An Exploratory Analysis of the German EEX Power Market

Jacob S Townson

May 17, 2018

Introduction

This paper covers my observations and analysis of the data given to me by Josef Spalenka of Genscape in order to determine my qualifications for a position at Genscape. The data used in this document is used in the German day-ahead power market called "EEX". It contains the power prices, fuel prices, emission prices, heating degrees, and wind and solar forecasts for the next day.

My goals with this data are to first and foremost, understand what exactly is this data presenting us with. Then, I will make some code to read in this data to make it usable in R. Then, I will look to see if there are any interesting correlations and relationships to be found in the data. I will present my findings here, and make conclusions based off of them.

The Data

The data given to me was split into 7 files, and some excel files with more sheets than others. My goal here is to explain precisely what this data is telling us, then to explore what possible correlations and relationships we may be able to find.

Coal Future Data

In this section we will be exploring the file "coal_futures_historie_2013.xls". This excel file has 8 sheets. The first tells us most of what we need to know, and that's that coal seems to be phasing out from the German market, if it already hasn't been now that we are well past 2013. The second sheet indicates similarly with it's page full of zeros.

The rest of the sheets in this excel file show the 2013 data for specific contracts and their trade unit along with prices. These bits of data make me curious, as the previously mentioned sheets seem to indicate that there were no trades through the year, however these sheets show the monthly, quarterly, and yearly trade units with their settlement prices. I can only assume this means that these prices are the prices that the trade units cost, however no contracts for these units were made.

Since we can see that most of this data is just zeros, we will ignore it from here. However, I would like to understand more of what exactly these data sets are saying, especially the differences between the "Coal-Futures Total" sheets and the contract sheets afterwards. Maybe if offered the job, I can learn more about these things!

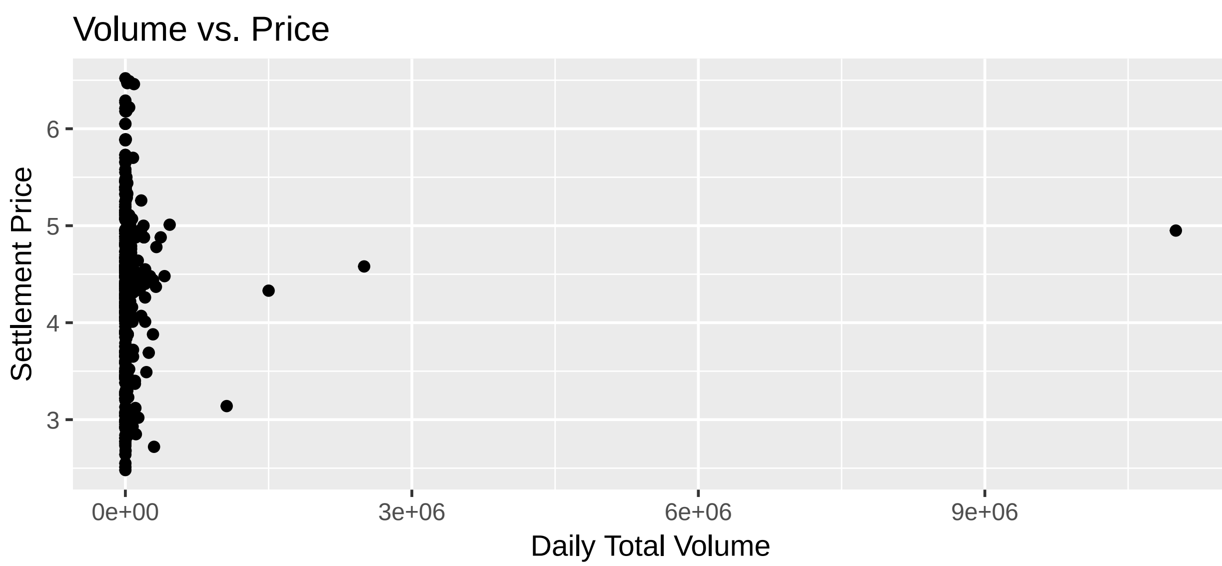
Emission History

This data set (in the file "emission_spot_historie_2013.xls") gives the emission prices based on specific contracts and their emission volumes. This excel file has 3 sheets, first explaining the contracts' emission allowance, the next showing reductions, and the final showing the market auction information.

For the most part, this data set is relatively clear in what it shows, however in the market auction sheet, we are given a column labeled "auction details". For my purposes, this information will be ignored as without more former knowledge of the system, I will not be able to accurately try to analyze anything about it.

For curiosity's sake, let's quickly look at the emission allowance data, and compare the settlement price with the total volume of emissions. We will use the "emission_allowance" dataframe I created in R to do this efficiently.

```
em_price = emission_allowance$Settlement.Price.EUR.EUA
em_vol = emission_allowance$Daily.Total.Volume..incl..OTC.
em_cont = emission_allowance$Contract
em_trades = emission_allowance$Trades
qplot(em_vol,em_price, xlab = 'Daily Total Volume', ylab = 'Settlement Price',
      main = 'Volume vs. Price')
```



It's clear there are some outliers here, so for now, let's remove them and see what the plot looks like then.

```
allow = data.frame(em_price,em_vol,em_cont,em_trades)
allow = arrange(allow, desc(em_vol))
```

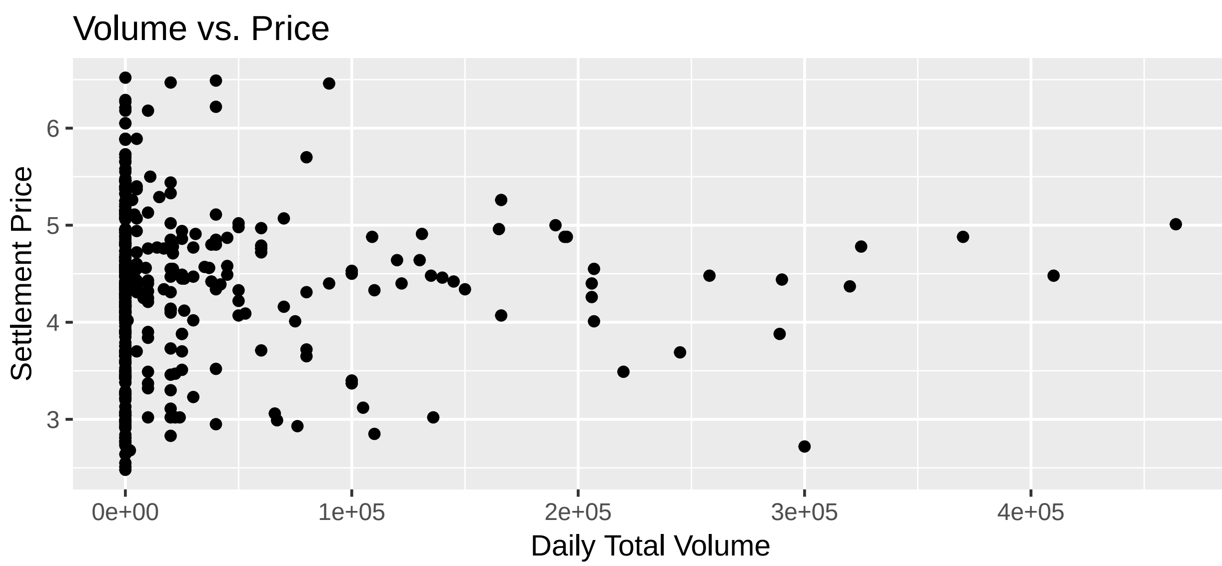
```
## Warning: package 'bindrcpp' was built under R version 3.4.4
```

```
head(allow)
```

```
##   em_price   em_vol em_cont em_trades
## 1     4.95 11000000    EUSP        11
## 2     4.58  2500000    EUSP         2
## 3     4.33  1500000    EUSP         2
## 4     3.14  1061000    EUSP        21
## 5     5.01   464000    EUSP         1
## 6     4.48   410000    EUSP        25
```

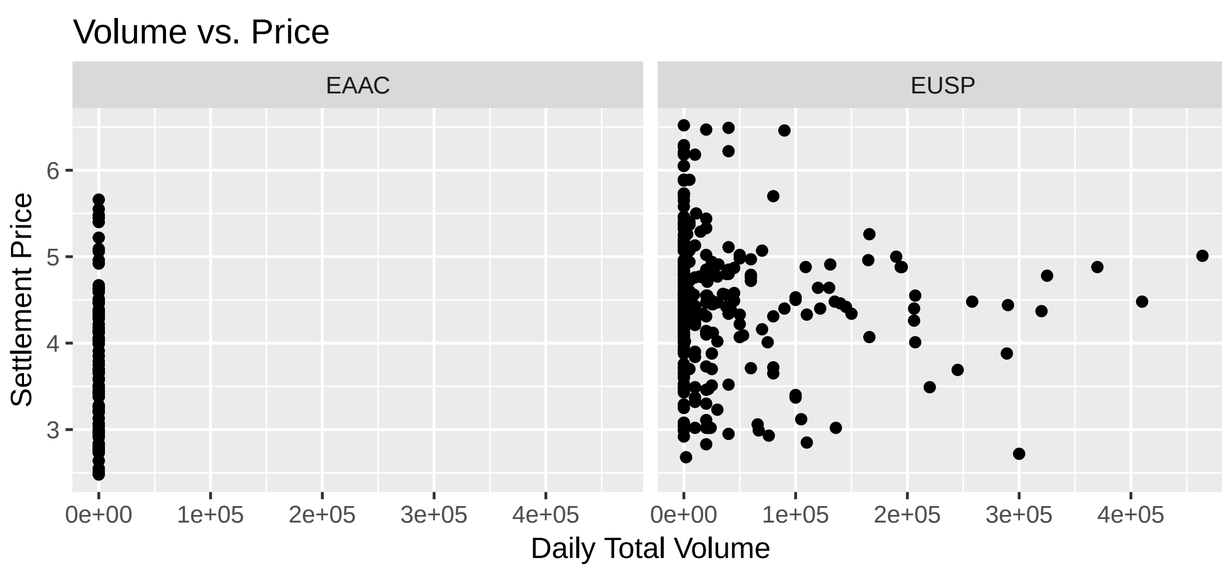
So from here we can see our outliers, and now we can remove them accordingly.

```
allow = filter(allow, em_vol < 1061000)
qplot(em_vol,em_price, data = allow, xlab = 'Daily Total Volume', ylab = 'Settlement Price',
      main = 'Volume vs. Price')
```



This plot still seems fairly noisy. Perhaps we should divide the information by contract to see if that has something to do with the noise in price. There are two types of contracts in this data set, EAAC, and EUSP.

```
qplot(em_vol,em_price, data = allow, xlab = 'Daily Total Volume', ylab = 'Settlement Price',
      main = 'Volume vs. Price', facets = .~em_cont)
```



This seems to show just more noise, no correlation seems to show for this information. I think we can then conclude that the daily volume of emissions doesn't affect the settlement price.

As a quick aside, note that emission volume seems to stay fairly constant over time, excluding a few outliers. Although no plot was included to show this, it can be seen fairly easily through a quick glance at the data as well. So let's move on!

Gas History

This data set (found in the gas_spot_historie_2013.xls) is by far the most dense of any of them all. With 10 sheets, it has a lot of information to offer. The first of these sheets gives us the reference price for each market area for each day. The strange thing about this collection of data is all of the missing values. I have to wonder if this is possibly because there were no trades on these given days for these market areas, or if whoever was collecting this data was unable to get it. Either way, for time's purpose, we will simply ignore the empty values for now.

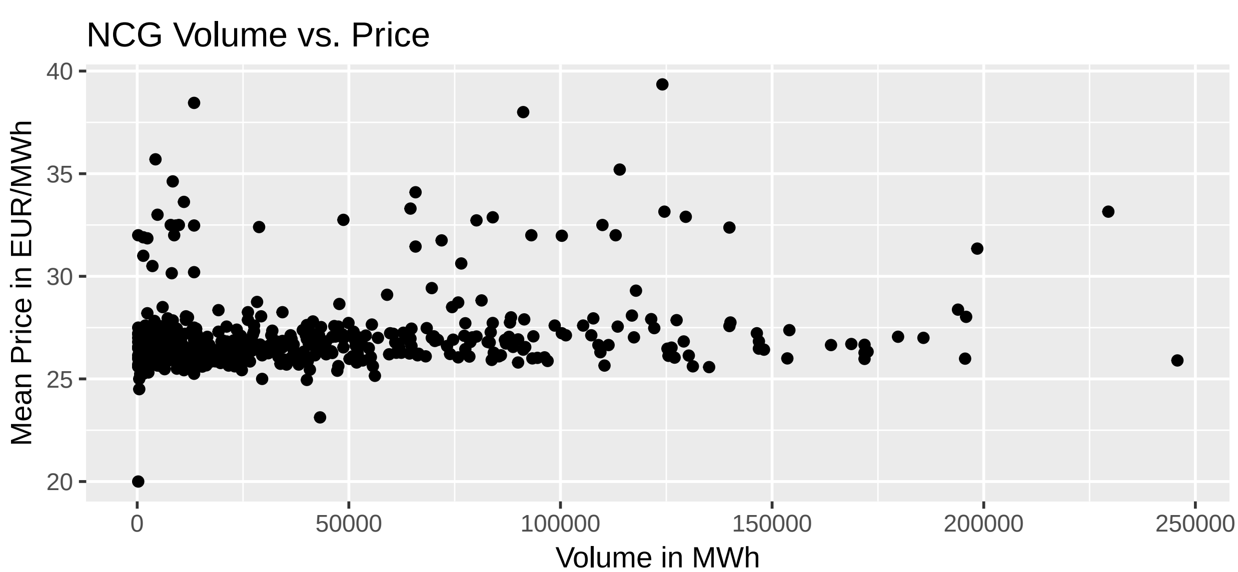
The following sheets give us data on what I have understood to be 3 European gas companies/traders, NCG (NetConnect Germany), GASPOOL, and TTF (Title Transfer Facility). For each one, we get the market information for the whole 2013 year. This has their number of trades, price information, and volume traded for every day.

Let's look and see if we can find any trends in this data! To keep things simple, let's start by looking at the NCG data, and if we find some interesting things, we'll dive down the rabbit hole. Otherwise, we'll move on. First things first, let's write a bit of code to simplify the low vs. high price problem in the data.

```
ncg.high = gas_market.ncg$High.Price.EUR.MWh
ncg.low = gas_market.ncg$Low.Price.EUR.MWh
ncg.mean_price = rowMeans(cbind(ncg.low,ncg.high))
```

Now that we have this mean price for each day, let's see if the volume vs. price plot shows us anything interesting.

```
ncg.trades = gas_market.ncg$Trades
ncg.vol = gas_market.ncg$Volume.MWh
ncg.day = gas_market.ncg$Trading.Date
NCG = data.frame(ncg.day,ncg.mean_price,ncg.vol,ncg.trades)
qplot(ncg.vol,ncg.mean_price, data = NCG, xlab = 'Volume in MWh', ylab = 'Mean Price in EUR/MWh',
      main = 'NCG Volume vs. Price')
```

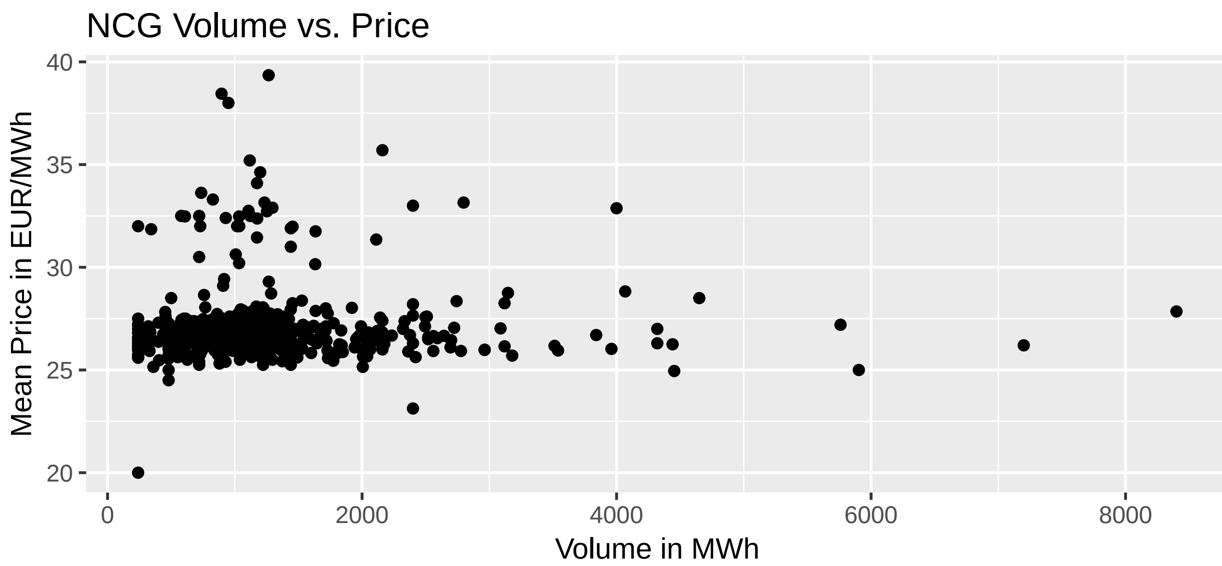


We seem to get a strange looking plot here. It almost looks as if no matter the volume, the price stays between 25 and 30 euros. This makes me wonder if possibly the number of trades has something to do with it. Let's take a look.

```
NCG = filter(NCG, ncg.trades > 0)
NCG = mutate(NCG, vol_trade = ncg.vol/ncg.trades)
```

What I have done in this code chunk is to first remove any zero trade values in the data, then make a new variable to show roughly the amount of volume per trade.

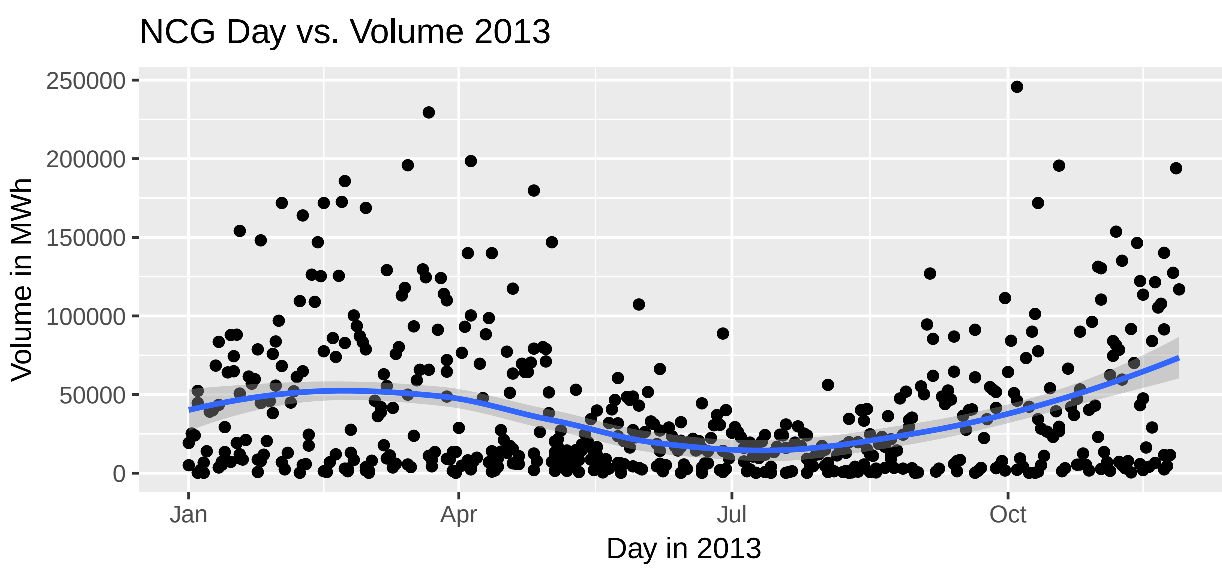
```
qplot(vol_trade,ncg.mean_price, data = NCG, xlab = 'Volume in MWh', ylab = 'Mean Price in EUR/MWh',
      main = 'NCG Volume vs. Price')
```



This didn't really clear much up for us. So seemingly the price in euros per MWh is not affected by the volume. Before we move on, let's take a quick look and see how it changes throughout the year though.

```
qplot(ncg.day, ncg.vol, data = NCG, xlab = 'Day in 2013', ylab = 'Volume in MWh',
      main = 'NCG Day vs. Volume 2013', geom = c("point", "smooth"))
```

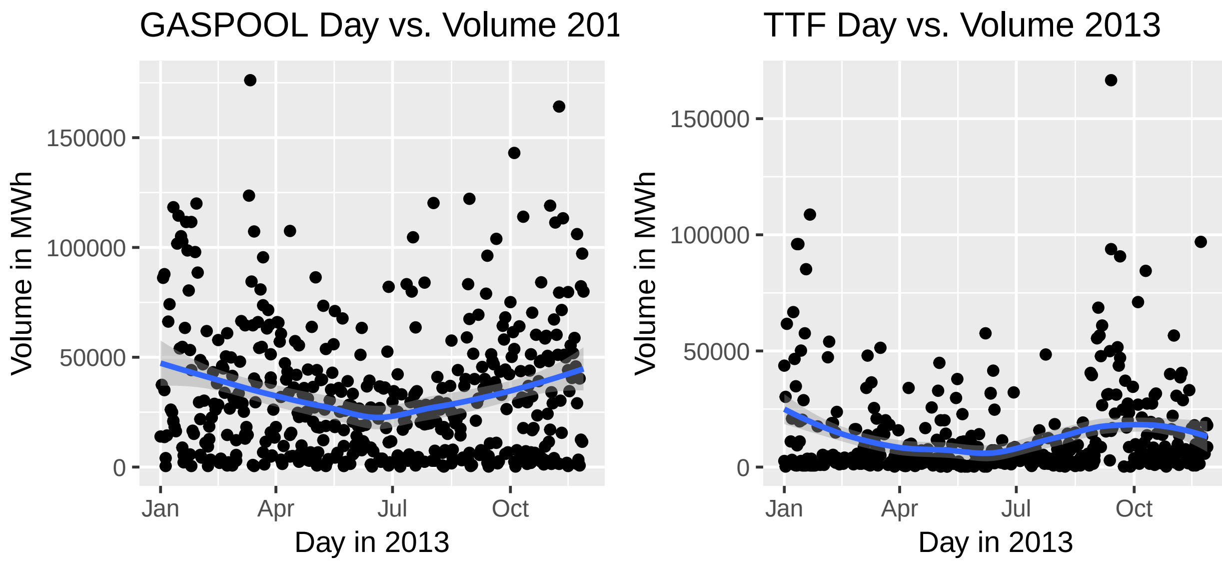
```
## `geom_smooth()` using method = 'loess'
```



Here we see an interesting bit of information. There was a bit of a surge of volume traded right before April, and going into December. This could mean a number of things, from people needing to heat their houses in the winter months, to people driving and using fuel to travel during spring. Without more data, we can't be sure, but we can make some interesting assumptions!

If we take this work and apply it to the GASPOOL and TTF data, and look at the time vs. volume plot, this is what we find:

```
## `geom_smooth()` using method = 'loess'
## `geom_smooth()` using method = 'loess'
```



It's strange, the GASPOOL data doesn't seem to tell us much when it comes to the time of year other than there is a dip in average volume around July, but even that is slight. The TTF data shows us a little more, however it strangely looks almost like the exact opposite of the NCG data. Why this is, I can't say from this data alone, but it is interesting to note nonetheless.

Heating Degrees

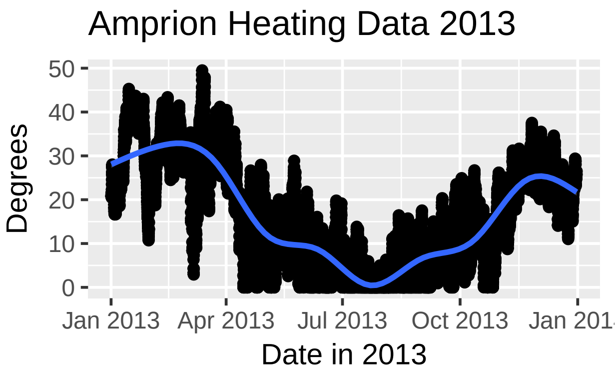
This data set (found in HeatingDegrees.xlsx) contains data for heating energy. The first sheet covers from 12/30/2009 - 07/31/2013, and the second sheet covers from 08/01/2013 - 01/06/2014. After a quick Google, we can understand that Amprion, Transnet BW, TenneT South, TenneT North, and 50 Hertz are all companies. Based off of the title of the data set, I believe we can assume that the values given for each then are degrees heated on that given day and time.

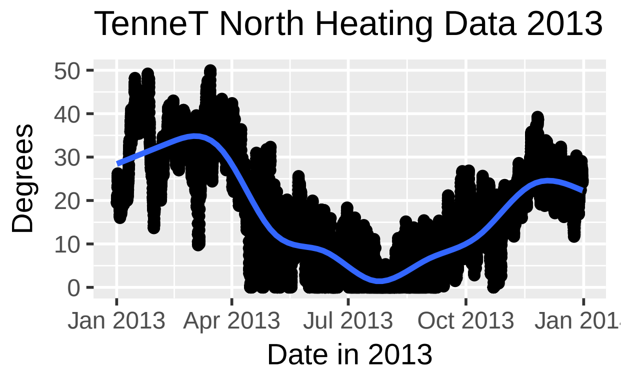
Since our focus is on the year 2013, let's shrink this data down a bit to get it usable four out purposes. Then we'll take a look at how much these companies heat things up.

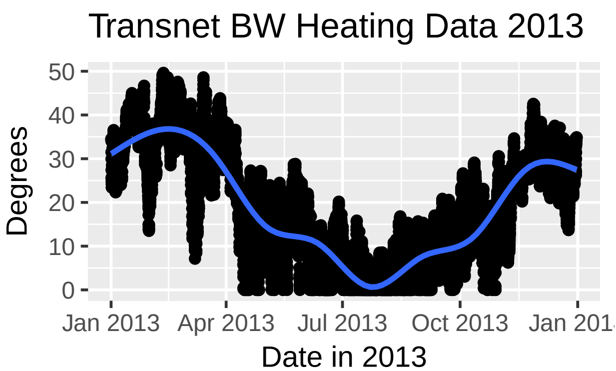
```r
colnames(heating2) = c('Date..GMT.', 'Amprion.EDDL','Transnet.BW','TenneT.South','TenneT.North',
                        'X50.Hertz')
heating = bind_rows(heating1,heating2)
heating13 = filter(heating, Date..GMT. >= "2013-01-01" & Date..GMT. <= "2013-12-31")
```

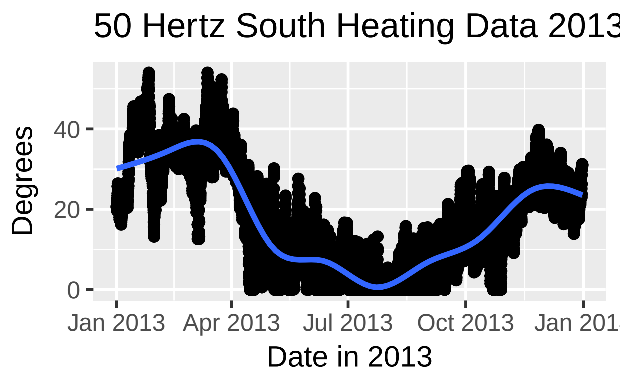
And now we have a complete dataframe with all of the data from 2013! So let's check it out some plots and see how things look for each company.

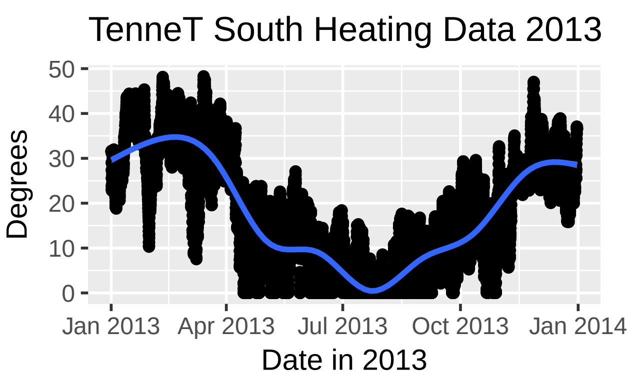
```
## `geom_smooth()` using method = 'gam'
## `geom_smooth()` using method = 'gam'
## `geom_smooth()` using method = 'gam'
## `geom_smooth()` using method = 'gam'
## `geom_smooth()` using method = 'gam'
```









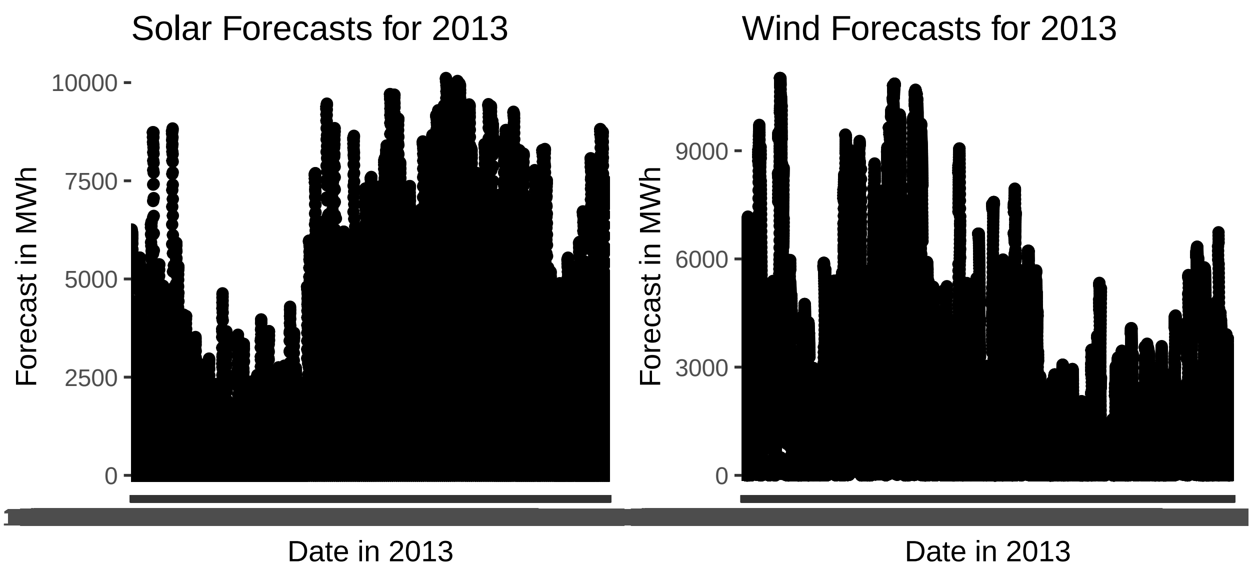


Wow! These plots are super interesting! So here, notice how similar all of these companies' heating data are! It's easier to see in the blue lines here how the growth is happening, and you can see how strikingly similar all of these are. From this we can assume that the time of year affects all of the companies the same temperature-wise. While this isn't all that surprising, it is interesting to see exactly how similar these plots are to each other.

Wind and Solar Forecasts

These data sets are simple enough in explanation, we have dates and the forecast for how many megawatt hours can be produces on given days at given times. Before we can play with this data, we have to get the dates and times in correctly. Now let's take a look at these to see what this data looks like exactly.

```
## `geom_smooth()` using method = 'loess'
## `geom_smooth()` using method = 'loess'
```



These are both pretty noisy, so I'm not so sure we'll get too much out of these for a quick exploratory analysis like this. So let's move on for now.

Power Prices

Finally we've made it to the real meat and potatoes of the data given to us. This data set has two sheets (contained in the file "energy_spot_historie_2013.xls"). The first contains the delivery date, information on the euro per megawatt hour for each date, and even each hour of each day. The second sheet contains the strict volumes in megawatt hours (not euro per megawatt hours) for each hour of each day. There is a lot of information in this set of data, but by the end of this paper I hope to find a relatively simple model that we can use to help make predictions for data such as this.

We will touch more on this data in the next section, but for now let's talk in more detail what's going on in this data. In sheet 1, the first thing we are given is the delivery date. This concept is simple enough, so we won't elaborate anymore on it. But next is the Phelix Day Base and Peak variables. Well, a quick Google will tell us that the *Phelix Base* is the average price of all of the hours in a day for the electricity traded on the spot market. And the *Phelix Peak* is the average price of during the hours 9-20 for electricity traded on the spot market. One can assume it get's it's name because these are the actice hours of the day. This method of calculation is used for the the Phelix day and month values, just for days and months respectively. **********

Sheet 2 is almost identical to sheet 1, except for 2 columns of difference, those being the EPEX Spot Volume columns. **********

Analysis

Conclusions

Appendix