Energy Generation Analysis Report

The Energy Generators

2022-11-10

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

- Does the cost of fuel influence the energy production rates?
- How has the reduction in renewables cost affected generation numbers?
- Which energy source produces the most energy?
- What is the average yearly production of each energy source by state?
- Which state produced the most of each energy type in the last 10 years?

Data Collection and Arrangement

```
energy = read_csv("organised_Gen.csv")
```

```
## 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")
```

```
## 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")
```

```
## 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")
```

```
## 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")
```

```
## Rows: 745 Columns: 6
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl  (5): Central Appalachia 12,500 Btu, 1.2 SO2, Northern Apppalachia 13,00...
## 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.
```

```
print(energy)
```

```
## # A tibble: 496,774 x 7
##     ...1  YEAR MONTH STATE `TYPE OF PRODUCER`                    ENERG~1 GENER~2
##    <dbl> <dbl> <dbl> <chr> <chr>                                 <chr>     <dbl>
##  1     0  2001     1 AK    Total Electric Power Industry         Coal      46903
##  2     1  2001     1 AK    Total Electric Power Industry         Petrol~   71085
##  3     2  2001     1 AK    Total Electric Power Industry         Natura~  367521
##  4     3  2001     1 AK    Total Electric Power Industry         Hydroe~  104549
##  5     4  2001     1 AK    Total Electric Power Industry         Wind         87
##  6     5  2001     1 AK    Total Electric Power Industry         Total    590145
##  7     6  2001     1 AK    Electric Generators, Electric Utilit~ Coal      18410
##  8     7  2001     1 AK    Electric Generators, Electric Utilit~ Petrol~   64883
##  9     8  2001     1 AK    Electric Generators, Electric Utilit~ Natura~  305277
```

```
## 10     9  2001     1 AK    Electric Generators, Electric Utilit~ Hydroe~  104549
## # ... with 496,764 more rows, and abbreviated variable names
## #   1: `ENERGY SOURCE`, 2: `GENERATION (Megawatthours)`
```

```
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
```

```
print(og)
```

```
## # A tibble: 5,785 x 3
##    Date       Price_oil Price_natural_gas
##    <date>         <dbl>             <dbl>
##  1 2000-01-04      24.4              2.18
##  2 2000-01-05      23.7              2.17
##  3 2000-01-06      23.6              2.20
##  4 2000-01-07      23.1              2.17
##  5 2000-01-10      23.7              2.22
##  6 2000-01-11      24.6              2.26
##  7 2000-01-12      24.8              2.24
##  8 2000-01-13      25.0              2.25
##  9 2000-01-14      25.5              2.32
## 10 2000-01-17      25.6             NA
## # ... with 5,775 more rows
```

```
print(uranium)
```

```
## # A tibble: 265 x 3
##     Year Month Price_uran
##    <dbl> <chr>      <dbl>
##  1  2000 Jan         9.55
##  2  2000 Feb         9.39
##  3  2000 Mar         9.31
##  4  2000 Apr         9.14
##  5  2000 May         8.5
##  6  2000 Jun         8.23
##  7  2000 Jul         8.1
##  8  2000 Aug         7.96
##  9  2000 Sep         7.57
## 10  2000 Oct         7.4
## # ... with 255 more rows
```

```
print(coal)
```

```
## # A tibble: 745 x 6
##    `Week Ended` Central Appalachia 12,500 Btu,~1 North~2 Illin~3 Powde~4 Uinta~5
##    <date>                                  <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
##  1 2022-10-28                               182.    186.    186.    17.2    41.8
##  2 2022-10-21                               177.    180.    186.    17.4    42.0
##  3 2022-10-14                               177.    180.    186.    17.4    42.0
##  4 2022-10-07                               205.    188.    197.    16.2    42.4
##  5 2022-09-30                               205.    188.    197.    16.2    42.4
##  6 2022-09-23                               199.    182.    197.    16.2    42.6
##  7 2022-09-16                               199.    182.    197.    16.2    42.6
##  8 2022-09-09                               193.    176.    197.    16.2    42.7
##  9 2022-09-02                               193.    176.    197.    16.2    42.7
## 10 2022-08-26                               187.    146.    190.    16.2    42.8
## # ... with 735 more rows, and abbreviated variable names
## #   1: `Central Appalachia 12,500 Btu, 1.2 SO2`,
## #   2: `Northern Apppalachia 13,000 Btu, <3.0 SO2`,
## #   3: `Illinois Basin 11,800 Btu, 5.0 SO2`,
## #   4: `Powder River Basin 8,800 Btu, 0.8 SO2`,
## #   5: `Uinta Basin 11,700 Btu, 0.8 SO2`
```

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?

Does the cost of fuel influence the energy production rates?

How has the reduction in renewables cost affected generation numbers?

Which energy source produces the most energy?

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

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

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