Power_Plant_Energy

Kyle Ruhl

5/1/2022

This is a sample R script assignment for exploring the PowerPlant dataset.

The dataset, from the paper "The environmental footprint of data centers in the United States" by Md Abu Bakar Siddik et al 2021, explores power plants across the lower 48 states + DC and their power type (coal, biomass, solar, etc), the CO2 Emissions associated with producing the electricity, the water and carbon intensity per MWh of power produced, and the water consumption of the plant.

Before performing the data analysis, for a quick review of renewable vs fossil fuel energy sources, read through this webpage from the US Energy Information Administration: https://www.eia.gov/energyexplained/renewable-sources/ (https://www.eia.gov/energyexplained/renewable-sources/)

Lets start by importing and cleaning the data set. After downloading the csv, run the cell block below.

```r
r library(tidyverse)
```

```
## — Attaching packages ———————————————————— tidyverse 1.3.1 —
```

```
## ✓ ggplot2 3.3.5     ✓ purrr   0.3.4 ## ✓ tibble  3.1.6     ✓ dplyr   1.0.8 ## ✓ tidyr   1.2.0     ✓ stringr 1.4.0 ## ✓ readr   2.1
```

```
## — Conflicts ———————————————————————— tidyverse_conflicts() — ## x dplyr::filter() masks stats::filter() ## x d
```

```r
```r PowerPlants <- read.csv("PowerPlants.csv", header = TRUE)
```

#Keep variables of interest: PowerPlants <- PowerPlants %>% select(c(Plant.state, Plant.name, Latitude, Longitude, Plant.primary.fuel.code, Net.generation..MW Water.intensity..m3.MWh., Carbon.intensity..Tons.MWh., Water.Consumption..m3.))

#Change variable names: names(PowerPlants) <- c("Plant.State", "Plant.Name", "Latitude", "Longitude", "Power.Type", "Net.Power.Generated", "CO2.Emissions" "Water.Consumption")

#Factor states and power type into levels: PowerPlants$Plant.State <- factor(PowerPlants$Plant.State) PowerPlants$Power.Type <- factor(PowerPlants$P

Task 1:

Lets explore the breakdown of power type in your home state (or favorite state if you are international!)

1.1: Subset the dataframe so that you are only looking at your chosen state's data. Name this dataframe State1. Output the head of this new dataframe.

```r
#Sample Answer
State1 <- PowerPlants %>% filter(PowerPlants$Plant.State == "NY")
head(State1)
```

```
##   Plant.State                       Plant.Name Latitude Longitude Power.Type
## 1          NY                       23rd and 3rd 40.66310 -74.00000        GAS
## 2          NY                       59th Street 40.77128 -73.99115        OIL
## 3          NY                       74th Street 40.76800 -73.95150        OIL
## 4          NY                          Al Turi 41.40432 -74.37671    BIOMASS
## 5          NY      Albany Medical Ctr Cogen Plant 42.65500 -73.77770        GAS
## 6          NY Albert Einstein College of Medicine 40.85132 -73.84830        OIL
##   Net.Power.Generated CO2.Emissions Water.Intensity Carbon.Intensity
## 1              117563         77664            0.80             0.66
## 2                 199           182            2.06             0.91
## 3                 344           377            2.06             1.09
## 4                5126             0            1.94             0.00
## 5               32054         10882            0.80             0.34
## 6                  47            36            2.06             0.77
##   Water.Consumption
## 1             93701
## 2               410
## 3               710
## 4              9932
## 5             25548
## 6                97
```

1.2 Now, lets examine the Power.Type variable in your state. Output a table of the # of power sites using each type of power in your state and save it as table1.

```r
State1$Power.Type = factor(State1$Power.Type) #refactor to remove power sources absent in state

#Sample Answer
table1 <- table(State1$Power.Type)
table1
```

```
## 
## BIOMASS    COAL     GAS  HYDRO NUCLEAR     OIL   SOLAR    WIND
##      32       2      88     158       5      50      96      26
```

1.3 Visualize this table by creating an appropartite plot with correct labels. Next, Describe the plot- what is the most popular power type in your state? Based off your knowledge of your state, is this distribution what you expected?

```
#Sample Answer
barplot(table1, las=2, xlab = 'Type of Power', ylab = '# Of Plants',
        main = 'Power Plant Sites by Power Type in NY')
```

Power Plant Sites by Power Type in NY
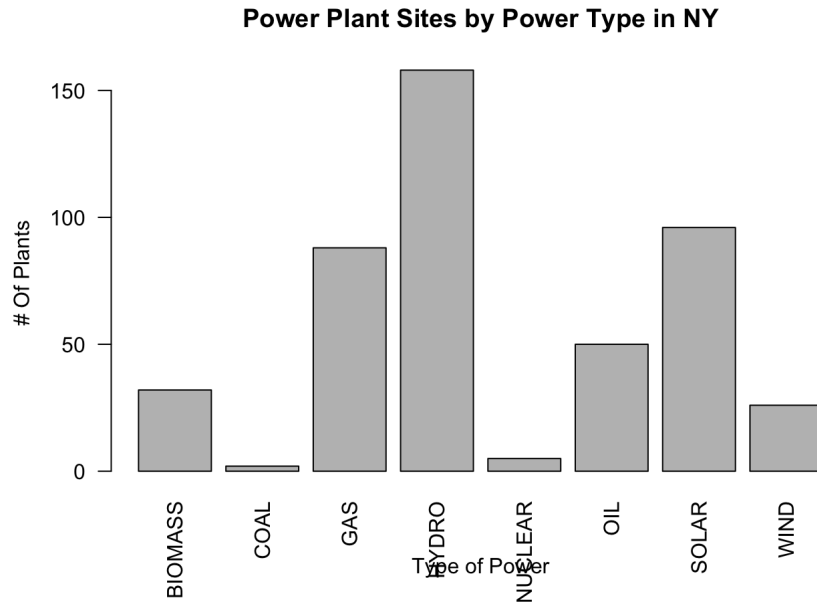
- ANSWER HERE: (Sample Answer- In NY, Hydroelectric power is the most popular power type with 158 sites. This distribution makes sense to me considering the lakes and rivers throughout NY, including Niagra falls, which gives opportunity for many hydro sites. I was suprised that there are still so many oil and natural gas sites in my state.)

1.4 Now, lets look at the net power generated by each power type. Use the group_by function to find the sum of the net power generated by each power type in your chosen state, and save this as State1Type. Output the head of this dataframe.

```
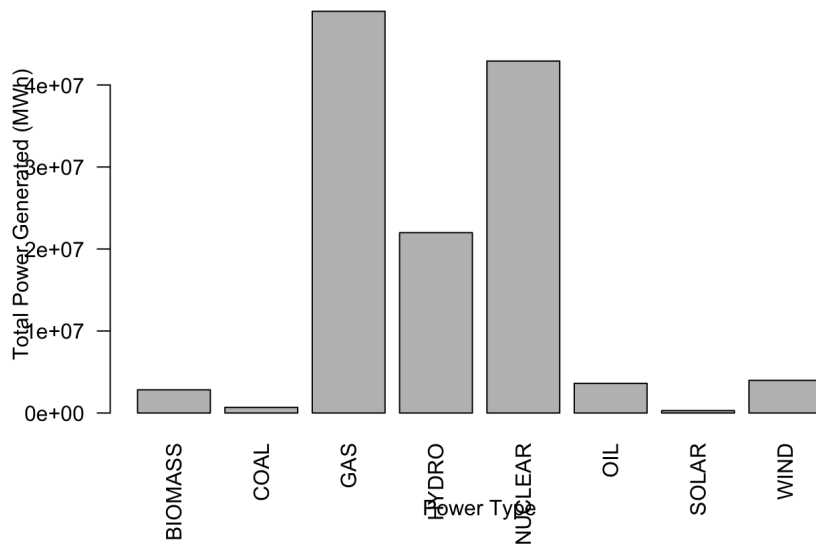#Sample Answer
State1Type <- State1 %>% group_by(Power.Type) %>% summarize(Net.Power.Generated = sum(Net.Power.Generated))
head(State1Type)
```

```
## # A tibble: 6 × 2
##   Power.Type Net.Power.Generated
##   <fct>                    <int>
## 1 BIOMASS                2829702
## 2 COAL                    688346
## 3 GAS                   48986406
## 4 HYDRO                 21990447
## 5 NUCLEAR               42919011
## 6 OIL                    3614776
```

1.5 Plot the resulting dataframe State1Type in an approparite visualization to show the sum of the net power generated by each site type in your chosen state. Then, answer the following question: Which site types generated the most power? Did this result match your expectations from 1.3?

```
#Sample Answer
barplot(State1Type$Net.Power.Generated, las=2, names.arg = State1Type$Power.Type,
        xlab = 'Power Type', ylab = 'Total Power Generated (MWh)', main = 'Net Power Generated by site in NY')
```

## Net Power Generated by site in NY



- Answer Q here: (Sample answer- Gas and Nuclear have the greatest total power generated in NY. From 1.3 I expected that hydro would have produced the most power, but it came in third. Suprisingly, nuclear, which only had 5 sites in NY produced almost 43 million MWh of power, the second most of the site types.)

1.6 Now we will assess the average power produced by each plant type. Calculate this by taking the net power generated (from 1.5) / by the number of sites in your state (from 1.2). Save this list as the variable By.Plant and plot it using an approparite visualiztion.

Then, Answer the questions: a) When averaging by the number of sites, did the distribution change for your state? b) Which power types produce the most energy per plant? c) Compare the energy produced per site for renewable energy sources vs fossil fuel sources. What does this tell us about the number of sites needed for renewable energy to match fossil fuels?

```
#Sample answer
By.Plant <- (State1Type$Net.Power.Generated / table1)
barplot(By.Plant, las=2, xlab = "Power Type", ylab = 'Total Power Generated (MWh)',
        main = "Average Power Generated By Site in NY")
```

Average Power Generated By Site in NY
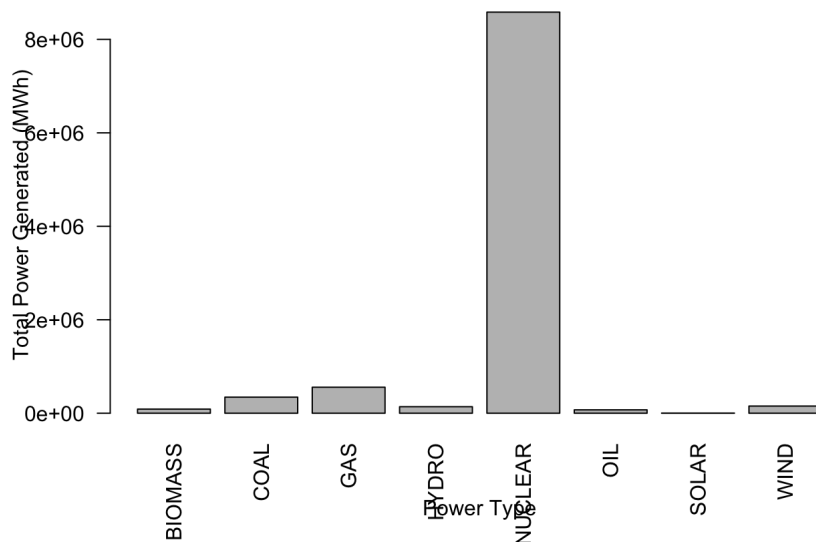
Q Answers: -a) (sample answer- Yes, when we averaged power generated by the number of sites in NY, we see that nuclear energy sites produce the most power per site, nearly x5 that of a natural gas site) -b) (sample answer- Nuclear and Natural Gas) -c) (sample answer- the power generated from solar and wind per site is minimal compared to natural gas power generated per site, meaning that more solar and wind sites would need to be constructed to match the current output of fossil fuel sites. Other net-zero carbon emission energy sources like nuclear power produce the most energy per site, however many Americans are opposed to the idea of building nuclear sites, citing health and environmental concerns.)

Optional: Consider re-running part 1 with a different state. Particularly, if your original state had 0 or 1 nuclear power plants, consider choosing a state that has 2 or more - good choices are AL,FL,GA,IL,LA,MI,MN,NC,NJ,NY,OH,PA,SC,TN, or VA. If you originally chose one of these, choose a state not on the list and see how it compares.

#Part 2

Now, lets explore carbon emissions by power type.

2.1: Use group_by to create a new dataframe called Carbon.by.Type that groups the state and power type by the mean of the carbon emissions. Then, create a stacked barplot to visualize, with each state having a bar on one axis, mean of state's carbon emissions on the other axis, and each bar filled in by power type.

For reference, the Carbon.by.Type data frame should look like this 1 AL BIOMASS 1.987808e+04 2 AL COAL 6.967619e+06 3 AL GAS 8.389327e+05 4 AL HYDRO 0.000000e+00 5 AL NUCLEAR 0.000000e+00 6 AL OIL 7.500000e+02 7 AL SOLAR 0.000000e+00 8 AR BIOMASS 1.512933e+04 9 AR COAL 6.573748e+06 … And so on

```
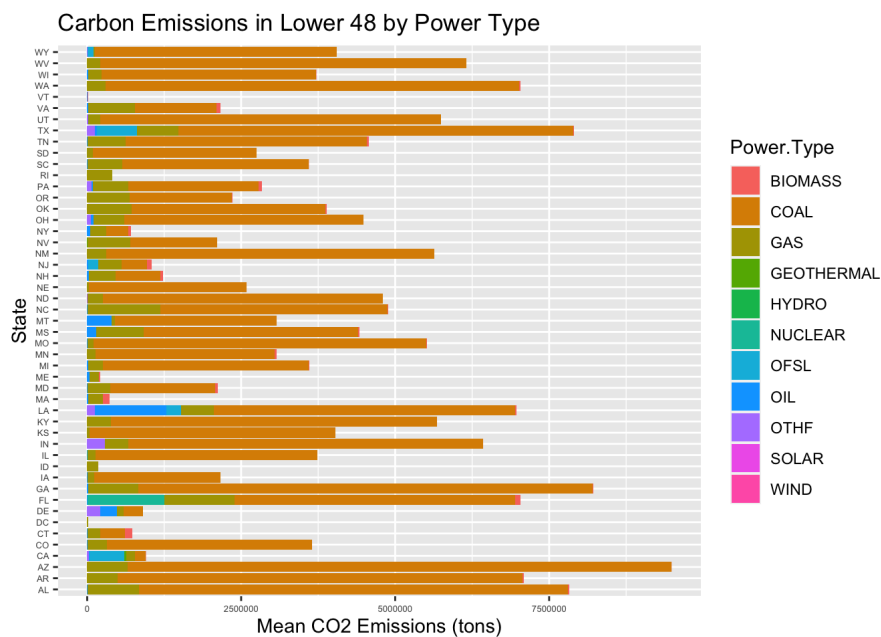#Sample answer
Carbon.by.Type <- PowerPlants %>% group_by(Plant.State, Power.Type) %>% summarize(CO2.Emissions = mean(CO2.Emissi
ons))
```

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

```
ggplot(data=Carbon.by.Type, aes(fill=Power.Type, y=CO2.Emissions, x=Plant.State)) +
    geom_bar(position="stack", stat="identity")  + coord_flip() + theme(axis.text=element_text(size=5)) +
    xlab('State') + ylab('Mean CO2 Emissions (tons)') + ggtitle('Carbon Emissions in Lower 48 by Power Type')
```


2.2 Use group_by to create a stacked barplot of the power plant distribution of the country, where each state is on one axis, the number of power plants on the other, and the fill is energy type.

```
#Sample Answer
EnergyMix <- PowerPlants %>% group_by(Plant.State, Power.Type)

ggplot(data=EnergyMix, aes(Plant.State, ..count..)) + geom_bar(aes(fill=Power.Type)) +
  coord_flip() + theme(axis.text=element_text(size=5)) +
  xlab('State') + ylab('Number of Plants') + ggtitle('Lower 48 Power Plant Type')
```

Lower 48 Power Plant Type
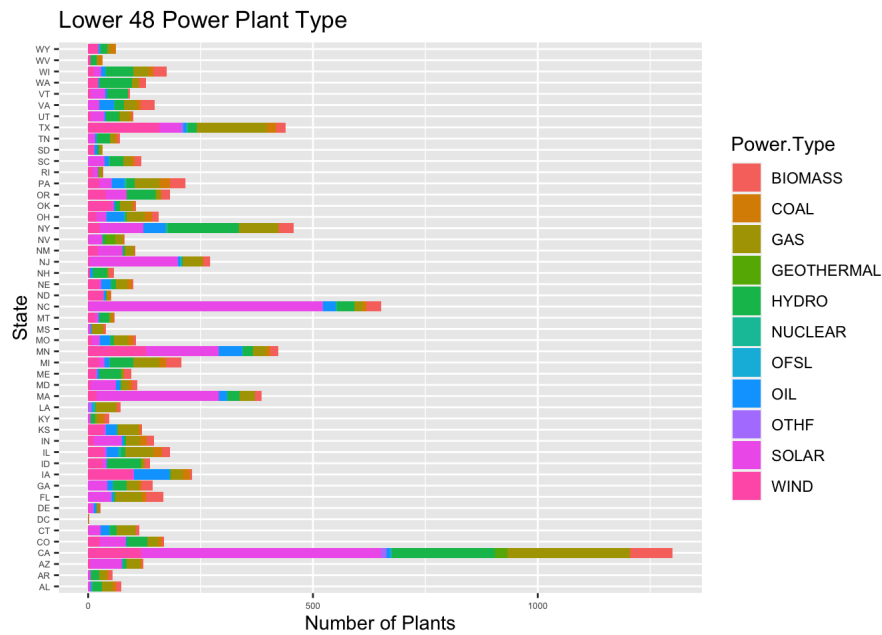
2.3 Based on the stacked barplots from 2.1 and 2.2, do you think carbon emissions from power generation are related to state size/population? Geography? Some other factor?

-Answer: (Sample Answer: Since large states (in both population and size) like CA and NY have smaller average CO2 emissions than many other larger and less populated states, it is likely that population doesn't have an influence on carbon emissions. A possible factor as to why some states have larger carbon emissions than others is their energy mix. States that rely on coal and natural gas have much larger total CO2 Emissions than states that do not. As we see from the barplot in 2.2, CA has >1000 power plants, almost all renewable energy, which is why their carbon footprint is much lower than other large states)