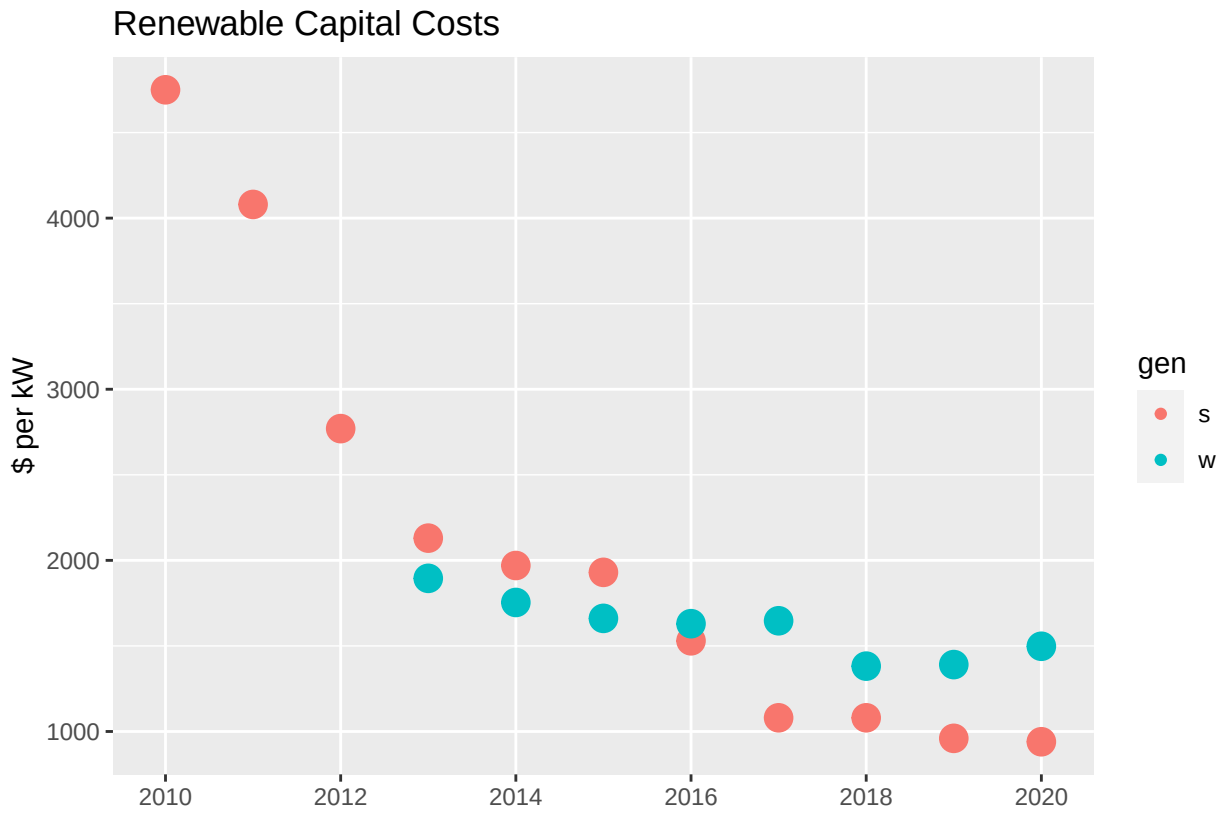
```
## # A tibble: 19 x 3
##     year kWcost gen
##    <dbl>  <dbl> <chr>
##  1  2010   4750 s
##  2  2011   4080 s
##  3  2012   2770 s
##  4  2013   2130 s
##  5  2014   1970 s
##  6  2015   1930 s
##  7  2016   1530 s
##  8  2017   1080 s
##  9  2018   1080 s
## 10  2019    960 s
## 11  2020    940 s
```

```
## 12  2013   1895 w
## 13  2014   1754 w
## 14  2015   1661 w
## 15  2016   1630 w
## 16  2017   1647 w
## 17  2018   1382 w
## 18  2019   1391 w
## 19  2020   1498 w
```

```
ggplot(renewableCost) +
  geom_point(aes(x = year,y = kWcost, color = gen, size=2))+
  labs(title = "Renewable Capital Costs",
       x = "",
       y = "$ per kW")+
  scale_x_discrete(limits = c(2010, 2012, 2014, 2016, 2018, 2020))+
  guides(size=FALSE)
```

```
## Warning: Continuous limits supplied to discrete scale.
## Did you mean `limits = factor(...)` or `scale_*_continuous()`?
```

```
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides(<scale> =
## "none")` instead.
```




Plots of Wind and solar costs data with overlay of production on same plot

Linear model of cost to generation

Table of data fit

Most of the dramatic drop in wind renewable capital costs occurred in the early 2000's.

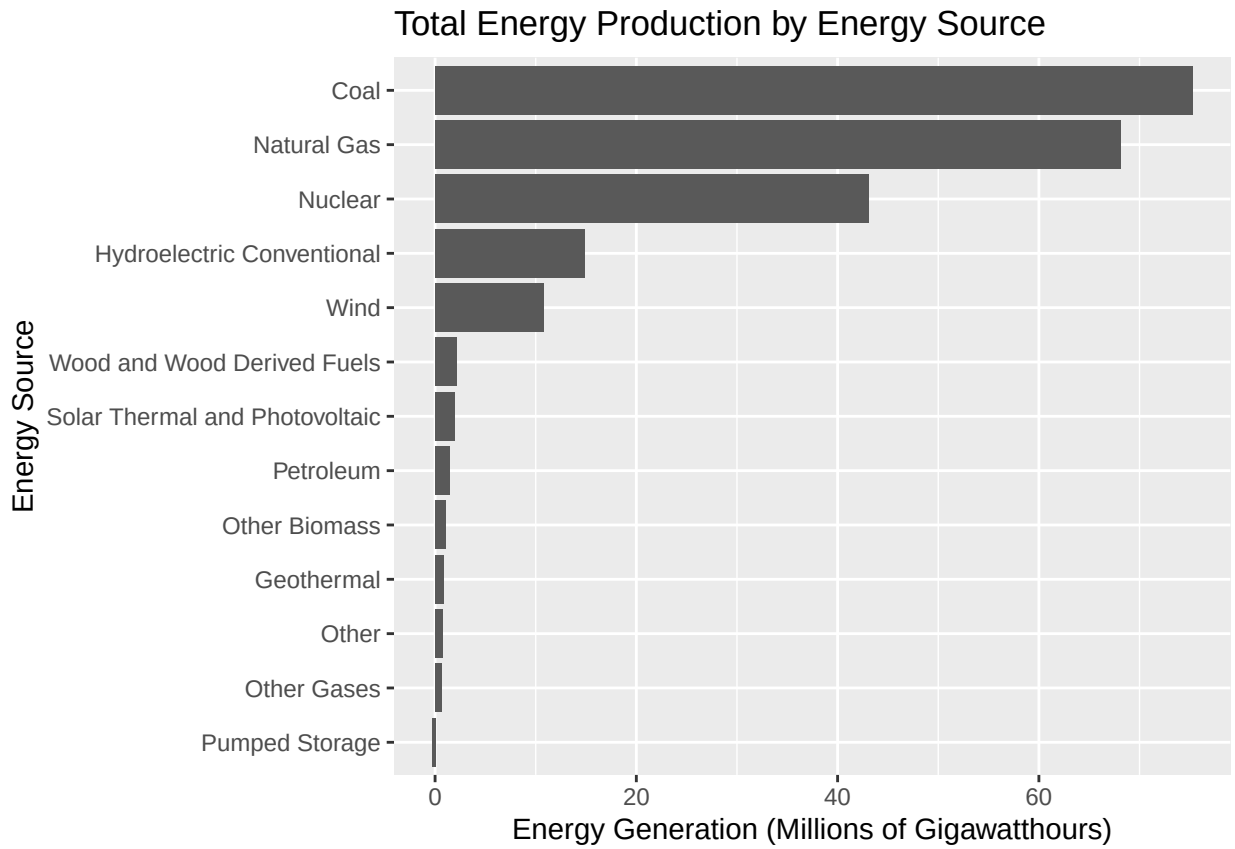
### Which energy source produces the most energy?

If you look at it simply by total amount generated is clear the number one energy source from 2008-2021 in the entire United States is Coal followed Natural Gas and Nuclear. Then the other types of energy sources drops of drastically. Note that the data shown in this bar graph is in the millions of Gigawatt hours because of the massive amounts of energy produce over the time period.

```
##Start with bar chart showing overall energy production
##clean up column names
energyWholeClean <- energyWhole
names(energyWholeClean)[names(energyWholeClean) == 'ENERGY SOURCE'] <- 'EnergySource'
names(energyWholeClean)[names(energyWholeClean) == 'GENERATION (Megawatthours)'] <- 'Generation'
energyWholeClean <- energyWholeClean %>%
  mutate(GenerationMilGig = Generation/1000000000) %>%
  mutate(GenerationGig = Generation/1000000)
##energyWholeClean
##Sum up enegry Generation by source in millions of Gig
totalBySourceGig <- energyWholeClean %>% group_by(EnergySource) %>%
  summarize(totalProductionGig = sum(GenerationMilGig), totalProduction = sum(Generation)) %>%
  filter(!EnergySource=="Total") ## remove Total category
##totalBySourceGig
##Sum up enegry Generation by source in gigawatt
totalBySource <- energyWholeClean %>% group_by(EnergySource,STATE, YEAR) %>%
  summarize(totalProduction = sum(GenerationGig)) %>%
  filter(!EnergySource=="Total") ## remove Total category
```

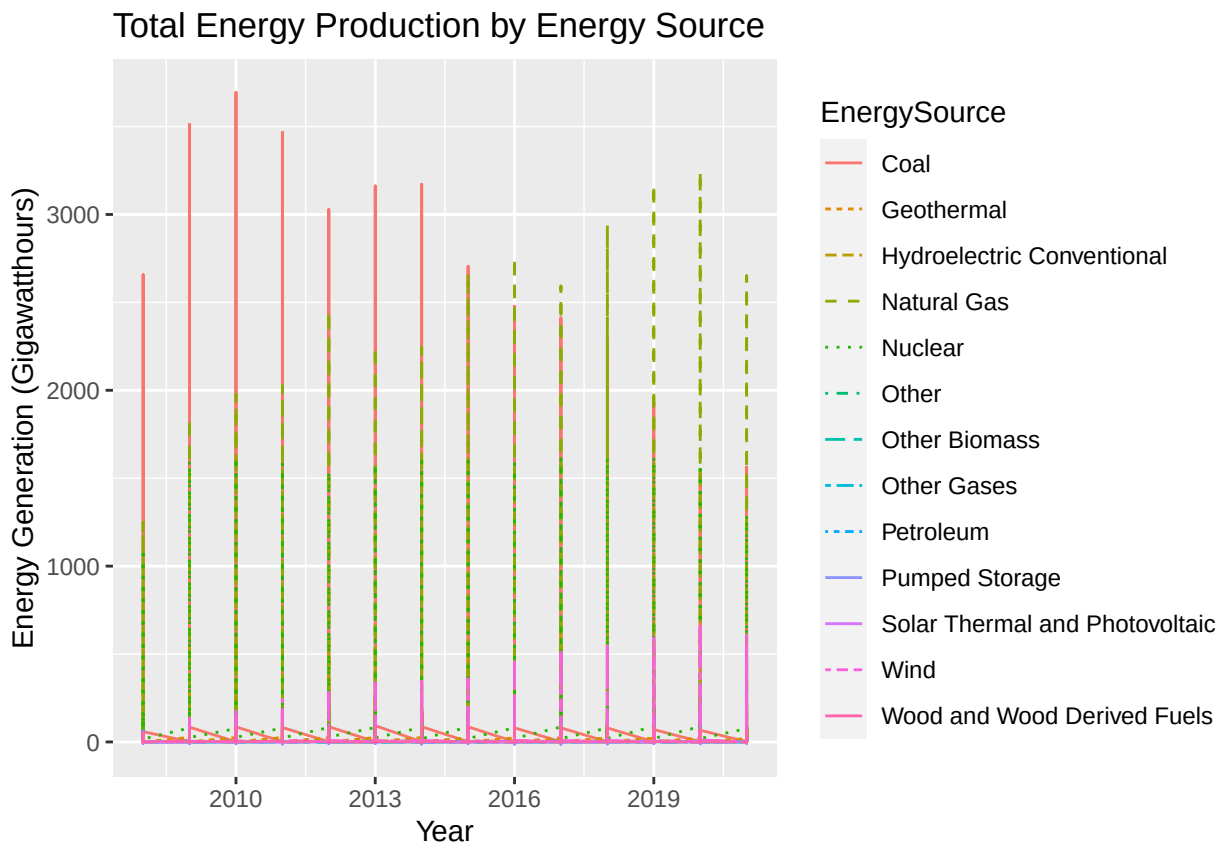
```
## `summarise()` has grouped output by 'EnergySource', 'STATE'. You can override
## using the `.groups` argument.
```

```
##totalBySource
## create bar graph for total production
ggplot(data = totalBySourceGig) +
  geom_bar(aes(y = reorder(EnergySource, totalProductionGig), x = totalProductionGig), stat = "identity
  labs(title = "Total Energy Production by Energy Source",
       x = "Energy Generation (Millions of Gigawatthours)",
       y = "Energy Source")
```




Looking at the data simply by total doesn't tell the whole story though. When we look at using a line graph, we find drastic changes in a few of the energy sources. Clearly there is a shift to move away from using coal as an energy source with the total production almost cuts in half since 2008. The exact opposite can be said for Natural Gas with energy production nearly doubling. Clearly the country is making efforts to protect the environment by moving from Coal to Natural Gas for its energy. It is also worth noting that Nuclear energy production has remained level over the entire time period. There are reasons preventing the US from moving to more nuclear production. The only other real item on note is the gradual increase in production of Wind energy. Yet more evidence the US is looking to increase cleaner energy. ##something happened and this chart looks goofy now - figure out why

```
##show data by year
ggplot(data = totalBySource) +
  geom_line(aes(y = totalProduction,x = YEAR, group = EnergySource, color = EnergySource, linetype = En
  labs(title = "Total Energy Production by Energy Source",
       y = "Energy Generation (Gigawatthours)",
       x = "Year")
```




### What is the average yearly production of each energy source by state?

##STILL WORKING on this section

```
library(rlang)
```

```
##
## Attaching package: 'rlang'
```

```
## The following objects are masked from 'package:purrr':
##
##     %@%, as_function, flatten, flatten_chr, flatten_dbl, flatten_int,
##     flatten_lgl, flatten_raw, invoke, splice
```

```
## create tibble for to get data for 5 key energy sources
avgProdbySourcebyState <- energyWholeClean %>%
##  filter(YEAR=='2008', !EnergySource=='Total', !STATE=='US-TOTAL') %>%
  filter(EnergySource=='Coal' | EnergySource=='Natural Gas'| EnergySource=='Nuclear'| EnergySource=='Hy
  group_by(EnergySource,STATE, YEAR) %>%
  summarise(avgGigHours = mean(GenerationGig))%>%
  filter(avgGigHours >0)
```

```
## `summarise()` has grouped output by 'EnergySource', 'STATE'. You can override
## using the `.groups` argument.
```

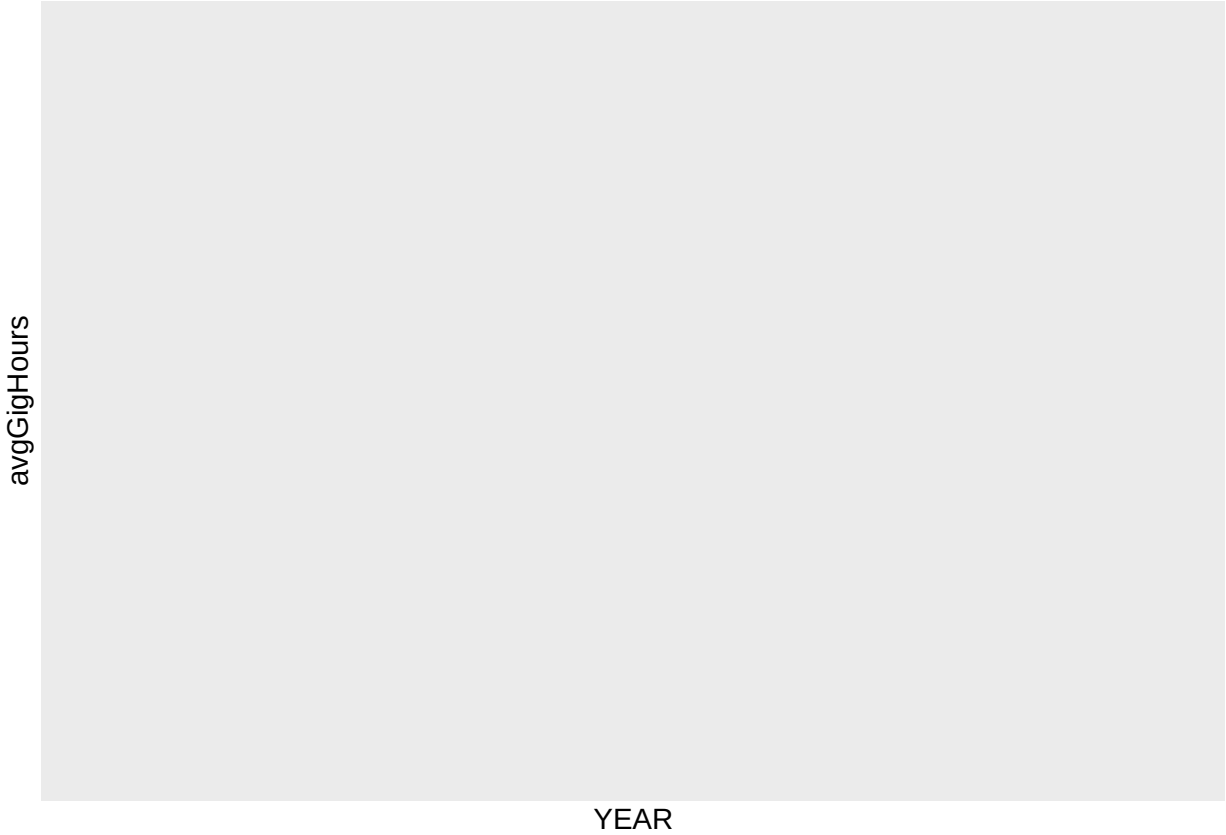
```
avgProdbySourcebyState
```

```
## # A tibble: 3,005 x 4
## # Groups:   EnergySource, STATE [219]
```

```
##    EnergySource STATE  YEAR avgGigHours
##    <chr>        <chr> <dbl>       <dbl>
##  1 Coal         AK     2008      0.0255
##  2 Coal         AK     2009      0.0263
##  3 Coal         AK     2010      0.0259
##  4 Coal         AK     2011      0.0273
##  5 Coal         AK     2012      0.0285
##  6 Coal         AK     2013      0.0260
##  7 Coal         AK     2014      0.0235
##  8 Coal         AK     2015      0.0278
##  9 Coal         AK     2016      0.0248
## 10 Coal         AK     2017      0.0232
## # ... with 2,995 more rows
```

```
##get list of unique states
variables <- avgProdbySourcebyState %>%
 group_by(STATE) %>%
  distinct(STATE)
##for each var (STATE) create a line chart showing avg Production by year
for (var in variables) {
print(var)
##get data for the STATE being looped on
stateAvgProd <-  avgProdbySourcebyState %>%
  filter(STATE=='!!sym(var)')
##Create a chart for state showing average production by energy source
plot <- ggplot(data = stateAvgProd, aes(x = YEAR, y = avgGigHours, group = EnergySource, color = Energy
    geom_line() +
    geom_point()
print(plot)
}
```

```
##  [1] "AK" "AL" "AR" "AZ" "CA" "CO" "CT" "DE" "FL" "GA" "HI" "IA" "ID" "IL" "IN"
## [16] "KS" "KY" "LA" "MA" "MD" "ME" "MI" "MN" "MO" "MS" "MT" "NC" "ND" "NE" "NH"
## [31] "NJ" "NM" "NV" "NY" "OH" "OK" "OR" "PA" "SC" "SD" "TN" "TX" "UT" "VA" "WA"
## [46] "WI" "WV" "WY" "RI" "VT" "DC"
```




**Which state produced the most of each energy type in the last 10 years?**

**Looking at 2011-2021, from end date above**

**NOT DONE**

```
library(rlang)

MostEnergyProdbySourcebyState <- energyWholeClean %>% filter(YEAR>='2011', YEAR<= '2021') %>% filter(EnergySource=='Coal' | EnergySource=='Natural Gas'| EnergySource=='Nuclear'| EnergySource=='Hydroelectric Conventional'| EnergySource=='Wind', !STATE=='US-TOTAL') %>% group_by(EnergySource,STATE, YEAR) %>% summarize(TotalProduction = sum(GenerationGig))%>% filter(TotalProduction >0) %>% arrange(TotalProduction) MostEnergyProdbySourcebyState
```

# Conclusions

# Appendix 1 - Data Source

https://www.kaggle.com/datasets/kevinmorgado/us-energy-generation-2001-2022

https://www.kaggle.com/code/bahadirumutiscimen/major-commodity-prices-data-analysis -> for Oil and Natural Gas

https://www.kaggle.com/datasets/timmofeyy/-metals-price-changes-within-last-30-years -> for Uranium

https://ourworldindata.org/grapher/coal-prices?country=~US+Central+Appalachian+coal+spot+price+index+%28BP%29 -> for Coal (may need a different set)

https://github.com/mtmccullough5/The_Energy_Generators/edit/main/README.md -> Project Repository

https://www.nrel.gov/docs/fy21osti/77324.pdf

https://www.eia.gov/electricity/generatorcosts/

https://www.nrel.gov/docs/fy12osti/54526.pdf https://apps.automeris.io/wpd/