EDA_CAPSTONE

Bodie Franklin

7/5/2021

EDA of ERCOT data. The data was compiled from ERCOT data Archives. Time period is 2010-2021.

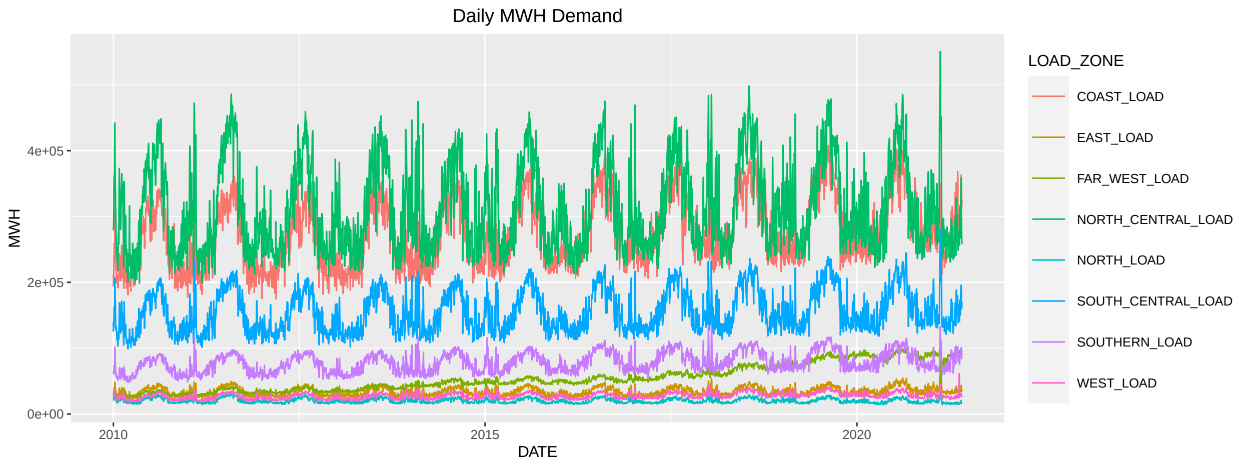
```r
#Getting Rid of Warning Messages
defaultW <- getOption("warn")
options(warn = -1)

library(tidyverse)
library(ggcorrplot)
library(naniar)
library(reshape2)

#Reading in the File
df <- read.csv(file="C:/users/bodie/Documents/CAPSTONE_DATA.csv")

#Change Day format
df$DAY <- as.Date(df$DAY,format="%m/%d/%Y")
```

Checking NA Values

Due to flow of document, I moved this to the end. We have a lot of missing Na values and output was tiresome. Analysis of Na's can be found at the end.

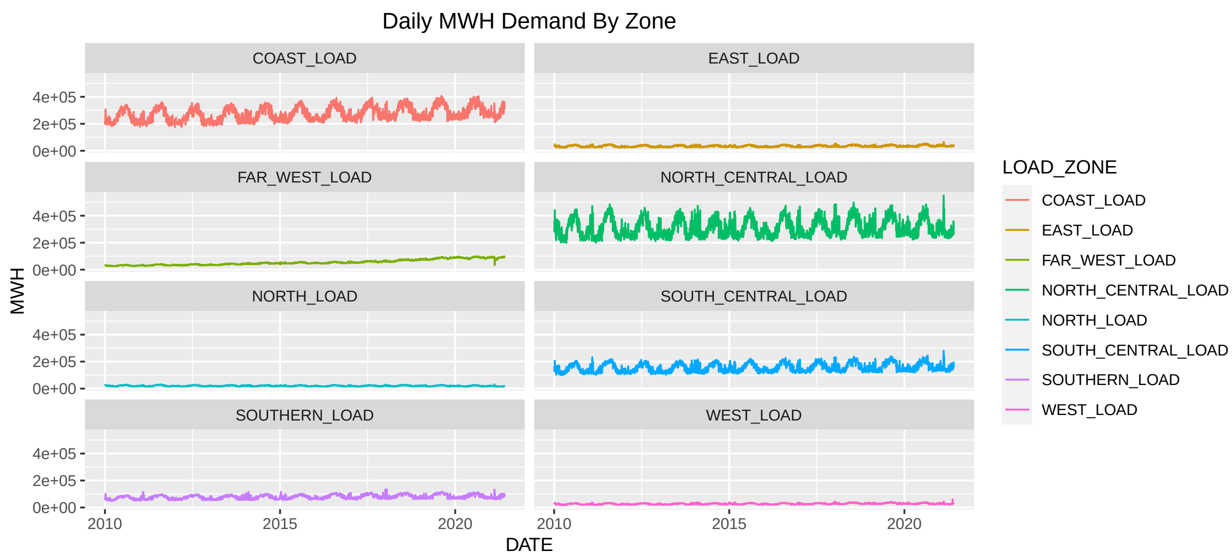
Exploring Daily MWH

```r
#Subsetting data in order to use gather function
LOAD_DF <-df %>% select('DAY','COAST_LOAD','EAST_LOAD','FAR_WEST_LOAD','NORTH_LOAD',
                        'NORTH_CENTRAL_LOAD','SOUTHERN_LOAD','SOUTH_CENTRAL_LOAD','WEST_LOAD')

#Data Wrangling
LOAD_DF <- LOAD_DF %>% gather("LOAD_ZONE","MWH",2:9)


LOAD_DF %>% ggplot(aes(x=DAY,y=MWH,color=LOAD_ZONE))+geom_line()+xlab("DATE") +
  ylab("MWH")+ggtitle("Daily MWH Demand")+theme(plot.title = element_text(hjust = 0.5))+
  theme(legend.key.size = unit(1.0, 'cm'))
```



Regardless of the location in TX, each load zone is following a cyclical nature. Intuition tells us that this makes sense since the MWH generated is a response due to weather cycles. Perhaps we will gain more insights by individually plotting the load zones.

```r
#Breaking the graph out by Zone
LOAD_DF %>% ggplot(aes(x=DAY,y=MWH,color=LOAD_ZONE))+geom_line()+xlab("DATE") +
  ylab("MWH")+ggtitle("Daily MWH Demand By Zone")+theme(plot.title = element_text(hjust = 0.5))+
  facet_wrap(~LOAD_ZONE,ncol=2)
```

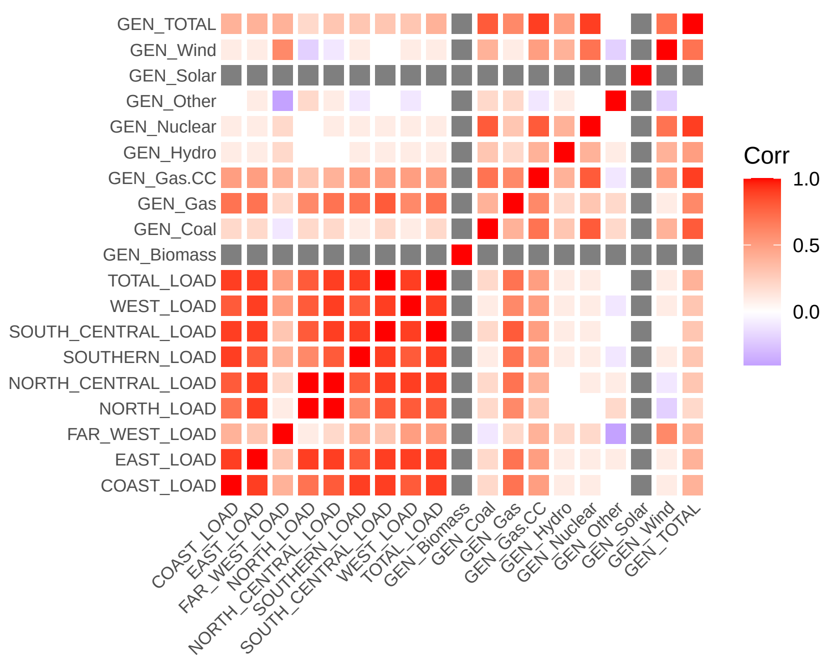


The break out of the individual load zones provides insights to population density. It seems that EAST, FAR_WEST, NORTH, SOUTHERN and WEST load zones are generating significantly lower Daily MWH than COAST, NORTH_CENTRAL, SOUTH_Central. Perhaps, these Load zones are less populated and thus have lower Energy demand requirements.

It is interesting that all the Load Zones with the exception of FAR_WEST are following a cyclical nature. Visually, we aren't seeing any impact of population growth over a 10 year period for the Load Zones. One would think that we would see some type of upward trend in the graphs.

Correlation plot

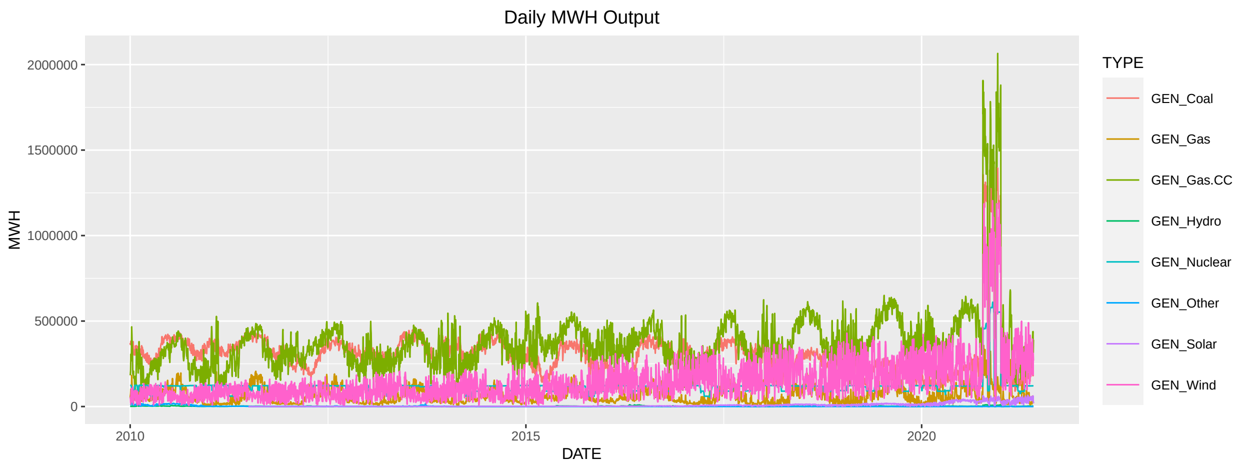
```r
corr_df <- df %>% select('COAST_LOAD','EAST_LOAD','FAR_WEST_LOAD',
                         'NORTH_LOAD','NORTH_CENTRAL_LOAD','SOUTHERN_LOAD',
                         'SOUTH_CENTRAL_LOAD','WEST_LOAD','TOTAL_LOAD',
                         'GEN_Biomass','GEN_Coal','GEN_Gas','GEN_Gas.CC',
                         'GEN_Hydro','GEN_Nuclear','GEN_Other','GEN_Solar','GEN_Wind','GEN_TOTAL')
corr <-round(cor(corr_df),1)

ggplot(melt(corr), aes(Var1, Var2, fill=value)) +
  geom_tile(height=0.8, width=0.8) +
  scale_fill_gradient2(low="blue", mid="white", high="red") +
  theme_minimal() +
  coord_equal() +
  labs(x="",y="",fill="Corr") +
  theme(axis.text.x=element_text(size=8, angle=45, vjust=1, hjust=1,
                                 margin=margin(-3,0,0,0)),
        axis.text.y=element_text(size=8, margin=margin(0,-3,0,0)),
        panel.grid.major=element_blank())
```



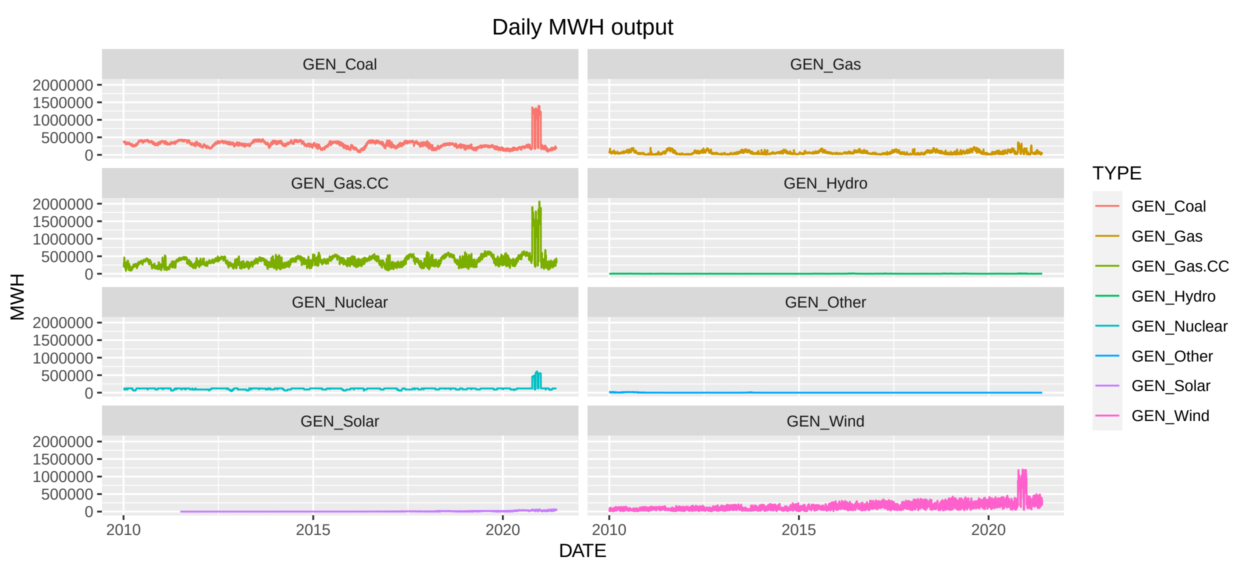
I figured certain Load Zones would be more reliant on certain generation methods. Surprisingly,the correlation plot really only shows strong correlation between Gas/Gas.CC and the load zones. This suggest that TX sources most of its' energy from Gas/Gas.CC. Although, FAR_WEST_LOAD is the exception. It seems that there is a strong correlation between GEN_WIND & FAR_WEST_LOAD. From my personal experience, I know there is a large population of Wind farms. Therefore, it makes sense that West TX is sourcing their energy from Wind.

```r
GEN_TYPE <- df %>% select('DAY','GEN_Coal','GEN_Gas','GEN_Gas.CC',
                          'GEN_Hydro','GEN_Nuclear','GEN_Other','GEN_Solar','GEN_Wind')
```

```r
GEN_TYPE <- GEN_TYPE %>% gather("TYPE","MWH",2:9)

GEN_TYPE %>% ggplot(aes(x=DAY,y=MWH,color=TYPE))+geom_line()+xlab("DATE") +
  ylab("MWH")+ggtitle("Daily MWH Output")+theme(plot.title = element_text(hjust = 0.5))+
  theme(legend.key.size = unit(1.0, 'cm'))
```




```r
GEN_TYPE %>% ggplot(aes(x=DAY,y=MWH,color=TYPE))+geom_line()+xlab("DATE") +
  ylab("MWH")+ggtitle("Daily MWH output")+theme(plot.title = element_text(hjust = 0.5))+
  facet_wrap(~TYPE,ncol=2)
```



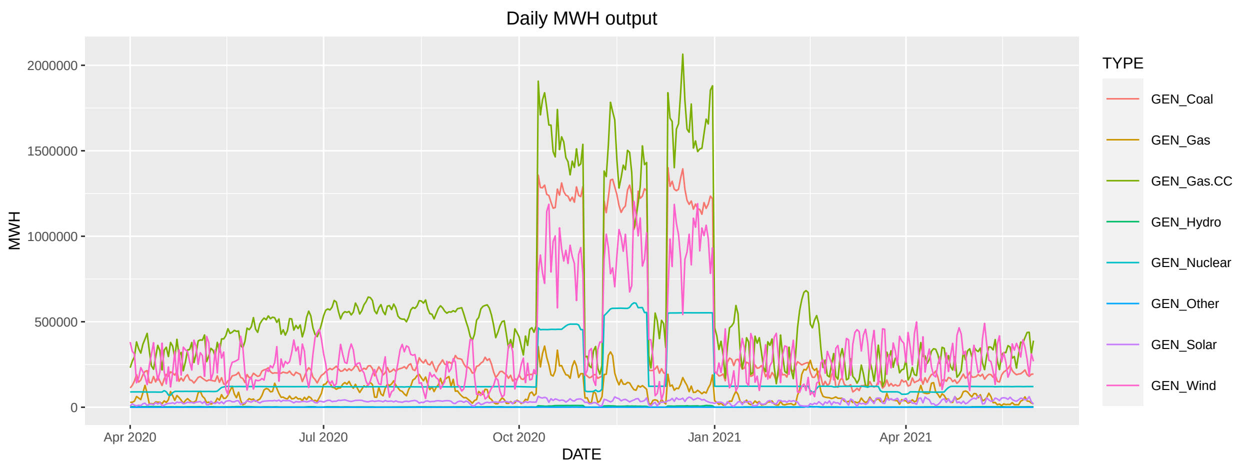
Above, we have Daily MWH output between 2010-2021. My initial thought is that Coal seems be a huge energy source. In the correlation plot, Coal didn't a strong correlation with any of the Load Zones. Therefore, it's surprising to see how large Coal's output is for TX. In contrast, we see that Gas.CC is one of the largest energy sources and that aligned with the output of the correlation plot.

The spike in output at right side of the graph is very interesting. I'm going to zoom in on that portion of the graph to see where this spike occurred. Perhaps the spike occurred during the TX Feb storm? I'll use 4/1/20 as the starting point.

Worth noting, that Solar,Hydro & Other don't reflect the output spike. Futher suggesting that these aren't huge energy sources for TX.

Zooming in on the spike!

```
GEN_2020 <- GEN_TYPE %>% filter(DAY >= as.Date("2020-04-01"))

GEN_2020 %>% ggplot(aes(x=DAY,y=MWH,color=TYPE))+geom_line()+xlab("DATE") +
  ylab("MWH")+ggtitle("Daily MWH output")+theme(plot.title = element_text(hjust = 0.5))+
  theme(legend.key.size = unit(1.0, 'cm'))
```

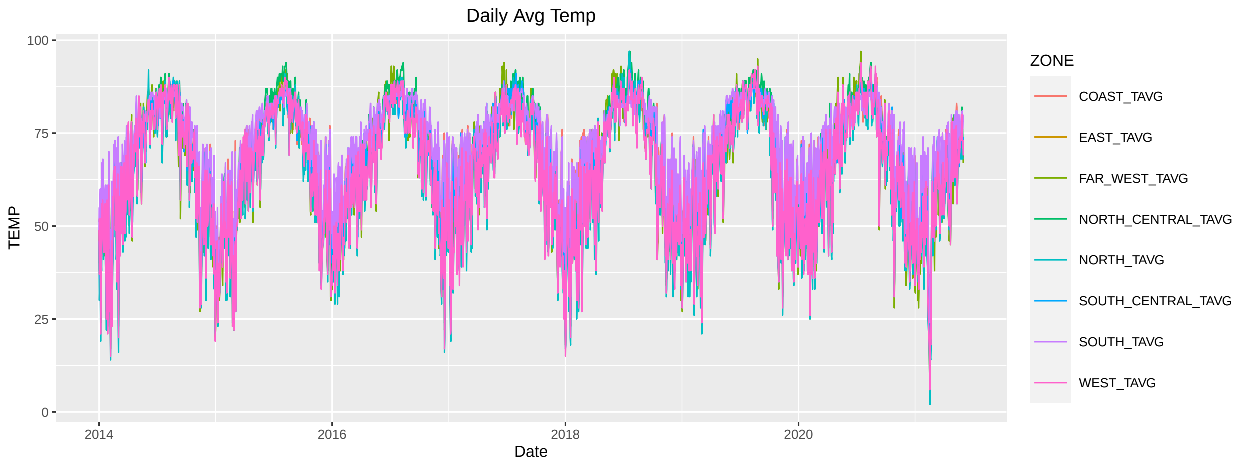


Wow! After zooming in on the graph, we see that the spike occurred in Nov/Dec 2020. This raises more ?'s for me. For starters, was this a historical cold Nov-Dec for TX? Was there a spike in output because transplants from coast states were not used to the cold? In other words, did people relocating due to COVID contribute to the spike? Could this spike have foreshadowed the problems that occurred in Feb 2021?

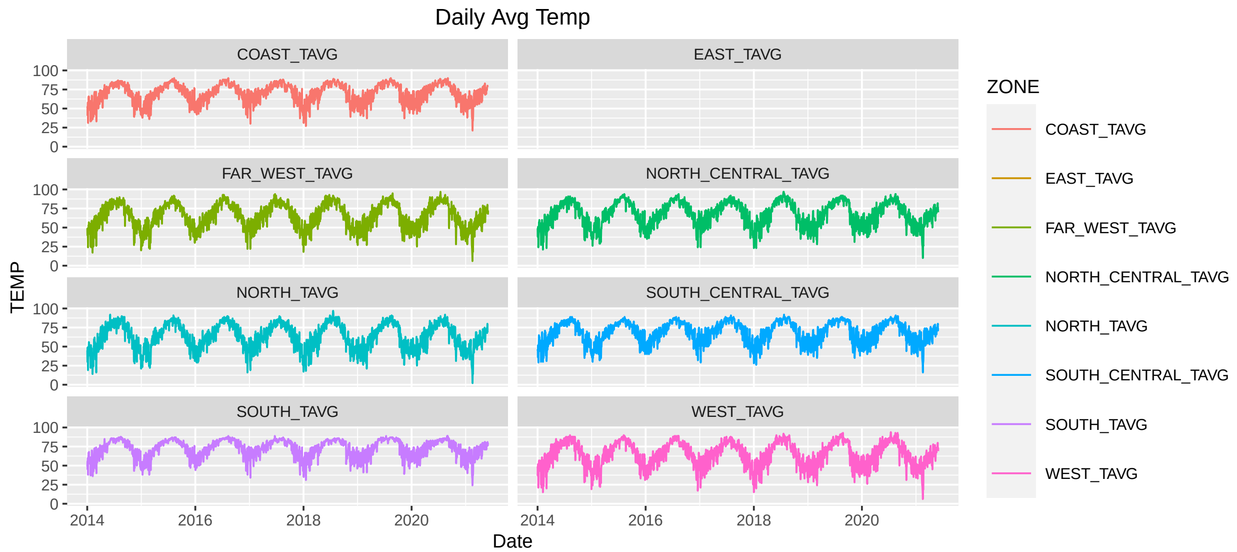
AVG Temp

Let's see if we can notice any large temp movements in Nov/Dec 2020 by exploring Avg Temp's between 2010-2021.

```
AVG_TEMP <- df %>% select('DAY','WEST_TAVG','SOUTH_CENTRAL_TAVG',
                          'SOUTH_TAVG','NORTH_CENTRAL_TAVG','COAST_TAVG','FAR_WEST_TAVG','EAST_TAVG','N

AVG_TEMP <- AVG_TEMP %>% gather("ZONE","TEMP",2:9)
##We know that a lot of temp data is missing, using 2014 as Starting Point per previous EDA not shown i
AVG_TEMP <- AVG_TEMP %>% filter(DAY >= as.Date("2014-01-01"))
AVG_TEMP %>% ggplot(aes(x=DAY,y=TEMP,color=ZONE))+geom_line()+ggtitle("Daily Avg Temp")+xlab("Date")+the
```




```
AVG_TEMP %>% ggplot(aes(x=DAY,y=TEMP,color=ZONE))+geom_line()+ggtitle("Daily Avg Temp")+xlab("Date")+the
  facet_wrap(~ZONE,ncol=2)
```

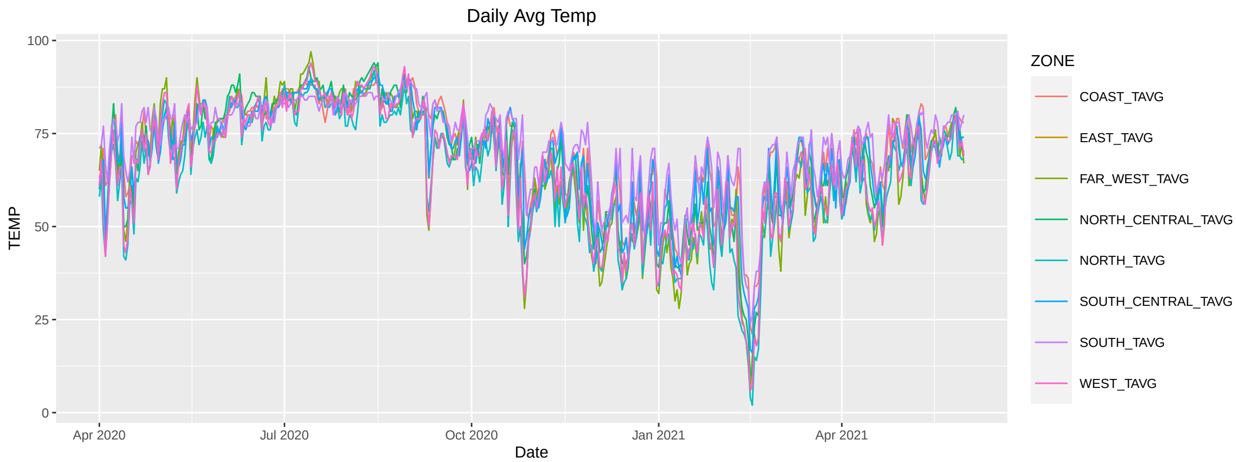


Note: Based on previous EDA not shown in this document, I realize a large amount of AVG temp data was missing from our dataset. I decided to use 2014 as the starting point.

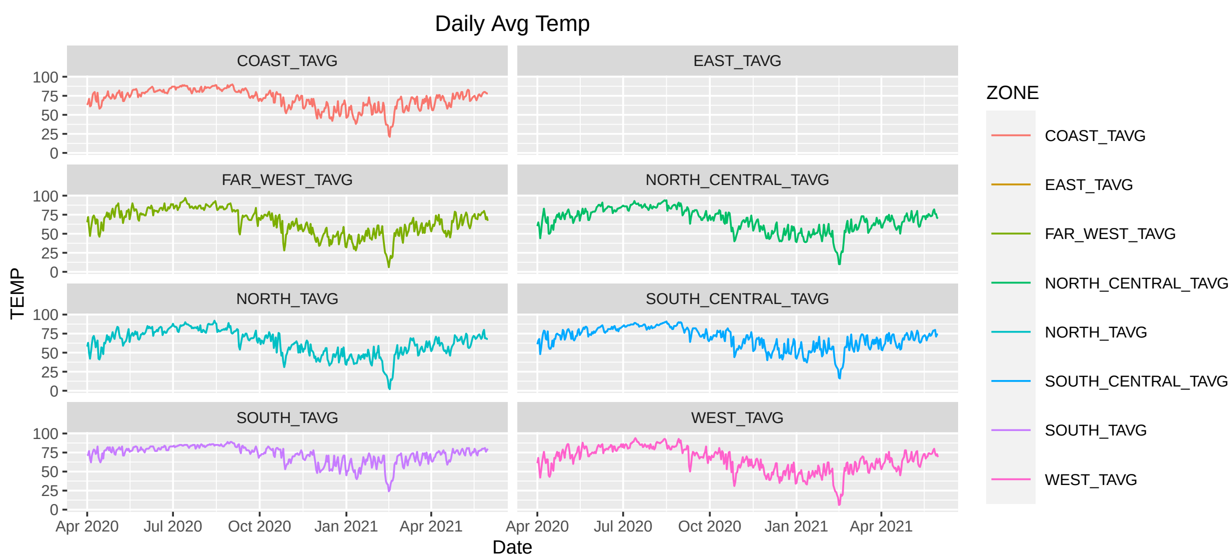
From first glance, we see that AVG temp(regardless of LOAD_ZONE) follows cyclical nature. I don't notice anything odd for Nov/Dec 2020. Perhaps if we zoom in on that time period, we will notice anything unusual.

With breaking out the Daily Avg Temp by Load_zone, the only interesting discovery is that we don't have any daily avg temp EAST_LOAD zone. This potentially may cause issues when modeling demand down the line.

```
AVG_TEMP <- AVG_TEMP %>% filter(DAY >= as.Date("2020-04-01"))
AVG_TEMP %>% ggplot(aes(x=DAY,y=TEMP,color=ZONE))+geom_line()+ggtitle("Daily Avg Temp")+xlab("Date")+the
```




```
AVG_TEMP <- AVG_TEMP %>% filter(DAY >= as.Date("2020-04-01"))
AVG_TEMP %>% ggplot(aes(x=DAY,y=TEMP,color=ZONE))+geom_line()+ggtitle("Daily Avg Temp")+xlab("Date")+the
  facet_wrap(~ZONE,ncol=2)
```



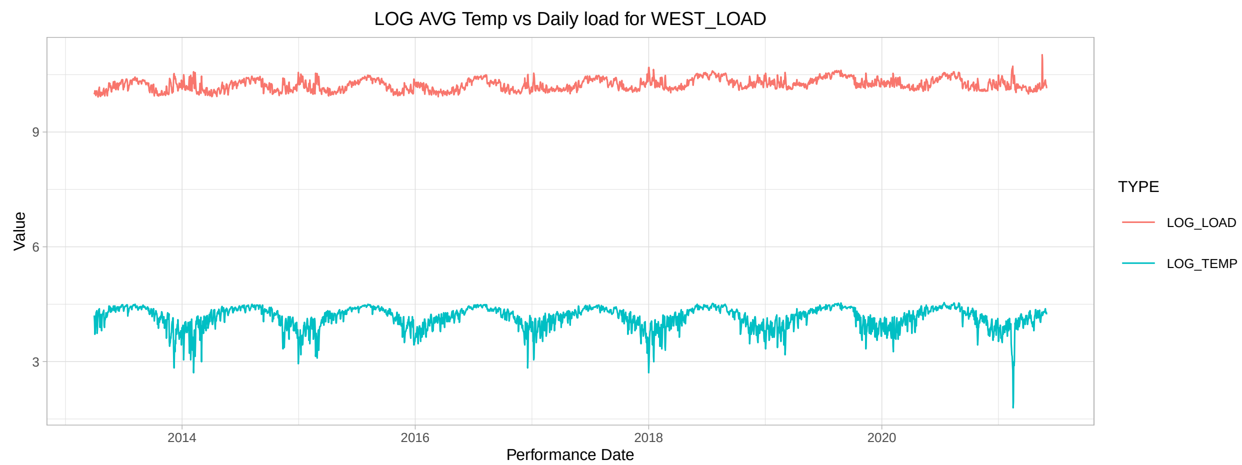
After zooming in on Daily AVG Temp, there doesn't appear to be anything unique about NOV/DEC 2020. Although,we see the huge dip in Daily AVG temp in Feb 2021 due to the winter storm.

Although there is nothing noticeable different for NOV/DEC 2020, my intuition tells me that this is due the transplants from coast states experiencing difficulty adjusting to the cold. Therefore, we need to research inflow of transplants into TX for 2020. Futhermore, the energy output spike for NOV/DEC 2020 needs to be explored in greater detail.

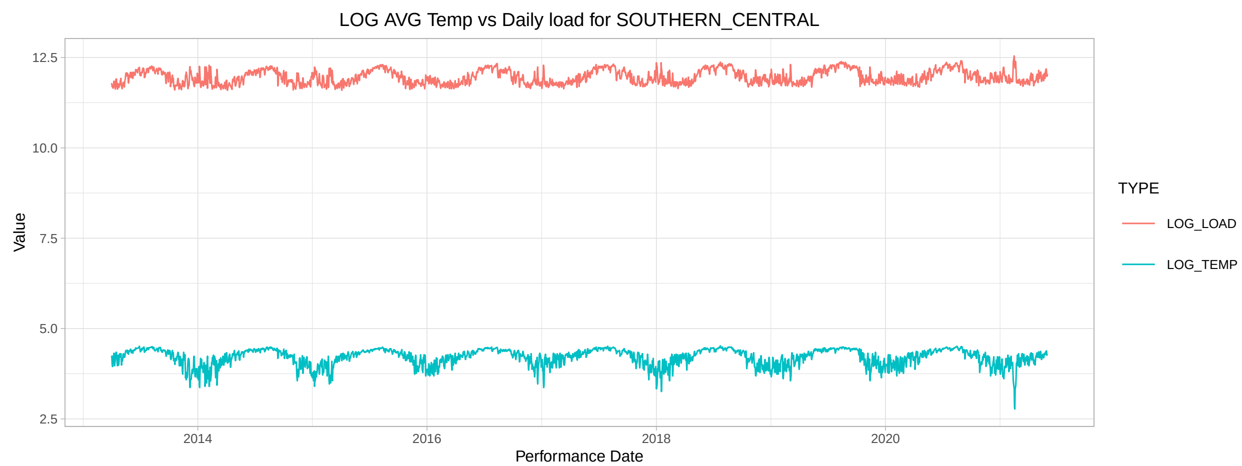
Relationship of AVG TEMP vs LOAD generation

In this section of the EDA, I wanted to explore the relationship between AVG Temp & Load generation. In order to scale the data and have a clean view of the shape, I took the Log of the AVG temp & Load generation respectively.

```
WEST_DF <- df %>% select('DAY','WEST_TAVG','WEST_LOAD')
WEST_DF <- WEST_DF %>% filter(DAY >= as.Date("2013-04-01"))
WEST_DF$LOG_LOAD <- log(WEST_DF$WEST_LOAD)
WEST_DF$LOG_TEMP <- log(WEST_DF$WEST_TAVG)
WEST_DF <- WEST_DF %>% select('DAY','LOG_LOAD','LOG_TEMP')
WEST_DF <- WEST_DF %>% gather("TYPE","VALUE",2:3)
WEST_DF %>% ggplot(aes(x=DAY,y=VALUE,color=TYPE))+geom_line()+xlab("Performance Date")+ylab("Value")+gg
```

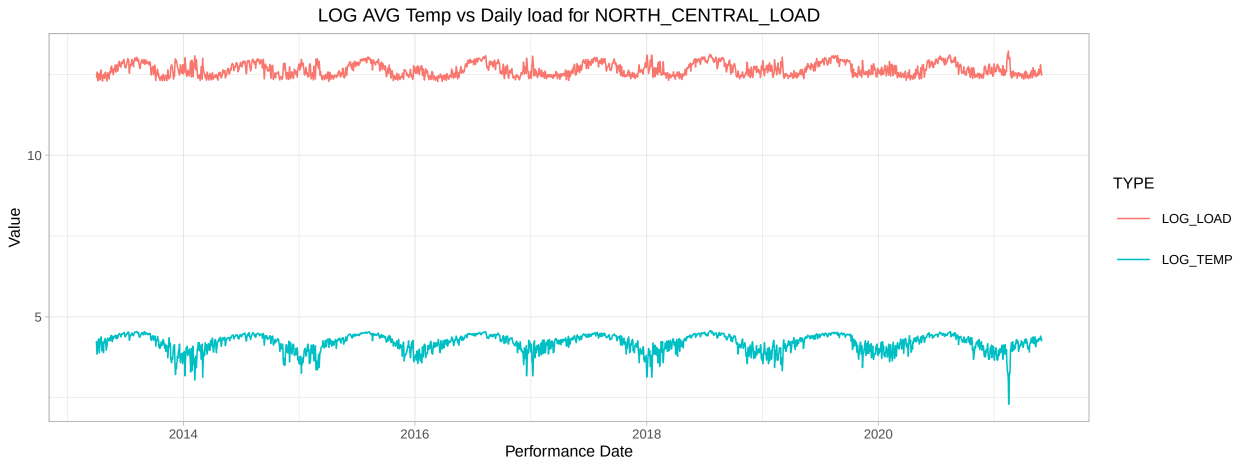



```
SOUTH_CENTRAL_DF <- df %>% select('DAY','SOUTH_CENTRAL_TAVG','SOUTH_CENTRAL_LOAD')
SOUTH_CENTRAL_DF <- SOUTH_CENTRAL_DF %>% filter(DAY >= as.Date("2013-04-01"))
SOUTH_CENTRAL_DF$LOG_LOAD <- log(SOUTH_CENTRAL_DF$SOUTH_CENTRAL_LOAD)
SOUTH_CENTRAL_DF$LOG_TEMP <- log(SOUTH_CENTRAL_DF$SOUTH_CENTRAL_TAVG)
SOUTH_CENTRAL_DF <- SOUTH_CENTRAL_DF %>% select('DAY','LOG_LOAD','LOG_TEMP')
SOUTH_CENTRAL_DF <- SOUTH_CENTRAL_DF %>% gather("TYPE","VALUE",2:3)
SOUTH_CENTRAL_DF %>% ggplot(aes(x=DAY,y=VALUE,color=TYPE))+geom_line()+xlab("Performance Date")+ylab("Va
theme(plot.title = element_text(hjust = 0.5))+  theme(legend.key.size = unit(1.0, 'cm'))
```




```
NORTH_CENTRAL_DF <- df %>% select('DAY','NORTH_CENTRAL_LOAD','NORTH_CENTRAL_TAVG')
NORTH_CENTRAL_DF <- NORTH_CENTRAL_DF %>% filter(DAY >= as.Date("2013-04-01"))
NORTH_CENTRAL_DF$LOG_LOAD <- log(NORTH_CENTRAL_DF$NORTH_CENTRAL_LOAD)
NORTH_CENTRAL_DF$LOG_TEMP <- log(NORTH_CENTRAL_DF$NORTH_CENTRAL_TAVG)
```

```
NORTH_CENTRAL_DF <- NORTH_CENTRAL_DF %>% select('DAY','LOG_LOAD','LOG_TEMP')
NORTH_CENTRAL_DF <- NORTH_CENTRAL_DF %>% gather("TYPE","VALUE",2:3)
NORTH_CENTRAL_DF %>% ggplot(aes(x=DAY,y=VALUE,color=TYPE))+geom_line()+xlab("Performance Date")+ylab("Va
theme(plot.title = element_text(hjust = 0.5))+  theme(legend.key.size = unit(1.0, 'cm'))
```

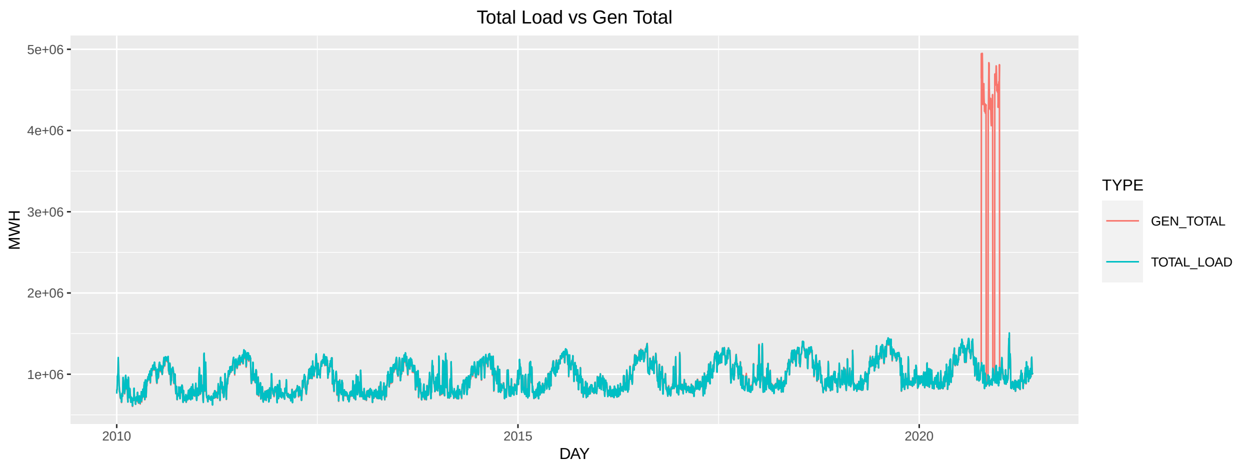


After spot checking a few of the LOAD ZONE's log transformation output and their Avg temp , we see that the both variables seem to mirror each other. The exception is that when large temp decreases occur, we see spikes in energy output. This makes sense since when people are cold, they are going to want to use their heating more.

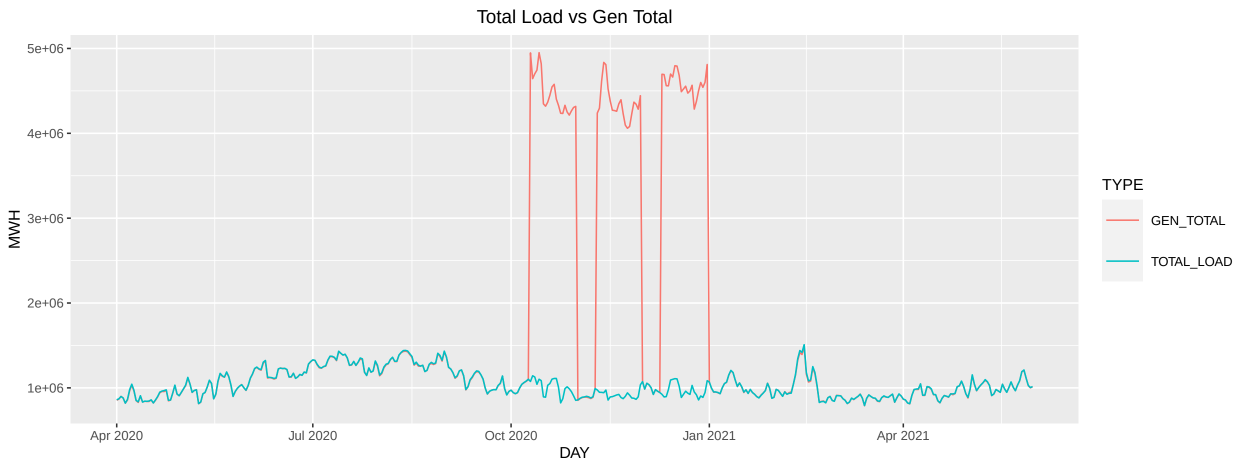
Later on the data will be encoded to check all of the Load zones effortlessly. But for now , this check will do.

Total Load vs Gen

```
Total_GEN <- df %>% select('DAY','TOTAL_LOAD','GEN_TOTAL')
Total_GEN <- Total_GEN %>% gather("TYPE","MWH",2:3)
Total_GEN %>% ggplot(aes(x=DAY,y=MWH,color=TYPE))+geom_line()+ggtitle("Total Load vs Gen Total")+theme(p
```



```r
Total_GEN <- df %>% select('DAY','TOTAL_LOAD','GEN_TOTAL')
Total_GEN <- Total_GEN %>% filter(DAY >= as.Date("2020-04-01"))
Total_GEN <- Total_GEN %>% gather("TYPE","MWH",2:3)
Total_GEN %>% ggplot(aes(x=DAY,y=MWH,color=TYPE))+geom_line()+ggtitle("Total Load vs Gen Total")+theme(p
```



From our previous plots, we see the same occurrences happen again in NOV/DEC 2020. Again this is very interesting QOI and will require further analysis.

Checking NA Values

```r
#Checking for Missing Values
sapply(df, function(x) sum(is.na(x)))
```

```
##                DAY          COAST_LOAD           EAST_LOAD
##                  0                   0                   0
##      FAR_WEST_LOAD          NORTH_LOAD  NORTH_CENTRAL_LOAD
##                  0                   0                   0
##      SOUTHERN_LOAD  SOUTH_CENTRAL_LOAD           WEST_LOAD
##                  0                   0                   0
##         TOTAL_LOAD         GEN_Biomass            GEN_Coal
##                  0                 546                   0
##            GEN_Gas          GEN_Gas.CC           GEN_Hydro
##                  0                   0                   0
##        GEN_Nuclear           GEN_Other           GEN_Solar
##                  0                   0                 546
##           GEN_Wind           GEN_TOTAL           WEST_AWND
##                  0                   0                   2
##          WEST_PGTM           WEST_PRCP           WEST_TAVG
##               3476                   0                1186
##          WEST_TMAX           WEST_TMIN           WEST_WDF5
##                  0                   0                  16
##          WEST_WSF5           WEST_WT01           WEST_WT02
##                 16                3329                4101
##          WEST_WT03           WEST_WT04           WEST_WT05
##               3804                4147                4085
##          WEST_WT06           WEST_WT07           WEST_WT08
```

```
##                4141                4104                3693
##           WEST_WT09           WEST_WT10           WEST_WT11
##                4166                4169                4160
##           WEST_WT13           WEST_WT14           WEST_WT15
##                4004                4167                4169
##           WEST_WT16           WEST_WT17           WEST_WT18
##                3909                4156                4148
##           WEST_WT19           WEST_WT21           WEST_WT22
##                4156                4169                4169
##  SOUTH_CENTRAL_AWND  SOUTH_CENTRAL_PGTM  SOUTH_CENTRAL_PRCP
##                   1                3531                   0
##  SOUTH_CENTRAL_TAVG  SOUTH_CENTRAL_TMAX  SOUTH_CENTRAL_TMIN
##                1186                   0                   0
##  SOUTH_CENTRAL_WDF5  SOUTH_CENTRAL_WSF5  SOUTH_CENTRAL_WT01
##                   5                   5                2444
##  SOUTH_CENTRAL_WT02  SOUTH_CENTRAL_WT03  SOUTH_CENTRAL_WT04
##                3962                3736                4149
##  SOUTH_CENTRAL_WT05  SOUTH_CENTRAL_WT06  SOUTH_CENTRAL_WT07
##                4087                4154                4142
##  SOUTH_CENTRAL_WT08  SOUTH_CENTRAL_WT09  SOUTH_CENTRAL_WT10
##                3894                4155                4169
##  SOUTH_CENTRAL_WT11  SOUTH_CENTRAL_WT13  SOUTH_CENTRAL_WT14
##                4165                3741                4048
##  SOUTH_CENTRAL_WT15  SOUTH_CENTRAL_WT16  SOUTH_CENTRAL_WT17
##                4169                3852                4169
##  SOUTH_CENTRAL_WT18  SOUTH_CENTRAL_WT19  SOUTH_CENTRAL_WT21
##                4165                4169                4160
##  SOUTH_CENTRAL_WT22          SOUTH_AWND          SOUTH_PGTM
##                4168                   2                3463
##          SOUTH_PRCP          SOUTH_TAVG          SOUTH_TMAX
##                   0                1186                   0
##          SOUTH_TMIN          SOUTH_WDF5          SOUTH_WSF5
##                   0                   8                   8
##          SOUTH_WT01          SOUTH_WT02          SOUTH_WT03
##                2520                3915                3697
##          SOUTH_WT04          SOUTH_WT05          SOUTH_WT06
##                4166                4072                4163
##          SOUTH_WT07          SOUTH_WT08          SOUTH_WT09
##                4148                3699                4151
##          SOUTH_WT10          SOUTH_WT11          SOUTH_WT13
##                4168                4163                3842
##          SOUTH_WT14          SOUTH_WT15          SOUTH_WT16
##                4097                4167                3842
##          SOUTH_WT17          SOUTH_WT18          SOUTH_WT19
##                4168                4169                4169
##          SOUTH_WT21          SOUTH_WT22  NORTH_CENTRAL_AWND
##                4118                4169                   0
##  NORTH_CENTRAL_PGTM  NORTH_CENTRAL_PRCP  NORTH_CENTRAL_TAVG
##                3533                   0                1186
##  NORTH_CENTRAL_TMAX  NORTH_CENTRAL_TMIN  NORTH_CENTRAL_WDF5
##                   0                   0                   6
##  NORTH_CENTRAL_WSF5  NORTH_CENTRAL_WT01  NORTH_CENTRAL_WT02
##                   6                3273                4111
##  NORTH_CENTRAL_WT03  NORTH_CENTRAL_WT04  NORTH_CENTRAL_WT05
```

```
##                3591                4148                4082
##  NORTH_CENTRAL_WT06  NORTH_CENTRAL_WT07  NORTH_CENTRAL_WT08
##                4149                4129                3622
##  NORTH_CENTRAL_WT09  NORTH_CENTRAL_WT10  NORTH_CENTRAL_WT11
##                4159                4167                4163
##  NORTH_CENTRAL_WT13  NORTH_CENTRAL_WT14  NORTH_CENTRAL_WT15
##                3955                4110                4167
##  NORTH_CENTRAL_WT16  NORTH_CENTRAL_WT17  NORTH_CENTRAL_WT18
##                3867                4169                4149
##  NORTH_CENTRAL_WT19  NORTH_CENTRAL_WT21  NORTH_CENTRAL_WT22
##                4169                4169                4166
##          COAST_AWND          COAST_PGTM          COAST_PRCP
##                   0                3513                   0
##          COAST_TAVG          COAST_TMAX          COAST_TMIN
##                1186                   0                   0
##          COAST_WDF5          COAST_WSF5          COAST_WT01
##                  28                  28                2569
##          COAST_WT02          COAST_WT03          COAST_WT04
##                3957                3542                4157
##          COAST_WT05          COAST_WT06          COAST_WT07
##                4072                4165                4157
##          COAST_WT08          COAST_WT09          COAST_WT10
##                3799                4167                4169
##          COAST_WT11          COAST_WT13          COAST_WT14
##                4167                3844                4159
##          COAST_WT15          COAST_WT16          COAST_WT17
##                4167                3746                4166
##          COAST_WT18          COAST_WT19          COAST_WT21
##                4168                4169                4144
##          COAST_WT22       FAR_WEST_AWND       FAR_WEST_PGTM
##                4169                   1                3429
##       FAR_WEST_PRCP       FAR_WEST_TAVG       FAR_WEST_TMAX
##                   0                1186                   0
##       FAR_WEST_TMIN       FAR_WEST_WDF5       FAR_WEST_WSF5
##                   0                   8                   8
##       FAR_WEST_WT01       FAR_WEST_WT02       FAR_WEST_WT03
##                3422                4037                3791
##       FAR_WEST_WT04       FAR_WEST_WT05       FAR_WEST_WT06
##                4166                4114                4154
##       FAR_WEST_WT07       FAR_WEST_WT08       FAR_WEST_WT09
##                4046                3342                4156
##       FAR_WEST_WT10       FAR_WEST_WT11       FAR_WEST_WT13
##                4169                4165                4022
##       FAR_WEST_WT14       FAR_WEST_WT15       FAR_WEST_WT16
##                4165                4169                3974
##       FAR_WEST_WT17       FAR_WEST_WT18       FAR_WEST_WT19
##                4166                4151                4160
##       FAR_WEST_WT21       FAR_WEST_WT22           EAST_AWND
##                4169                4156                   1
##           EAST_PGTM           EAST_PRCP           EAST_TAVG
##                3473                   1                4169
##           EAST_TMAX           EAST_TMIN           EAST_WDF5
##                   7                   5                  14
##           EAST_WSF5           EAST_WT01           EAST_WT02
```

```
##                   14                 2946                 4062
##            EAST_WT03            EAST_WT04            EAST_WT05
##                 3630                 4161                 4155
##            EAST_WT06            EAST_WT07            EAST_WT08
##                 4158                 4164                 3797
##            EAST_WT09            EAST_WT10            EAST_WT11
##                 4168                 4169                 4167
##            EAST_WT13            EAST_WT14            EAST_WT15
##                 3817                 4169                 4169
##            EAST_WT16            EAST_WT17            EAST_WT18
##                 3819                 4164                 4156
##            EAST_WT19            EAST_WT21            EAST_WT22
##                 4163                 4169                 4167
##           NORTH_AWND           NORTH_PGTM           NORTH_PRCP
##                    1                 3531                    0
##           NORTH_TAVG           NORTH_TMAX           NORTH_TMIN
##                 1189                    0                    0
##           NORTH_WDF5           NORTH_WSF5           NORTH_WT01
##                   11                   11                 3144
##           NORTH_WT02           NORTH_WT03           NORTH_WT04
##                 4029                 3647                 4160
##           NORTH_WT05           NORTH_WT06           NORTH_WT07
##                 4073                 4131                 4139
##           NORTH_WT08           NORTH_WT09           NORTH_WT10
##                 3763                 4151                 4168
##           NORTH_WT11           NORTH_WT13           NORTH_WT14
##                 4158                 3904                 4154
##           NORTH_WT15           NORTH_WT16           NORTH_WT17
##                 4168                 3893                 4166
##           NORTH_WT18           NORTH_WT19           NORTH_WT21
##                 4144                 4161                 4169
##           NORTH_WT22 ALL_ZONES_Total.AWND ALL_ZONES_Total.PGTM
##                 4158                    0                 3424
## ALL_ZONES_Total.PRCP ALL_ZONES_Total.TAVG ALL_ZONES_Total.TMAX
##                    0                 1186                    0
## ALL_ZONES_Total.TMIN ALL_ZONES_Total.WDF5 ALL_ZONES_Total.WSF5
##                    0                    0                    0
## ALL_ZONES_Total.WT01 ALL_ZONES_Total.WT02 ALL_ZONES_Total.WT03
##                 1143                 3402                 2519
## ALL_ZONES_Total.WT04 ALL_ZONES_Total.WT05 ALL_ZONES_Total.WT06
##                 4096                 3902                 4098
## ALL_ZONES_Total.WT07 ALL_ZONES_Total.WT08 ALL_ZONES_Total.WT09
##                 3928                 1986                 4113
## ALL_ZONES_Total.WT10 ALL_ZONES_Total.WT11 ALL_ZONES_Total.WT13
##                 4165                 4129                 3419
## ALL_ZONES_Total.WT14 ALL_ZONES_Total.WT15 ALL_ZONES_Total.WT16
##                 3980                 4164                 3421
## ALL_ZONES_Total.WT17 ALL_ZONES_Total.WT18 ALL_ZONES_Total.WT19
##                 4145                 4122                 4143
## ALL_ZONES_Total.WT21 ALL_ZONES_Total.WT22
##                 4089                 4143
```

Wow, we have a lot of missing values. Given the similar counts, perhaps we are just missing a few of years' worth of data. Therefore, we may only have complete records for only the last 5 years or so. Further investigation is required to determine the cut off point for the date.