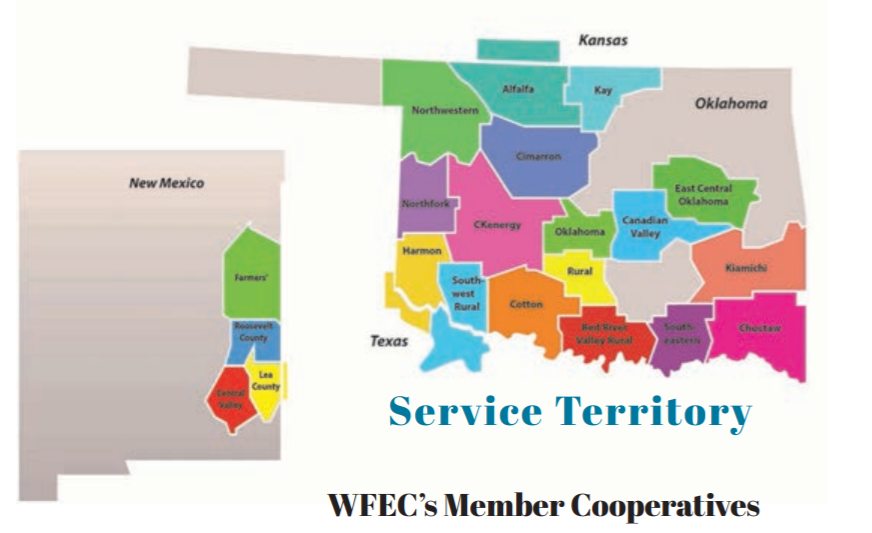
**Geographical and Climate Context**

The territory covered by the Western Farmers Electric cooperative comprises the near entirety of Oklahoma State (OK), a small Kansas (KS) district to the north of Cherokee, small Texas (TX) district to the south of Tipton, and three non-negligible New-Mexico (NM) around Clovis, NM, and Portales, NM. The geography of this territory is dominated by the ecological features characterizing Oklahoma and its great diversity of landscape and climate. It is situated in the Great Plains and borders four primary mountain ranges that are the Arbuckle Mountains, the Wichita Mountains, the Ozark Mountains and the Ouachita Mountains. Given its mountain ranges to the West and the great plains to the East, the WFEC territory slopes downward from its western to its eastern boundaries. WFEC presents semi arid high plains in the northwestern corner. The southwestern part is dotted of partial plains and small mountain ranges like Antelope Hills and the Wichita Mountains. The central portion of WFEC territory is made up of prairies and woodlands. Given its geography between the Great Plains stretching from Canada to the Gulf of Mexico and the mountain ranges bordering the WFEC territory, WFEC’s climate is representative of a diverse climate. It sits at the crossroads of three different air masses: humid air from the Gulf of Mexico, warm and dry hair from Mexico and the SouthWestern U.S, and cold dry air from Canada. WFEC’s region can see frequent air mass changes, that produce dramatic change in both temperature and humidity, during the same season and sometimes, even, the same day. The territory is subjected to extreme meteorological conditions, from temperature, wind, rainfall, and drought. WFEC areas are also characterized by frequent interaction between cold and warm air masses, producing severe conditions that lead to the highest of tornadoes between April and June. Due to its location as an intersection between zones of differing temperature and winds, the weather patterns in the territory can change significantly between relatively short distances. Precipitation levels are lowest in winter months and then rise dramatically from March to a peak in May (owing to frequent thunderstorm activity) before decreasing by mid summer when long periods of dry weather occur in July and August. Precipitation then reach a second yearly maximum from early to mid fall before decreasing from October to December. The eastern part of the WFEC region has a humid subtropical climate that is heavily influenced the moistures from the Gulf of Mexico with hot and humid summers with mild winters. The western part is a semi arid zone where a drier climate prevails, with cold winters and hot summers with much lower humidity. As a general observation, Precipitation and temperatures fall from east to the west.

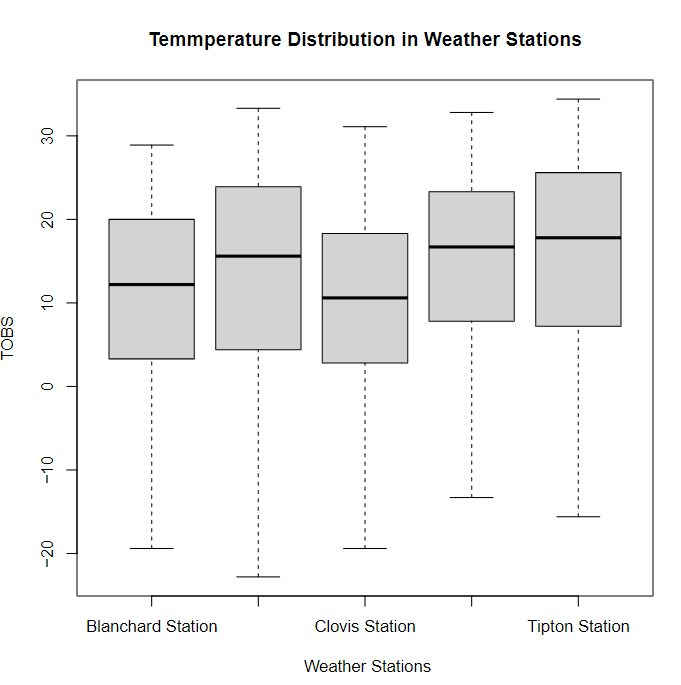
**WFEC Energy Supply**

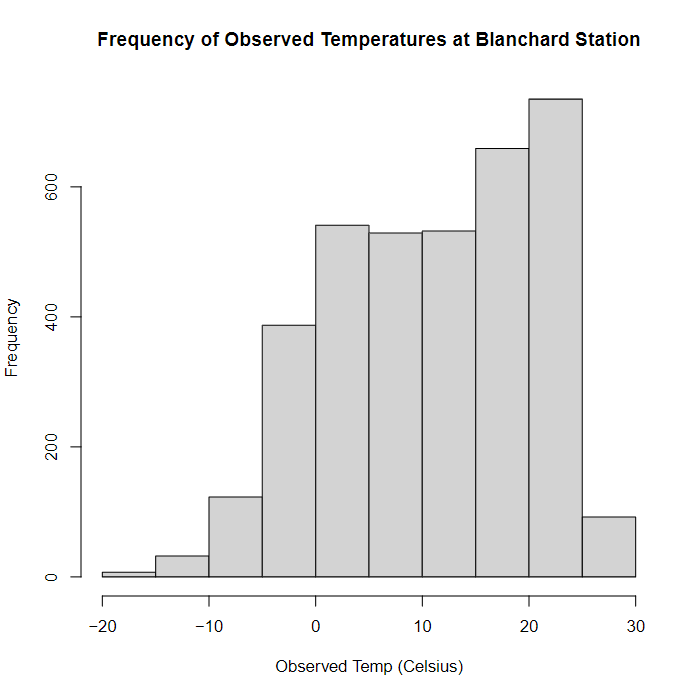
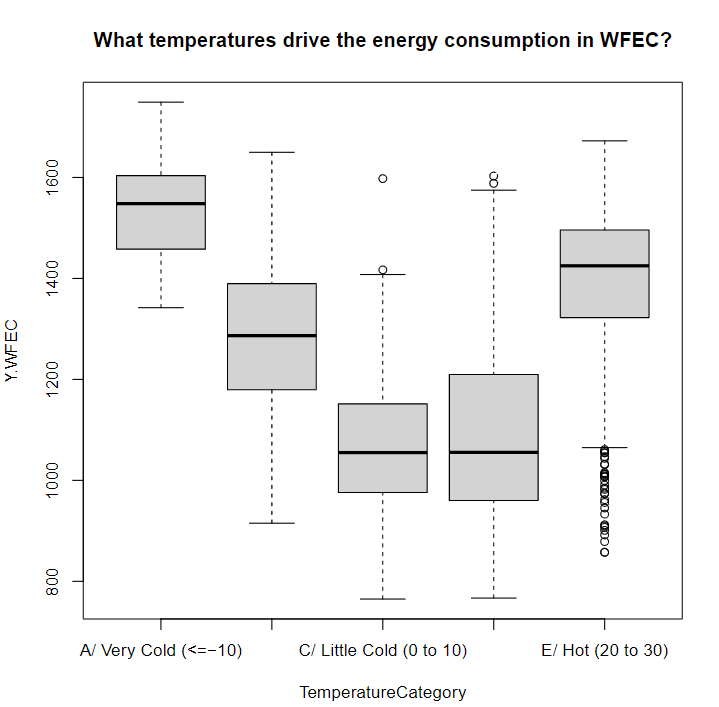
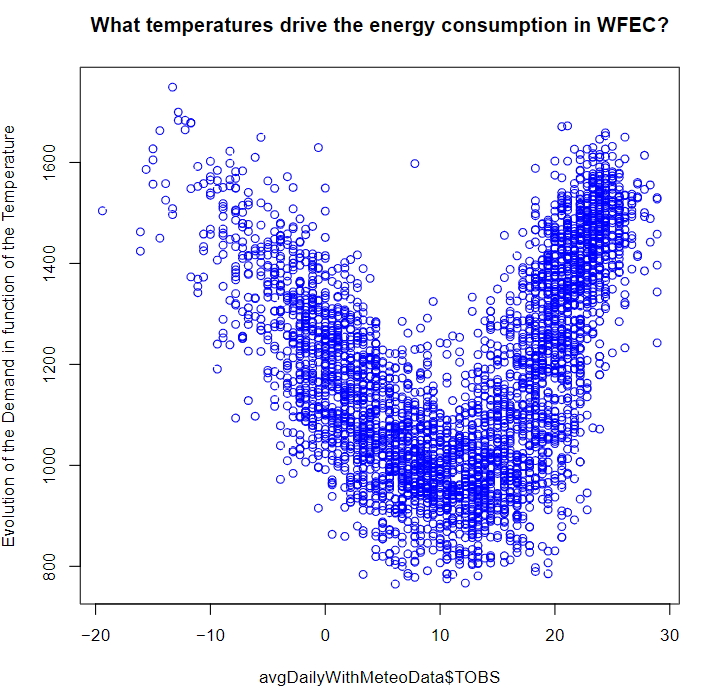
The energy mix of WFEC is made up of power purchase agreement, hydropower, wind energy, and solar. It maintains a diversified portfolio of generation resources that includes owned facilities and capacity. Several technologies and fuel types are part of that mix and ensure that WFEC can flexibly adjust the demand to its supply. As we can also see on WFEC website: “ The generation and transmission (G&T) cooperative's energy resources include a coal-based power plant, as well as numerous natural gas units. Wind resources, solar energy and hydropower allocations also make up a part of the overall energy mix, as well. “

**Data Wrangling**

Once the *rdata* is read thanks to the *load* function, we create a data frame and slice the data frame based on the WFEC region. Then we inspect the data and search for missing values. We discover the Y.WFEC target variable (that represents energy demand in WFEC region) has 6 missing values. It has one missing value on Sep 28 2017 at midnight (GMT time) and 5 missing values on Dec 12 2018 from 8am to 1pm (GMT time). After applying the Central American time zone to the date time variable of the data frame, we then impute the missing values by calculating its average value (at that specific hour where the value is missing) over the last 3 previous years. Finally, the aggregation of the hourly data into daily peaks is operated by using the *aggregate* function on the demand variable and the datetime variable (YYYY-mm-dd HH:MM:SS) that has been converted as a date variable instead (YYYY-mm-dd instead). The aggregating is then saved into the current directory under the data/ folder.

**Data Exploration of Potential Explanatory variables with Y.WFEC**

For the data exploration, we decided to extract daily meteorological summaries (wind, temperature, snowfall, rainfall) from the NOAA website at <https://www.ncdc.noaa.gov/cdo-web/search> for the 2011 through 2021 period for specific weather stations situated in the WFEC territory, that is stations located in different parts of Oklahoma, Kansas, New Mexico, and Texas. We also extracted the daily OK WTI Cushing Oil Price per Barel for our period on the EIA site at <https://www.eia.gov/dnav/pet/hist/RWTCD.htm>. We extracted the daily summaries because meteorological data are the first driver of energy demand, especially for WFEC where industrial and agricultural clients dominate. Daily Oil Price per Barel (WTI Cushing) is used because it could serve as a proxy variable for the economic health of our region, which is usually correlated with the energy demand variable, Y.WFEC. Further, given the oil and gas industry in the rural belt of Oklahoma (main parts of WFEC) represents a significant portion of the economic activity and thus a significant part of the energy demand, we also used the daily oil price per barrel.

Choosing the right weather station or a combination of the weighted average of the weather stations meteorological summaries is then the next challenge, given the large size of the WFEC region and the diversity of its climate. After considering several stations that span the different locations of WFEC, we explored whether the stations has drastic differences in their temperature distribution. As per the box plot below, we notice the distributions of observed temperature are relatively similar. The average and median are relatively similar. However, only the Blanchard station has close to complete data with only 2 missing rows. Given its central position in WFEC and a relative temperature distribution similarity to other stations, the Blanchard Weather Stations is chosen only. Once the proper manipulation are completed, we then merge the daily meteorological summaries to daily oil WTI oil prices. First, we want to see initial relationship between energy demand and observed temperature TOBS. After binning temperatures in intervals of 10 degrees from -20 degree Celsius to 30+ degree Celsius, we see, as showed in the box plot below, that the energy demand is the lowest in the 0 to 10 and 10 to 20 degree Celsius distribution and the highest when the temperature is below 0 degree Celsius or over 20 degrees. This is potentially the sign of a heating and cooling factor respectively amongst WFEC clients. That trend is confirmed with the effective temperature. This is also confirmed by the scatter plot below.

We also observe the following distribution for the observed temperature, rain, and oil prices over the 2011 – 2021 period.

Observed Temperature

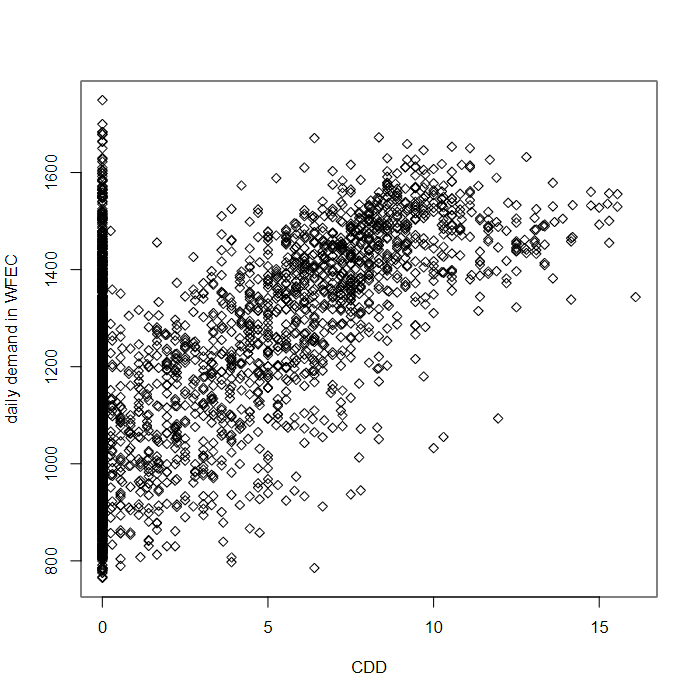
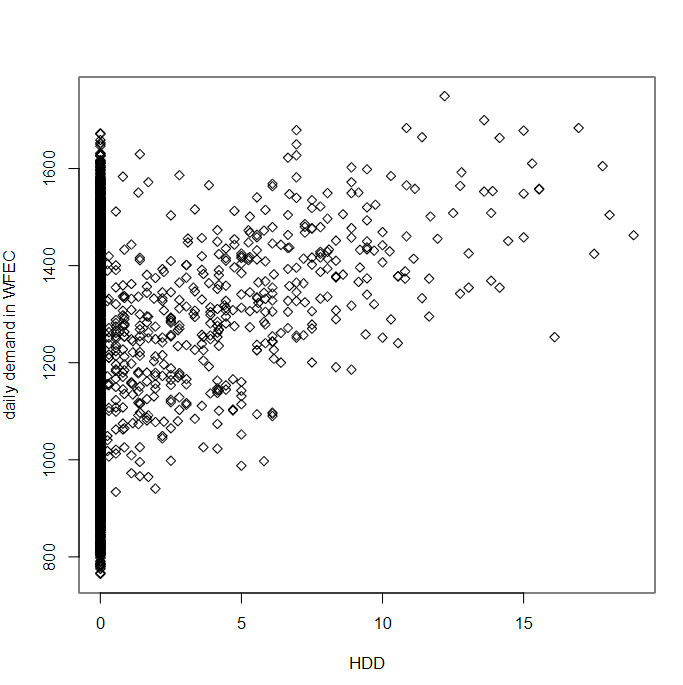
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Min | 1st Quartile | Median | Mean | 3rd Quartile | Max | NA’s |
| -19.40 | 3.30 | 12.20 | 11.34 | 20.00 | 28.90 | 17 |

Rainfall

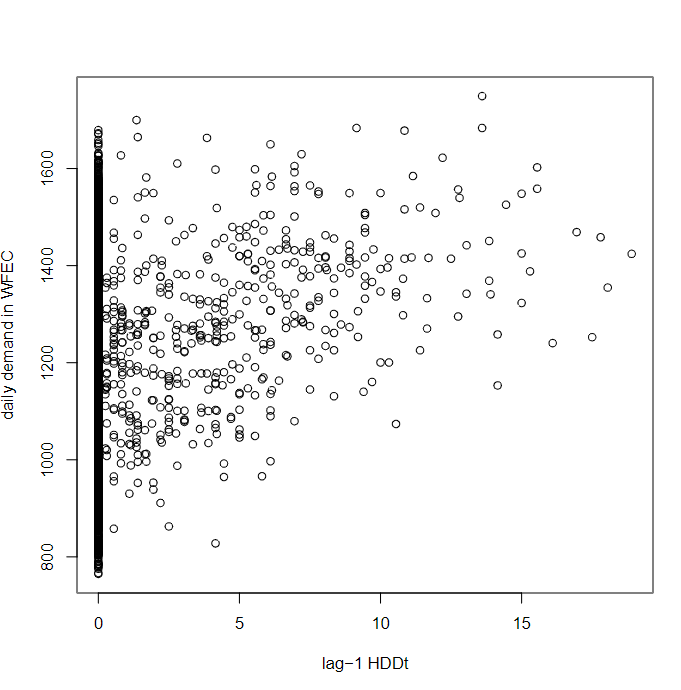
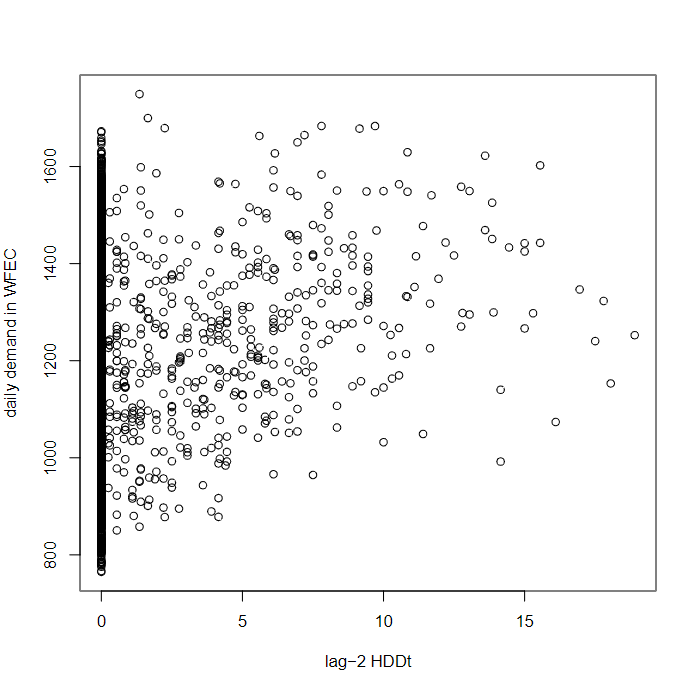
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Min | 1st Quartile | Median | Mean | 3rd Quartile | Max | NA’s |
| 0 | 0 | 0 | 2.547 | 0 | 158.2 | 15 |

Oil price (1141 missing data because no data on Saturday and Sundays)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Min | 1st Quartile | Median | Mean | 3rd Quartile | Max | NA’s |
| -36.98 | 48.49 | 60.88 | 68.45 | 93.49 | 113.39 | 1141 |

We then searched for the most appropriate temperature of reference for HDD and CDD. After looking for the temperature, on the cold and warm ends of the temperature-energy demand relationship, that would make energy demand significantly increase, we concluded that the respective Tref for HDD and CDD are 5 and 20 degree Celsius. The following charts show us that there is a cooling and heating effect on the energy demand with a clear upward relationship for both.

The lag -1 and lag-2 HDD and CDD charts show us there is also a cooling and heating effect with a lag of 1 and 2 days prior the effective date!

HDD lag effect

CDD lag Effect

