CM_Capstone_Final

Cynthia Mansbridge

August 30, 2018

Capstone Project: UK Energy Pricing, Temperature, and Energy Resources

Introduction

The goal of this project is to analyze data from the average day ahead energy prices and energy resources from across the UK. The average monthly UK temperature from the same time period will also be plotted to see if there is any correlation between temperature, price and resource mix.

Initial DataSets

The datasets utilized are from the OFGEM and UK government weather websites. All the data is downloadable, however, the Mean UK Temperature file was a flat file with tab delimination that required some additional cleansing.
There are three separate data sets that were put together: Mean UK Temperature, UK Day Ahead Price, and UK Resource Mix.

DataSet 1: **UKprice** *OFGEM* UK Monthly Average Day Ahead Energy Prices 1/2015 - 2/2018

For the UK Day Ahead Price [UKprice], I have collected the monthly Day Ahead Price information

```
##    Month Year Price
## 1      2 2018 53.68
## 2      1 2018 50.84
## 3     12 2017 56.18
## 4     11 2017 51.45
## 5     10 2017 46.68
## 6      9 2017 47.17
## 7      8 2017 43.34
## 8      7 2017 43.05
## 9      6 2017 39.68
## 10     5 2017 41.24
```

DataSet 2: **UKtemp** *UK National Weather* UK Mean Temperature 1/2015 - 2/2018

For the Mean UK Temperature [UKtemp], I have collected the Monthly Mean UK Temperature from 2015 - 2018.

```
##    Month Amount Year
## 1      1    3.7 2015
## 2      2    3.5 2015
## 3      3    5.5 2015
## 4      4    7.9 2015
## 5      5    9.6 2015
## 6      6   12.7 2015
## 7      7   14.4 2015
## 8      8   14.7 2015
## 9      9   11.9 2015
## 10    10   10.0 2015
```

DataSet 3: **UKmix** *OFGEM* UK Average Energy Resource Mix 1/2015 - 2/2018

For the UK Resource Mix [UKmix], I have the percentages by month and quarter from 2017 forward.

```
##    Year Quarter  Coal  Oil   Gas Nuclear Hydro  Wind  Bio PumpStorage
## 1  2006       1 47.56 2.06 29.98   20.52  1.11  1.09 2.37       -0.34
## 2  2007       1 37.02 1.22 42.82   14.12  1.94  1.65 2.26       -0.28
## 3  2008       1 33.42 1.03 47.61   13.14  1.97  2.22 2.17       -0.29
## 4  2009       1 37.81 1.88 37.27   15.40  1.56  2.60 2.52       -0.33
## 5  2010       1 30.98 1.29 47.38   16.52  0.82  2.45 2.72       -0.29
## 6  2011       1 33.36 0.81 37.66   17.67  1.30  3.35 2.92       -0.26
## 7  2012       1 40.00 0.63 26.72   15.62  1.82  5.22 3.49       -0.26
## 8  2013       1 39.27 0.50 26.71   16.61  1.24  7.04 3.62       -0.27
## 9  2014       1 32.78 0.49 21.39   15.01  2.23 11.55 3.94       -0.26
## 10 2015       1 28.03 0.47 23.17   16.51  2.02 12.80 6.02       -0.23
##    Other
## 1   0.87
## 2   0.76
## 3   0.81
## 4   0.78
## 5   0.60
## 6   0.66
## 7   0.68
## 8   0.81
## 9   0.98
## 10  1.11
```

Data Wrangling

Initially, before performing the graphical analysis, I did the following Data Cleansing. I looked at all three datasets and determined the best dataset structure to include all three datasets for analysis.

I imported each data set separately and made a master dataset from UKtemp. I added columns from UKprice and UKmix.

I then utilized R code to create a fourth DataSet which is UKenergy and is a combination of the three initial DataSets.

This is the R Code I used to create a combined dataset for graphical/statistical analysis.

```
#rename amount column in UKtemp to AvgTemp
names(UKtemp)[names(UKtemp) == "Amount"] <- "AvgTemp"

#add quarter column to UKPrice and update
UKprice <- mutate(UKprice,"Quarter" = "NA")

UKprice[UKprice$Month == 1,"Quarter"] <- 1
UKprice[UKprice$Month == 2,"Quarter"] <- 1
UKprice[UKprice$Month == 3,"Quarter"] <- 1
UKprice[UKprice$Month == 4,"Quarter"] <- 2
UKprice[UKprice$Month == 5,"Quarter"] <- 2
UKprice[UKprice$Month == 6,"Quarter"] <- 2
UKprice[UKprice$Month == 7,"Quarter"] <- 3
UKprice[UKprice$Month == 8,"Quarter"] <- 3
UKprice[UKprice$Month == 9,"Quarter"] <- 3
UKprice[UKprice$Month == 10,"Quarter"] <- 4
UKprice[UKprice$Month == 11,"Quarter"] <- 4
UKprice[UKprice$Month == 12,"Quarter"] <- 4

#add quarter column to UKtemp and update
UKtemp <- mutate(UKtemp,"Quarter" = 0)

UKtemp[UKtemp$Month == 1,"Quarter"] <- 1
UKtemp[UKtemp$Month == 2,"Quarter"] <- 1
UKtemp[UKtemp$Month == 3,"Quarter"] <- 1
UKtemp[UKtemp$Month == 4,"Quarter"] <- 2
UKtemp[UKtemp$Month == 5,"Quarter"] <- 2
UKtemp[UKtemp$Month == 6,"Quarter"] <- 2
UKtemp[UKtemp$Month == 7,"Quarter"] <- 3
UKtemp[UKtemp$Month == 8,"Quarter"] <- 3
UKtemp[UKtemp$Month == 9,"Quarter"] <- 3
UKtemp[UKtemp$Month == 10,"Quarter"] <- 4
UKtemp[UKtemp$Month == 11,"Quarter"] <- 4
UKtemp[UKtemp$Month == 12,"Quarter"] <- 4

#create new dataframe for final dataset
rm("UKenergy")
UKenergy <- UKtemp
View(UKenergy)

#Remove from UKenergy
UKenergy<-UKenergy[!(UKenergy$Month == "3" & UKenergy$Year==2018),]
UKenergy<-UKenergy[!(UKenergy$Month == "4" & UKenergy$Year==2018),]


#Add all fields from UKPrice and UKMix to master dataframe
UKenergy <- left_join(UKenergy, select(UKprice, Month, Year, Price))
UKenergy <- left_join(UKenergy, select(UKmix,Quarter,Year,Coal))
UKenergy <- left_join(UKenergy, select(UKmix,Quarter,Year,Oil))
UKenergy <- left_join(UKenergy, select(UKmix,Quarter,Year,Gas))
UKenergy <- left_join(UKenergy, select(UKmix,Quarter,Year,Nuclear))
UKenergy <- left_join(UKenergy, select(UKmix,Quarter,Year,Hydro))
UKenergy <- left_join(UKenergy, select(UKmix,Quarter,Year,Wind))
UKenergy <- left_join(UKenergy, select(UKmix,Quarter,Year,Bio))
UKenergy <- left_join(UKenergy, select(UKmix,Quarter,Year,PumpStorage))
UKenergy <- left_join(UKenergy, select(UKmix,Quarter,Year,Other))
```

```
#Export Master Dataframe
write.csv(UKenergy,'~/Downloads/UKenergy.csv')
```

The final dataset named UKenergy is structured as follows:

Field	Field Type
Month	Int
Year	Int
Quarter	Int
Price	Dbl
AvgTemp	Dbl
Coal	Dbl
Oil	Dbl
Gas	Dbl
Nuclear	Dbl
Hydro	Dbl
Wind	Dbl
Bio	Dbl
PumpStorage	Dbl
OtherResource	Dbl

This is a sample of the UKenergy dataset.

```
##   X Month AvgTemp Year Quarter Price  Coal  Oil   Gas Nuclear Hydro  Wind
## 1 1     1     3.7 2015       1 39.86 28.03 0.47 23.17   16.51  2.02 12.80
## 2 2     2     3.5 2015       1 42.86 28.03 0.47 23.17   16.51  2.02 12.80
## 3 3     3     5.5 2015       1 40.91 28.03 0.47 23.17   16.51  2.02 12.80
## 4 4     4     7.9 2015       2 43.97 15.11 0.33 22.95   15.37  1.44 11.45
##    Bio PumpStorage Other
## 1 6.02       -0.23  1.11
## 2 6.02       -0.23  1.11
## 3 6.02       -0.23  1.11
## 4 6.08       -0.23  1.07
```

Graphical Analysis and Further Data Wrangling

When I embarked on the graphical analysis, I found that the UKenergy data set was not in the format I needed for all graphs.

Therefore, I had to gather the data for the resource usage percentage into a different data set UKResGrp which is gathered by resource.

I also determined that 2018 data needed to be removed leaving only 2017 data. This allows for a clean graphical display that is easy to analyze.

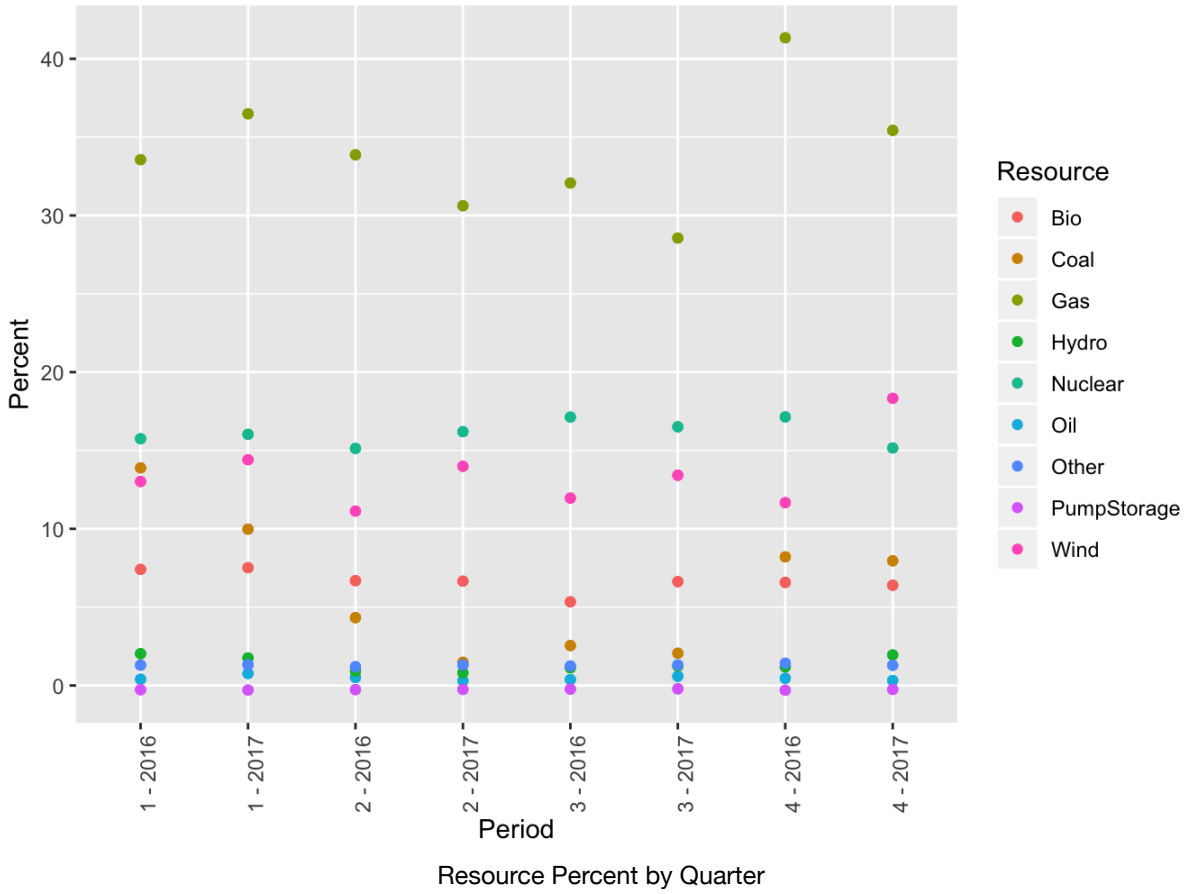
This is the R code used to create the additional dataframes for graphical analysis.

```
#remove UKenergy so that is is cleanly added
rm("UKenergy")

#load UKenergy so that it is fresh
UKenergy <- read.csv("~/Downloads/UKenergy.csv")

#add Period column to UKenergy and update
UKenergy <- mutate(UKenergy,"Period" = "NA")
UKenergy$Period <- paste(UKenergy$Quarter,'-',UKenergy$Year)
UKenergy$Period <- factor(UKenergy$Period, levels = UKenergy$Period)
UKenergy <- mutate(UKenergy, "MonthYear" = "NA")
UKenergy$MonthYear <- paste(UKenergy$Month,'-',UKenergy$Year)
UKenergy$MonthYear <- factor(UKenergy$MonthYear, levels = UKenergy$MonthYear)

#reorg by resource
UKResGRP <- gather(UKenergy,key = "Resource","Percent",Coal, Oil, Gas,Nuclear, Hydro,Wind,Bio,PumpS
torage, Other)

#Remove from UKResGRP where earlier than 2016
UKResGRP<-UKResGRP[!(UKResGRP$Year == 2015),]
UKResGRP<-UKResGRP[!(UKResGRP$Year == 2018),]

#Create new dataset with just 2017 data from UK Energy
UKRes2017<-UKenergy[(UKenergy$Year == 2017),]

#Create new dataset with just 2017 data from grouped dataframe
UKResGRP2017A<-UKResGRP[(UKResGRP$Year == 2017),]

#Export Master Dataframe
write.csv(UKResGRP,'~/Downloads/UKResGRP.csv')

write.csv(UKRes2017,'~/Downloads/UKRes2017.csv')

write.csv(UKResGRP2017A,'~/Downloads/UKResGRP2017A.csv')
```

Graphical Analysis

1. *Can you count something interesting?*

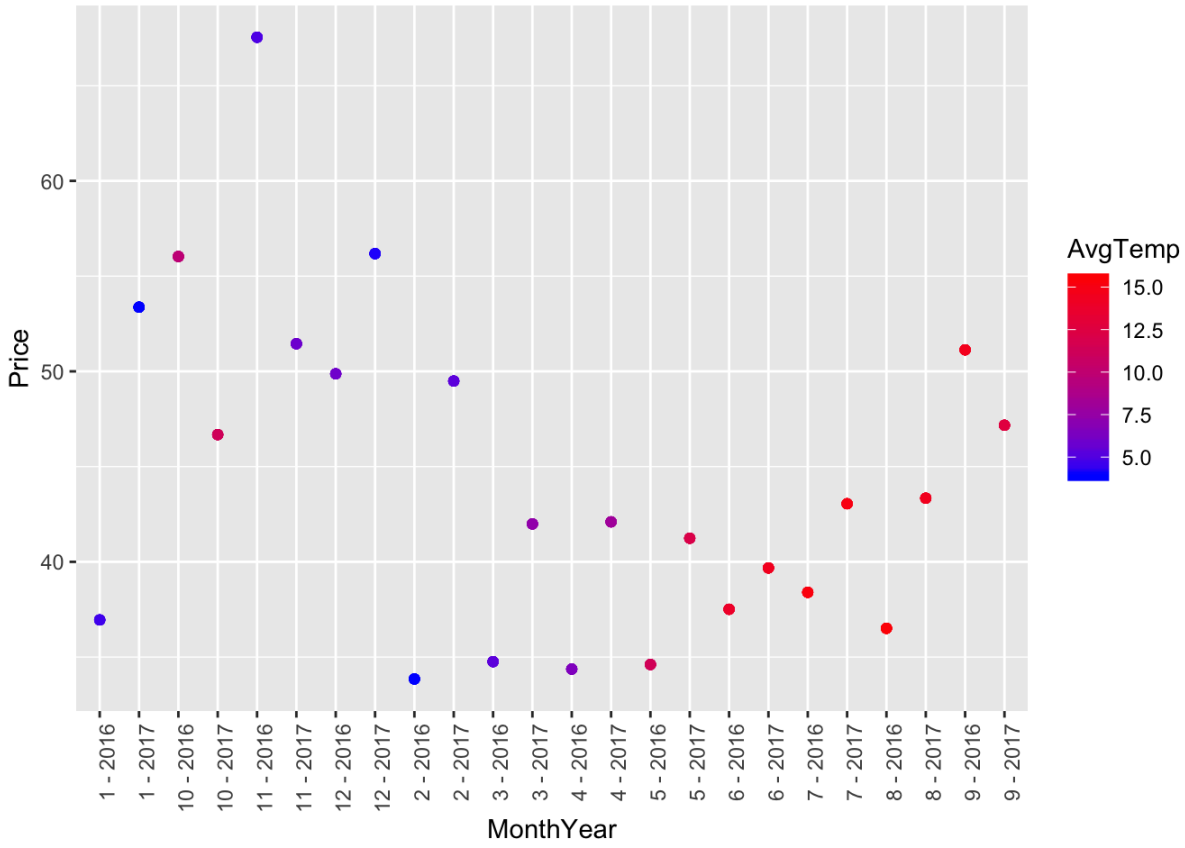
I looked at which energy resource had dominance per quarter. What I found was that the resource percentages didn't change drastically from quarter to quarter.
This was unexpected. I expected to see larger changes from one quarter to the next.



Resource Percent by Quarter

2. *Can you find some trends (high, low, increase, decrease, anomalies)?*

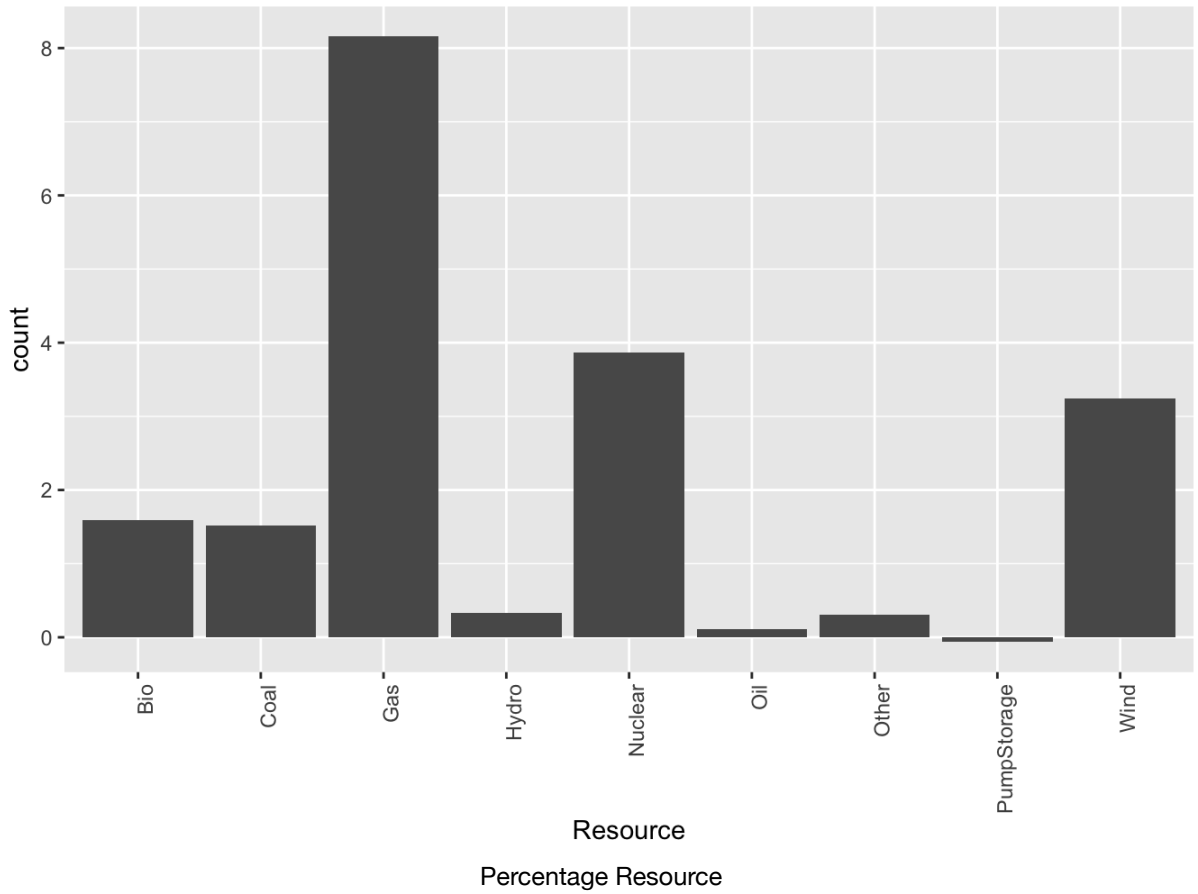
Interestingly, I found that the eight highest prices, six or 75% of them were in the lowest Average Temperature range. In the United States, many of the highest prices are in the warmer Average Temperature range. This shows that the UK is less Heat Weather sensitive than the US.



Price by Average Temperature

3. *Can you make a bar plot or a histogram?*

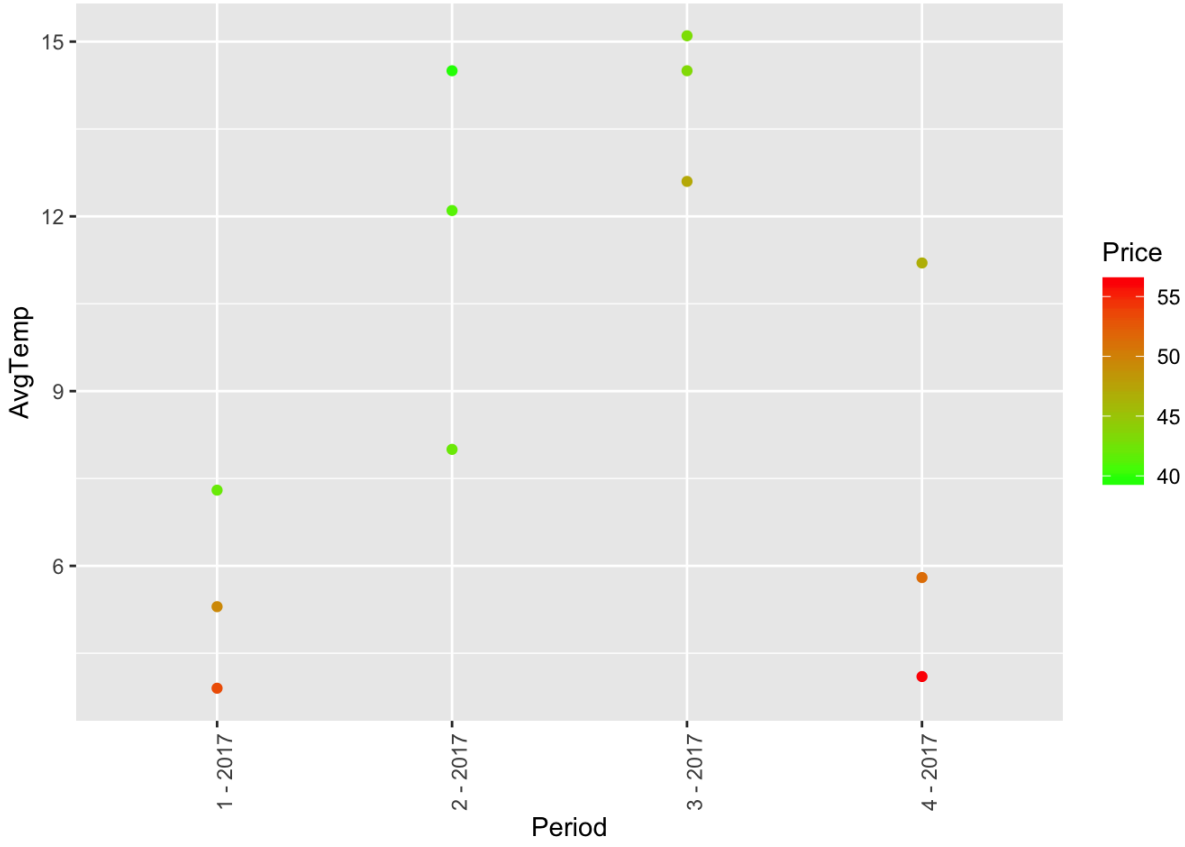
I plotted the total percent of each resource.



Percentage Resource

4. *Can you compare two related quantities?*

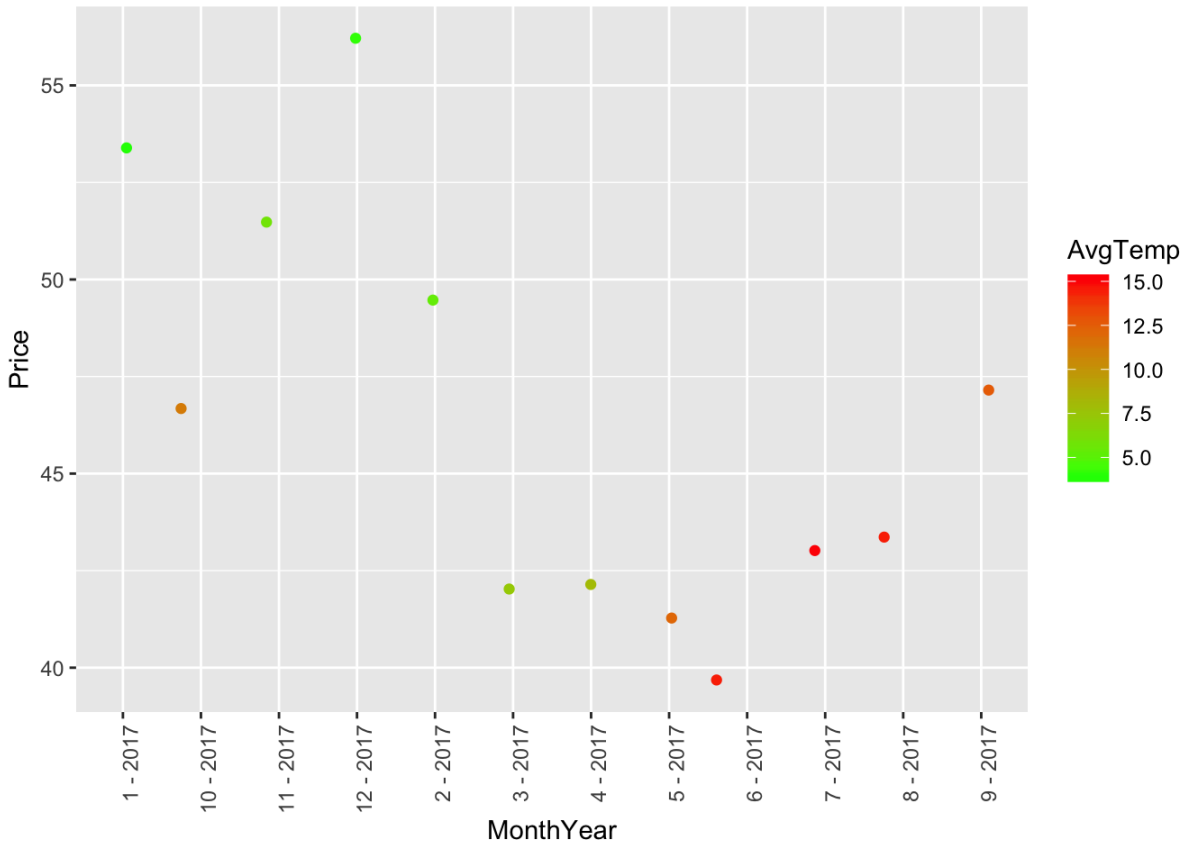
I plotted Price versus Average Temperature by Quarter.



Price versus Avg Temp by Quarter

5. *Can you make a scatterplot?*

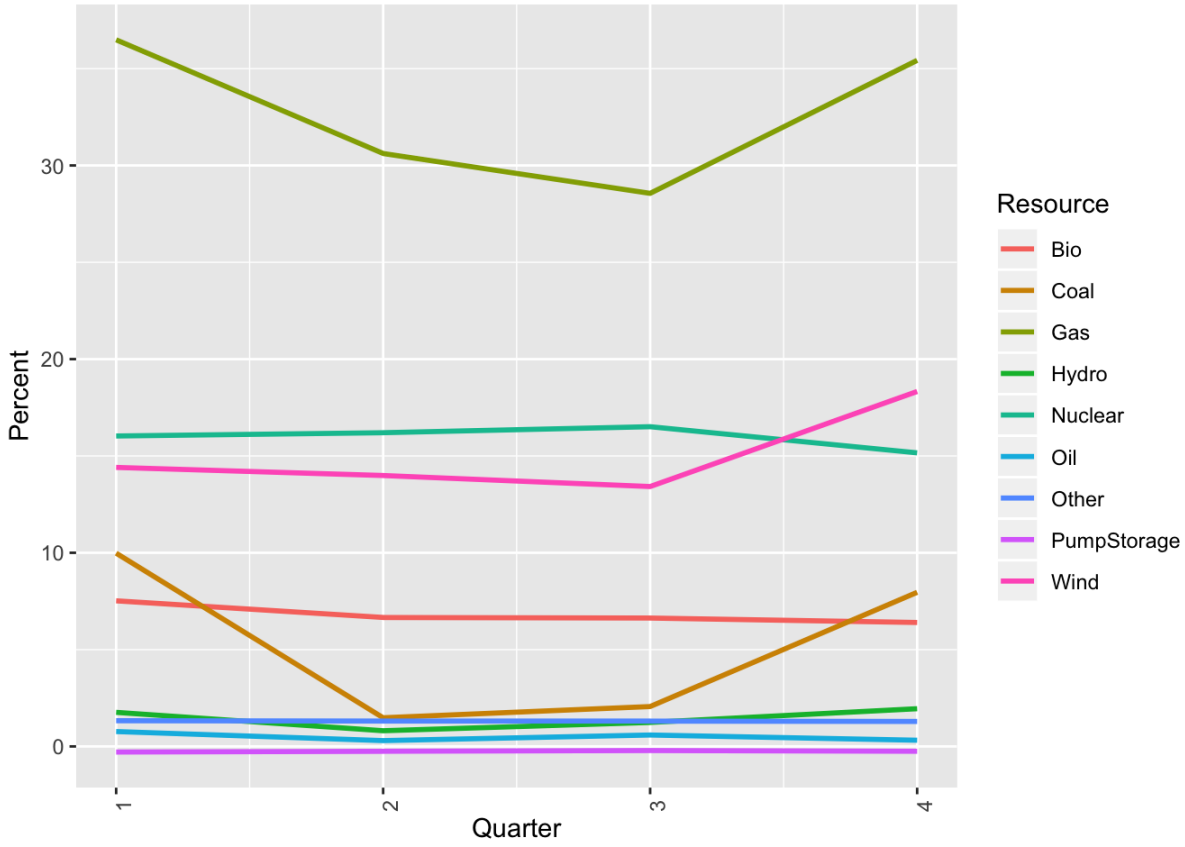
I plotted Average Temp by Price by Month for 2017.



Avg Temp versus Price by Month

6. *Can you make a time-series plot?*

I plotted Resource by Percent by Quarter for 2017. This shows that certain resources are definitely seasonal, with Coal and Gas being the resources that are more in demand in colder quarters.



Resource Percentage by Quarter

Limitations of the Data Sets

Because I don't have specific geographic data (the temperature is Mean UK temperature), I can't do more detailed analysis on the effects of temperature on resource mix. The resource mix and Day Ahead price were also not available at a low enough granularity to do the detailed analysis I might have been able to perform if I had been able to dive into one specific geographic location.

I would have preferred to obtain the cost to run price for the resource mix data set, but it was not available. This would have provided some further analysis on the temperature sensitivity. It also would have allowed me to overlay average DA price with the cost to run to predict the percentage resource usage.

Findings and Recommendations

Findings

1. While I was expecting to see some energy resources change their dominance in specific seasons, that did not end up being the case. The resource percentage distribution didn't change drastically from quarter to quarter as expected. The outlier in this was the fourth quarter 2017 increase in percentage of Wind Powered energy. This could be due to weather or additional wind capacity being brought online.

2. I found that the Highest Day Ahead Prices appeared to occur in correlation to the coldest mean Temperature. This finding would lend support to modeling the UK energy pricing market to mimic the NorthEastern US market (specifically, the PJM Independent System Operator area) as that area tends to have the highest prices during the coldest temperatures as well. In the PJM market, this is due to the prevalence of natural gas for central heating. Since central heating has a higher demand in the colder temperatures, then the natural gas demand is higher leading to higher input costs to power resources utilizing natural gas.

3. I found that certain resources are definitely seasonal, with Coal and Gas being the resources that are utilized more heavily in the colder quarters (1,4).

Recommendations

1. Additional research I could pursue is to overlay capacity online dates (new power plants) to provide a frame of reference for increased resource percentage. This could provide forecasted information for resource mix.

2. I would like to be able to dig deeper geographically to provide by congestion area instead of the entire country. This would allow for a more detailed pricing forecast.

3. The results of this analysis can be added to an energy pricing model.