Key Factors Driving Load Growth in PJM and ERCOT

https://github.com/jazpritch/KarneiKogaWargoPritchett_ENV872_EDA_FinalProject

Hanna Karnei, Haru Koga, Jazmine Pritchett, Jaimie Wargo

Contents

Rationale and Research Questions **3**

Dataset Information **4**

Exploratory Analysis of Raw Data **5**

Regression Analysis **16**

PJM Simple Linear Regression 16

ERCOT Multiple Linear Regression 19

Summary and Conclusions **25**

References **26**

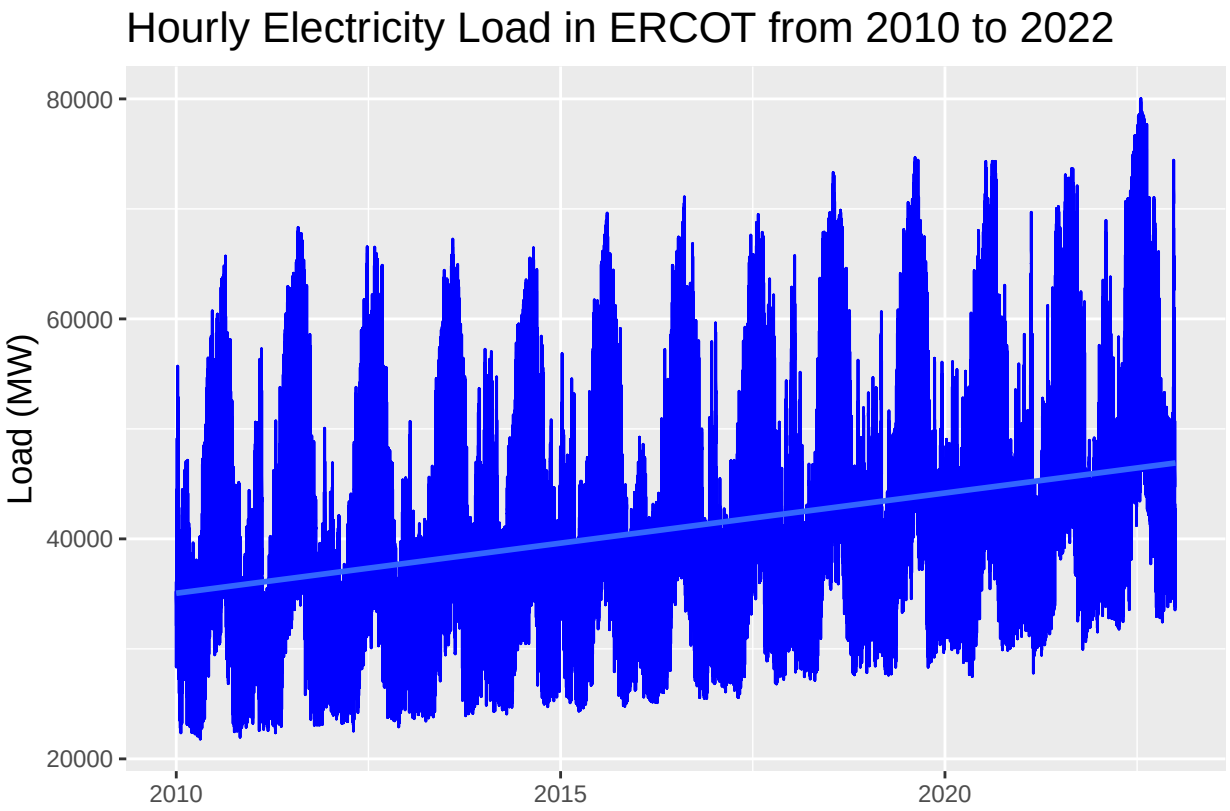
Rationale and Research Questions

Dataset Information

Exploratory Analysis of Raw Data

Distribution of ERCOT hourly load data, as well as aggregated by day and month.

```
## # A tibble: 6 x 2
##   Hour_Ending  ERCOT
##   <date>       <dbl>
## 1 2010-01-01  32094.
## 2 2010-01-01  32171.
## 3 2010-01-01  32242.
## 4 2010-01-01  32459.
## 5 2010-01-01  33124.
## 6 2010-01-01  34315.
```

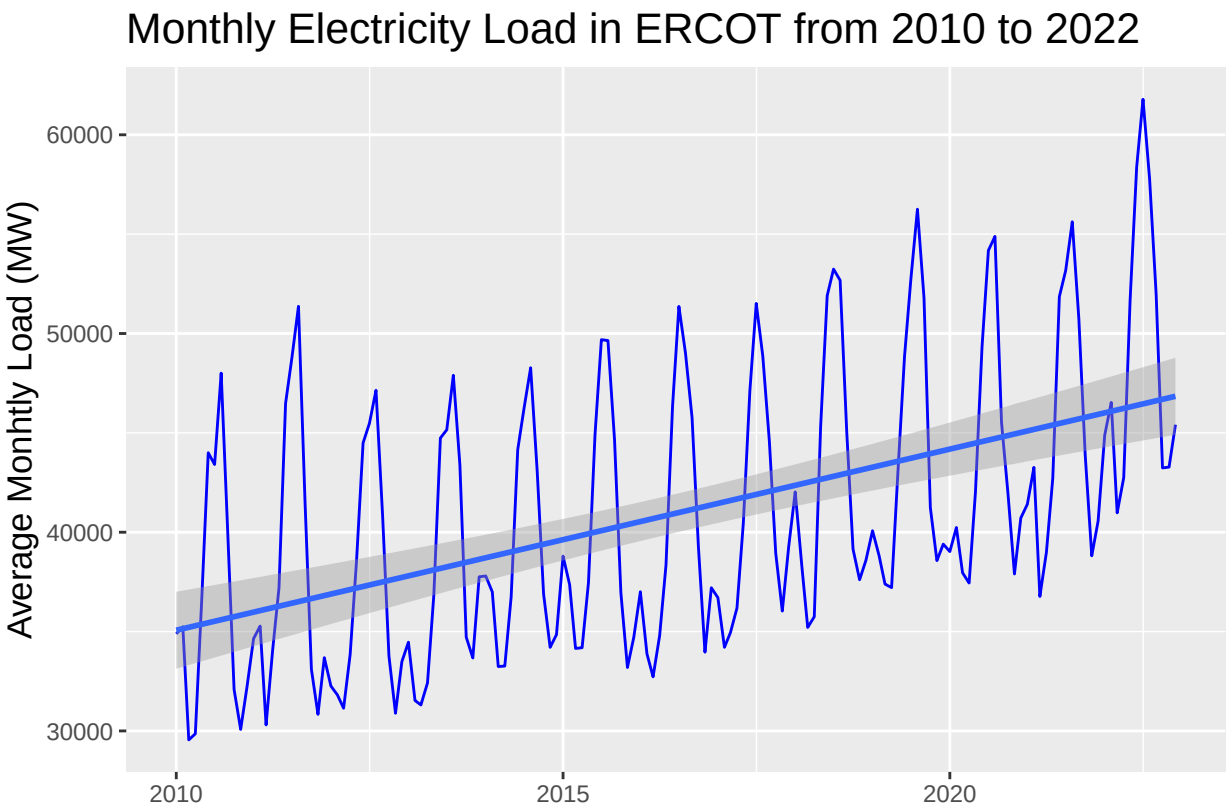


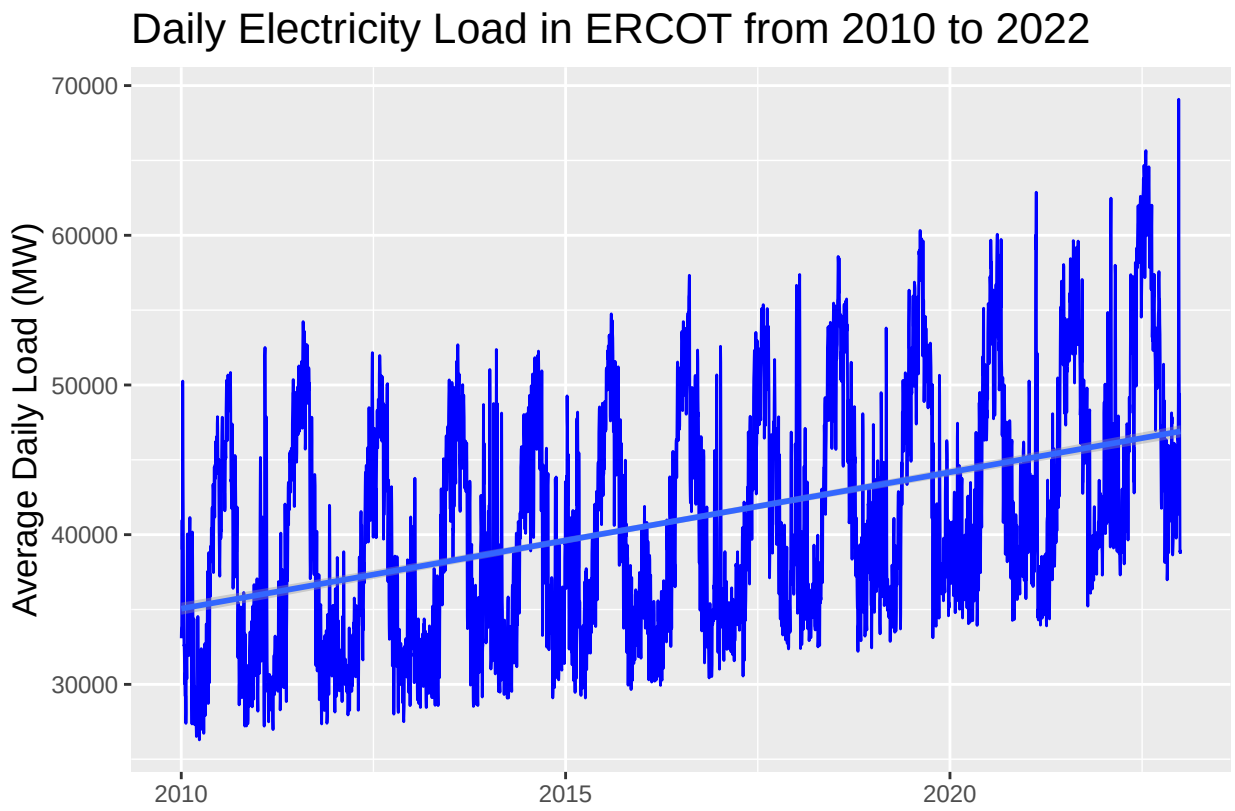

```
## # A tibble: 13 x 2
##     year Mean_Load
##    <dbl>     <dbl>
##  1  2010    36336.
##  2  2011    38127.
##  3  2012    37000.
##  4  2013    37869.
##  5  2014    38828.
##  6  2015    39669.
##  7  2016    40006.
##  8  2017    40782.
```

```
##  9  2018    42949.
## 10  2019    43818.
## 11  2020    43463.
## 12  2021    44829.
## 13  2022    49074.
```

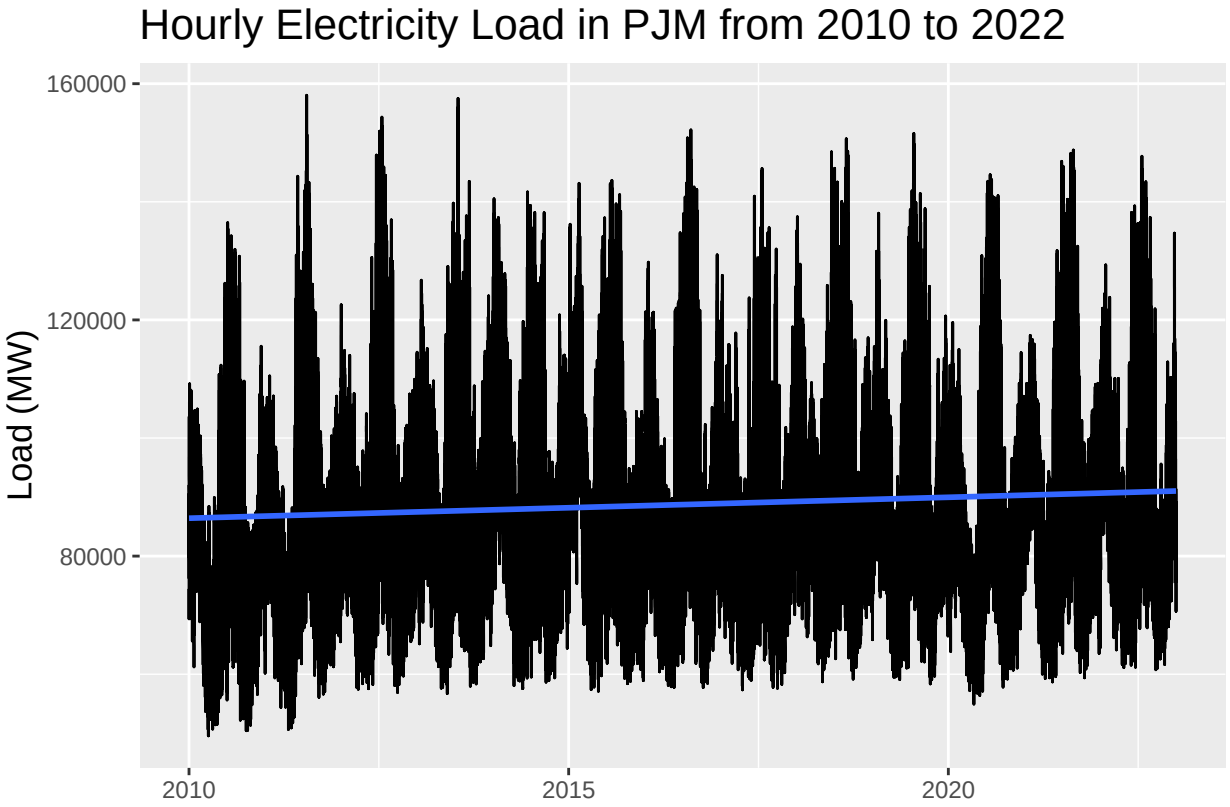
```
## # A tibble: 6 x 4
## # Groups:   Year [1]
##    Year Month Mean_Load NewDate
##   <dbl> <dbl>     <dbl> <date>
## 1  2010     1    34884. 2010-01-01
## 2  2010     2    35240. 2010-02-01
## 3  2010     3    29550. 2010-03-01
## 4  2010     4    29856. 2010-04-01
## 5  2010     5    36547. 2010-05-01
## 6  2010     6    44000. 2010-06-01
```

```
## # A tibble: 6 x 5
## # Groups:   Year, Month [1]
##    Year Month   Day Mean_Load NewDate
##   <dbl> <dbl> <int>     <dbl> <date>
## 1  2010     1     1    33080. 2010-01-01
## 2  2010     1     2    33780. 2010-01-02
## 3  2010     1     3    33892. 2010-01-03
## 4  2010     1     4    38857. 2010-01-04
## 5  2010     1     5    40971. 2010-01-05
## 6  2010     1     6    39005. 2010-01-06
```


Monthly Electricity Load in ERCOT from 2010 to 2022
Average Monhtly Load (MW)
60000
50000
40000
30000
2010
2015
2020

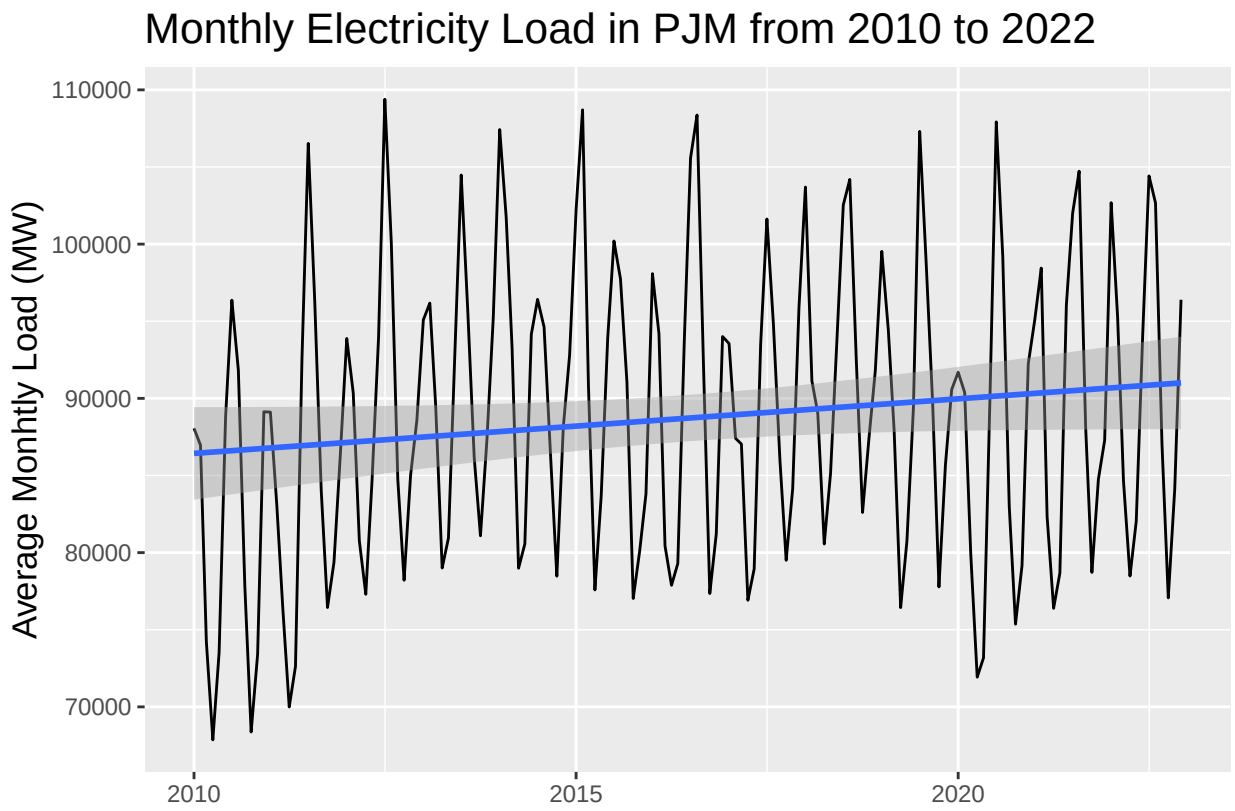


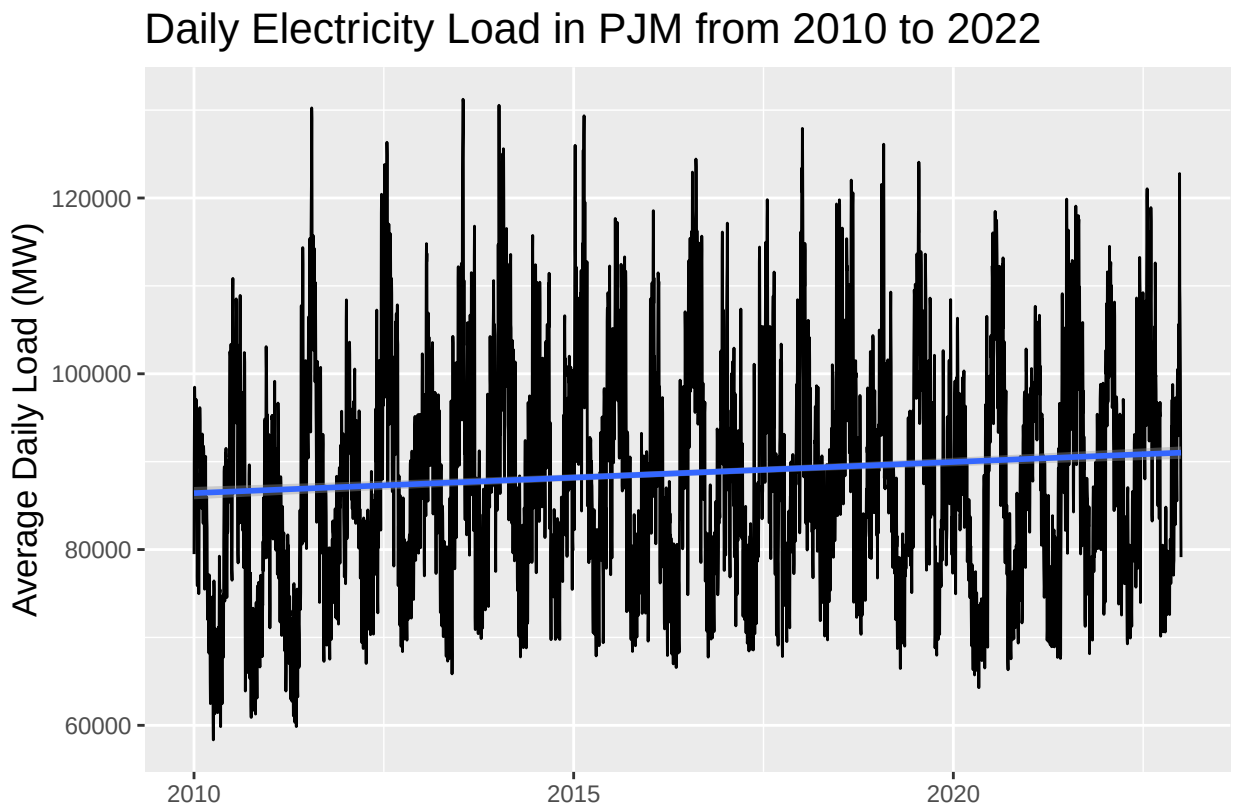
Distribution of PJM hourly load data, as well as aggregated by day and month.



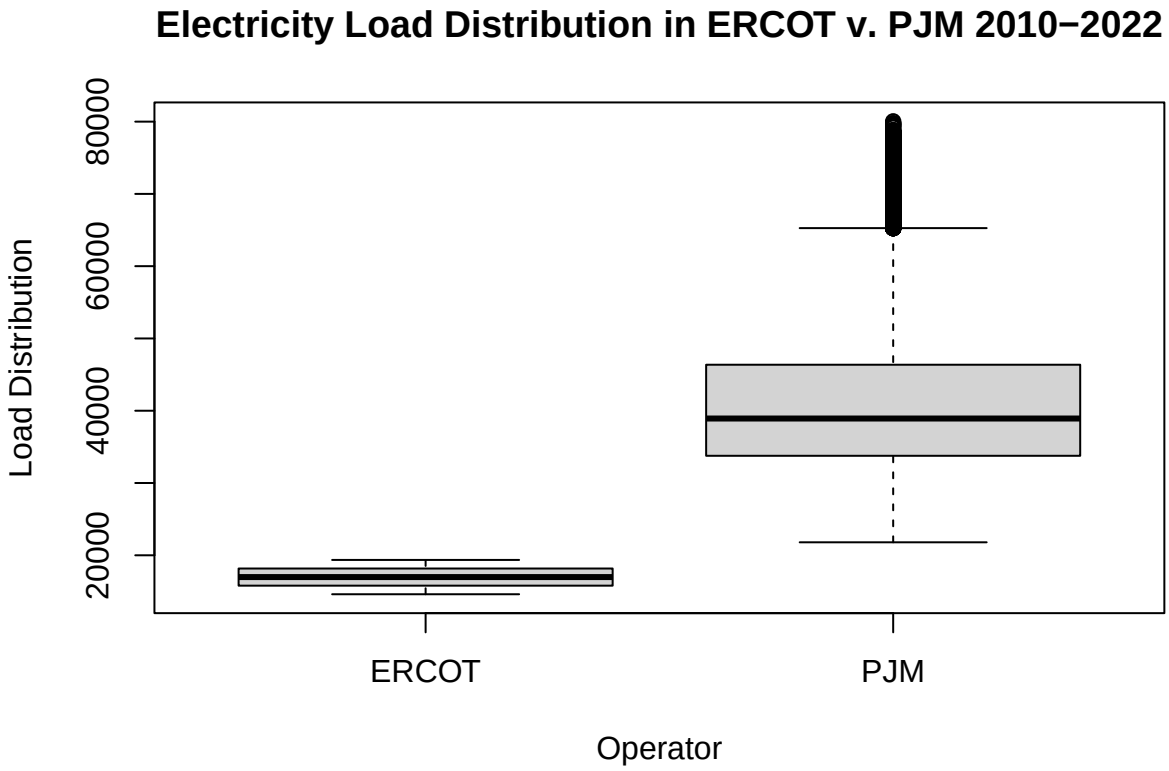

```
## # A tibble: 6 x 4
## # Groups:   Year [1]
##    Year Month Mean_Load NewDate
##   <dbl> <dbl>     <dbl> <date>
## 1  2010     1    88058. 2010-01-01
## 2  2010     2    86967. 2010-02-01
## 3  2010     3    74220. 2010-03-01
## 4  2010     4    67861. 2010-04-01
## 5  2010     5    73517. 2010-05-01
## 6  2010     6    88797. 2010-06-01
```

```
## # A tibble: 6 x 5
## # Groups:   Year, Month [1]
##    Year Month   Day Mean_Load NewDate
##   <dbl> <dbl> <int>     <dbl> <date>
## 1  2010     1     1    79483. 2010-01-01
## 2  2010     1     2    89741. 2010-01-02
## 3  2010     1     3    95061. 2010-01-03
## 4  2010     1     4    98474. 2010-01-04
## 5  2010     1     5    97021. 2010-01-05
## 6  2010     1     6    95613. 2010-01-06
```


Monthly Electricity Load in PJM from 2010 to 2022
Average Monhtly Load (MW)
110000
100000
90000
80000
70000
2010
2015
2020

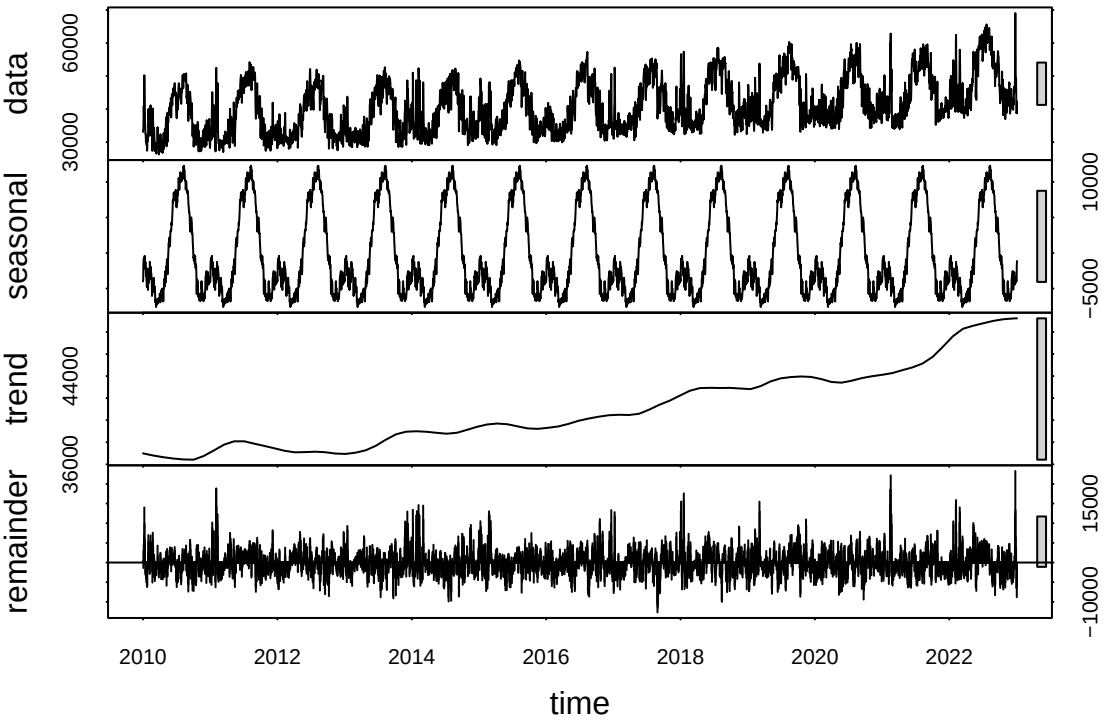


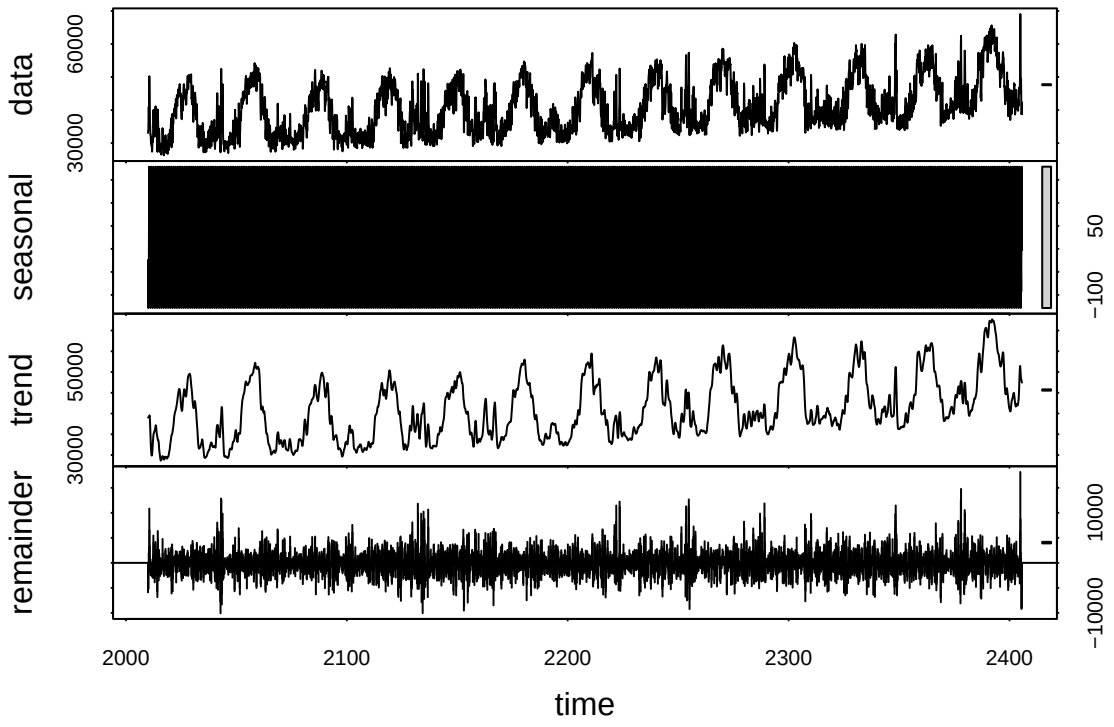
Using a boxplot to compare electricity load distribution between ERCOT versus PJM.


Electricity Load Distribution in ERCOT v. PJM 2010−2022
Load Distribution
20000
40000
60000
80000
ERCOT
PJM
Operator

#Time Series Analysis

The null hypothesis: Electricity load in ERCOT remained the same from 2010 to 2022. Alternative hypothesis: Electricity load in ERCOT has changed from 2010 to 2022.





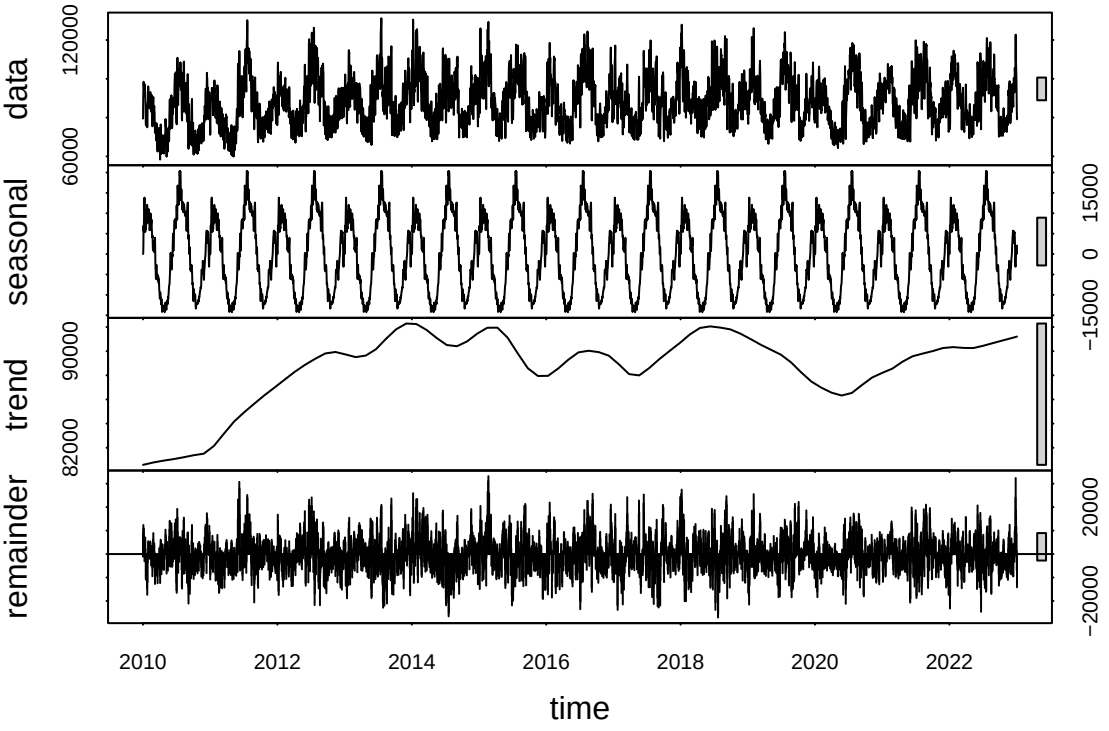

```
## Score =  287577 , Var(Score) = 82954344
## denominator =  937335.1
## tau = 0.307, 2-sided pvalue =< 2.22e-16
```

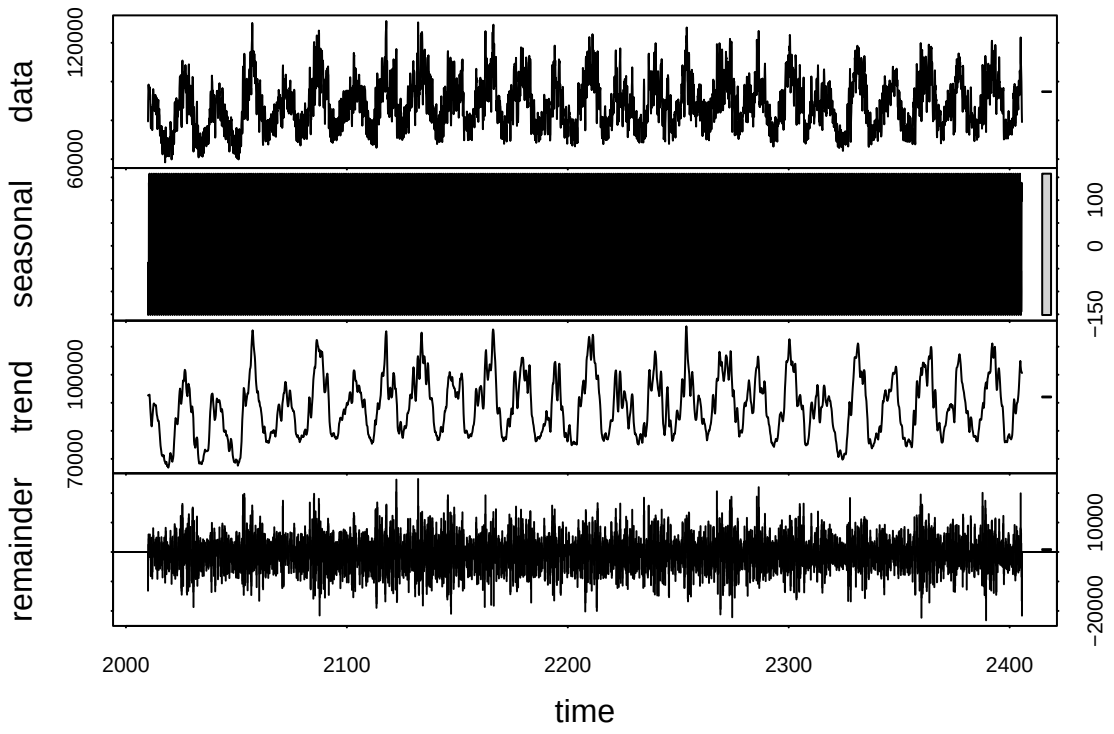
The seasonal Mann-Kendall is the most appropriate test because it is designed to analyze seasonal, non-parametric data, and there is a clear seasonal trend in our data. All other tests we learned about could only be used for non-seasonal data. A positive score (361963) and tau (0.341) indicate an increasing trend in the time series data. The p-value smaller than 0.05 allows us to reject the null hypothesis. Therefore, we can conclude that electricity demand has changed over time in ERCOT.

```
## Score =  3444826 , Var(Score) = 11904225280
## denominator =  11274125
## tau = 0.306, 2-sided pvalue =< 2.22e-16
```

Based on the results of the Mann Kendall test, we are able to reject the null hypothesis ($p<0.05$) and identify an increasing trend in the data (tau=0.34, score=4339158). Hence, the load in ERCOT has indeed changed over time; it was not simply a seasonal change. A p-value <0.05 in the seasonal Mann Kendall test also led us to the same conclusion.

The null hypothesis: Electricity load in PJM remained the same from 2010 to 2022. Alternative hypothesis: Electricity load in PJM has changed from 2010 to 2022.






```
## Score =  63799 , Var(Score) = 82902071
## denominator =  936939.6
## tau = 0.0681, 2-sided pvalue =2.4352e-12
```

The seasonal Mann-Kendall is the most appropriate test because it is designed to analyze seasonal, non-parametric data, and there is a clear seasonal trend in our data. All other tests we learned about could only be used for non-seasonal data. A positive score (82902071) and tau (0.0681) indicate an increasing trend in the time series data. The p-value smaller than 0.05 allows us to reject the null hypothesis. Therefore, we can conclude that electricity demand has changed over time in PJM.

```
## Score =  774030 , Var(Score) = 11896706048
## denominator =  11269379
## tau = 0.0687, 2-sided pvalue =< 2.22e-16
```

Mann Kendall test rejects the null hypothesis, which indicates that there is an increasing trend in the data (tau=0.0687, score=11896706048). This means that PJM load has changed over time from 2010 to 2022 without seasonality components. A p-value <0.05 in the seasonal Mann Kendall test also led us to the same conclusion.

Regression Analysis

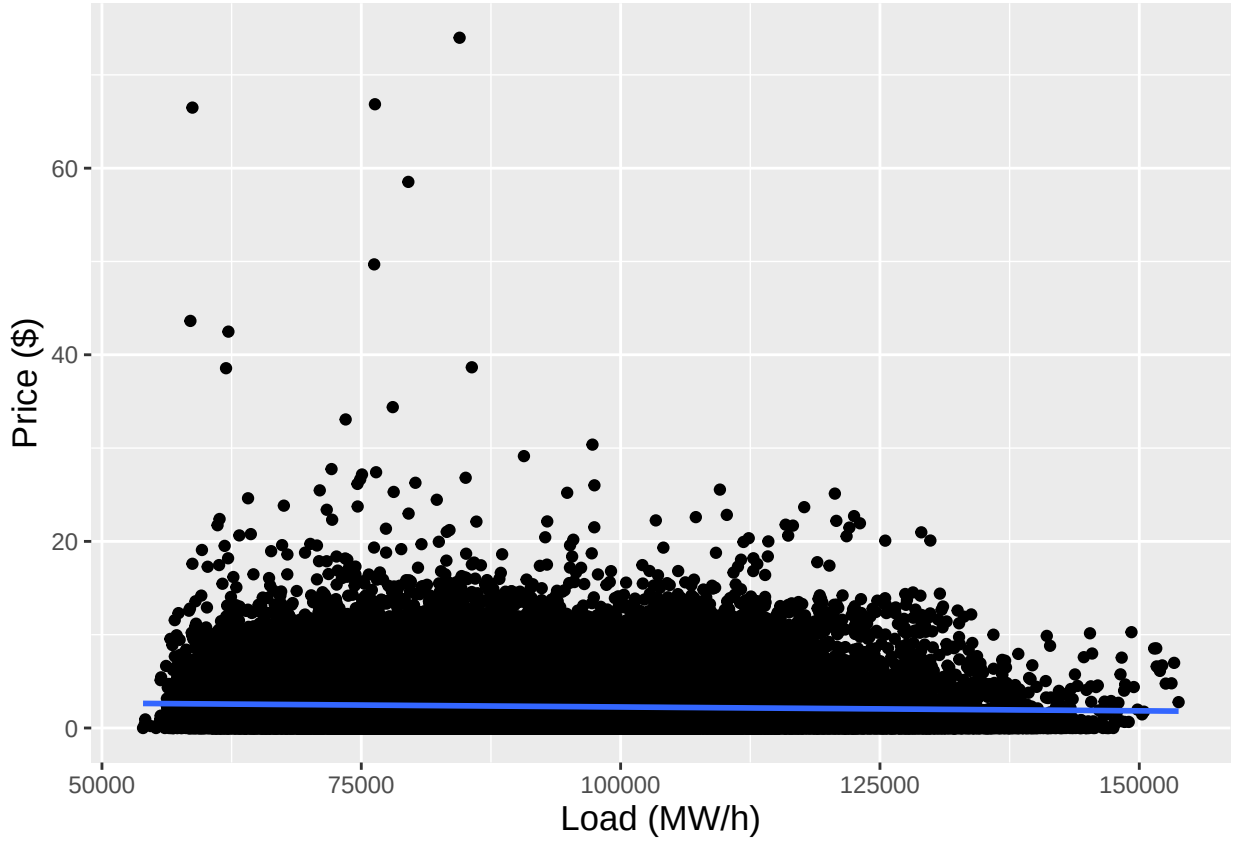
PJM Simple Linear Regression

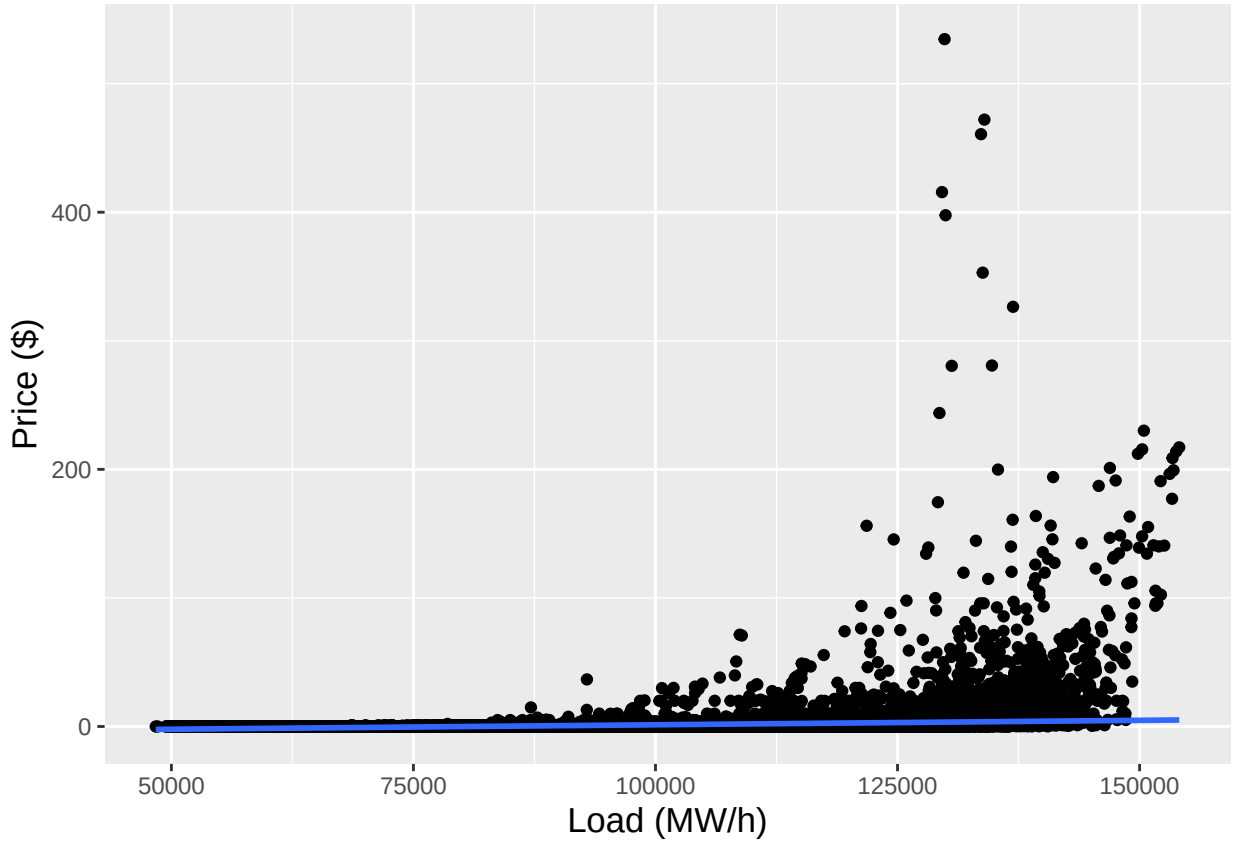
Simple Linear Regression

The null hypothesis: There is no change in electricity load with price in PJM from years...

Alternative hypothesis: Electricity load in PJM has changed in relation to price from years...

Real-time pricing reflects the actual market conditions, while day-ahead pricing provides insight into the expected future market conditions. Both pricing data used to understand regression analysis. Market Performance Clearing Price ($/MWh)used reflecting the cost of the movement of load in response to regulation signals, which is related to the performance of the resource in meeting the demand.






```
##
## Call:
## lm(formula = Price ~ Load_MW, data = rt_load)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -2.632 -1.810 -0.975  1.017 71.599
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.078e+00  5.320e-02   57.85   <2e-16 ***
## Load_MW     -8.254e-06  5.960e-07  -13.85   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.711 on 81086 degrees of freedom
##   (8760 observations deleted due to missingness)
## Multiple R-squared:  0.002359,   Adjusted R-squared:  0.002347
## F-statistic: 191.8 on 1 and 81086 DF,  p-value: < 2.2e-16

##
##  Pearson's product-moment correlation
##
## data:  rt_load$Price and rt_load$Load_MW
## t = -13.848, df = 81086, p-value < 2.2e-16
```

```
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.05543753 -0.04170422
## sample estimates:
##         cor
## -0.04857317
```

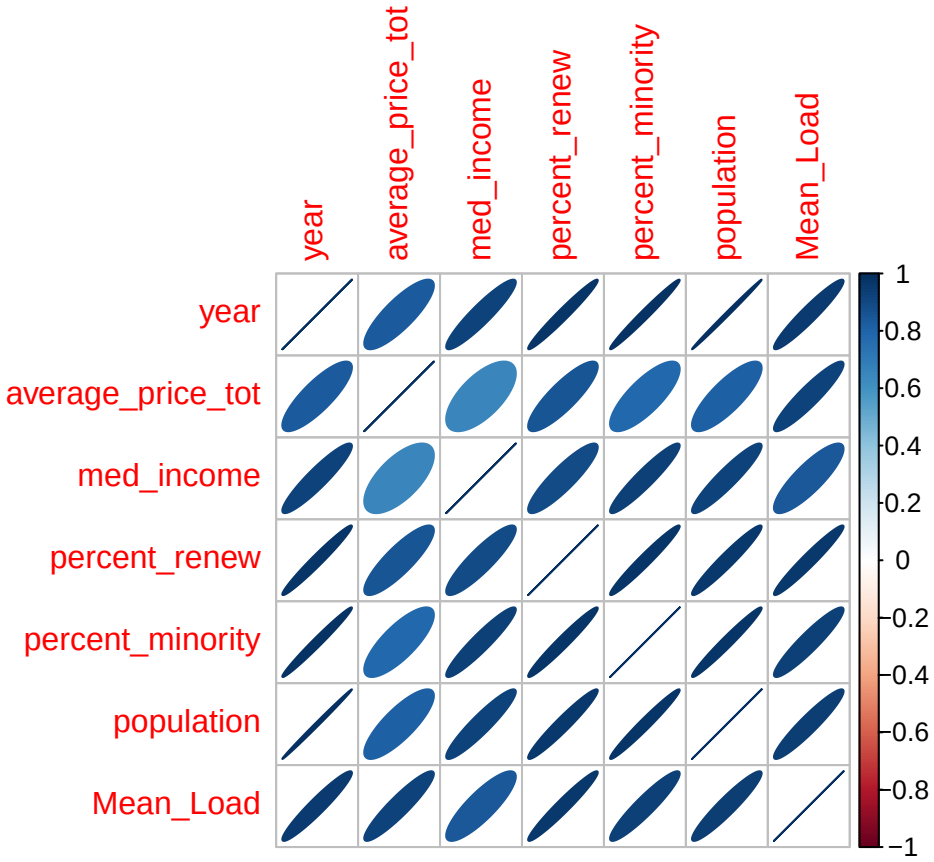
```
##
## Call:
## lm(formula = Price ~ Load_MW, data = dasr_load)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -4.07  -0.95  -0.21   0.45 531.28
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.607e+00  9.390e-02  -59.71   <2e-16 ***
## Load_MW      6.921e-05  1.061e-06   65.22   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.777 on 111728 degrees of freedom
## Multiple R-squared:  0.03668,    Adjusted R-squared:  0.03667
## F-statistic:  4254 on 1 and 111728 DF,  p-value: < 2.2e-16
```

```
##
##  Pearson's product-moment correlation
##
## data:  dasr_load$Price and dasr_load$Load_MW
## t = 65.224, df = 111728, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1858631 0.1971601
## sample estimates:
##      cor
## 0.191518
```

ERCOT Multiple Linear Regression

```
##    year average_price_tot med_income percent_renew percent_minority population
## 1  2010              9.83      61680        6.6849             54.7   25241897
## 2  2011              9.90      62080        7.1507             55.1   25645504
## 3  2012              9.84      64480        7.7938             55.6   26084120
## 4  2013             10.07      63070        8.4260             56.0   26479646
## 5  2014             10.44      65170        9.2940             56.0   26963092
## 6  2015             10.41      68360       10.2685             57.0   27468531
## 7  2016             10.27      69770       13.1272             57.0   27914064
## 8  2017             10.48      70860       15.5285             58.0   28291024
## 9  2018             10.53      69100       16.7659             58.4   28624564
## 10 2019             10.54      76820       18.5142             58.8   28986794
## 11 2020             10.59      77110       21.5531             60.3   29232474
## 12 2021             11.10      72680       23.9658             59.7   29558864
```

```
## 13 2022             12.36       74640        26.2289              60.2    30029572
##    Mean_Load
## 1   36335.65
## 2   38127.12
## 3   37000.04
## 4   37869.12
## 5   38828.39
## 6   39669.23
## 7   40006.36
## 8   40781.92
## 9   42949.25
## 10  43817.93
## 11  43463.07
## 12  44828.81
## 13  49074.36
```




```
##
## Call:
## lm(formula = Mean_Load ~ average_price_tot + population + percent_renew +
##     med_income + percent_minority, data = ercot_linear_model_data_totprice)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -749.5 -379.8 -316.2  185.2 1089.5
##
## Coefficients:
```

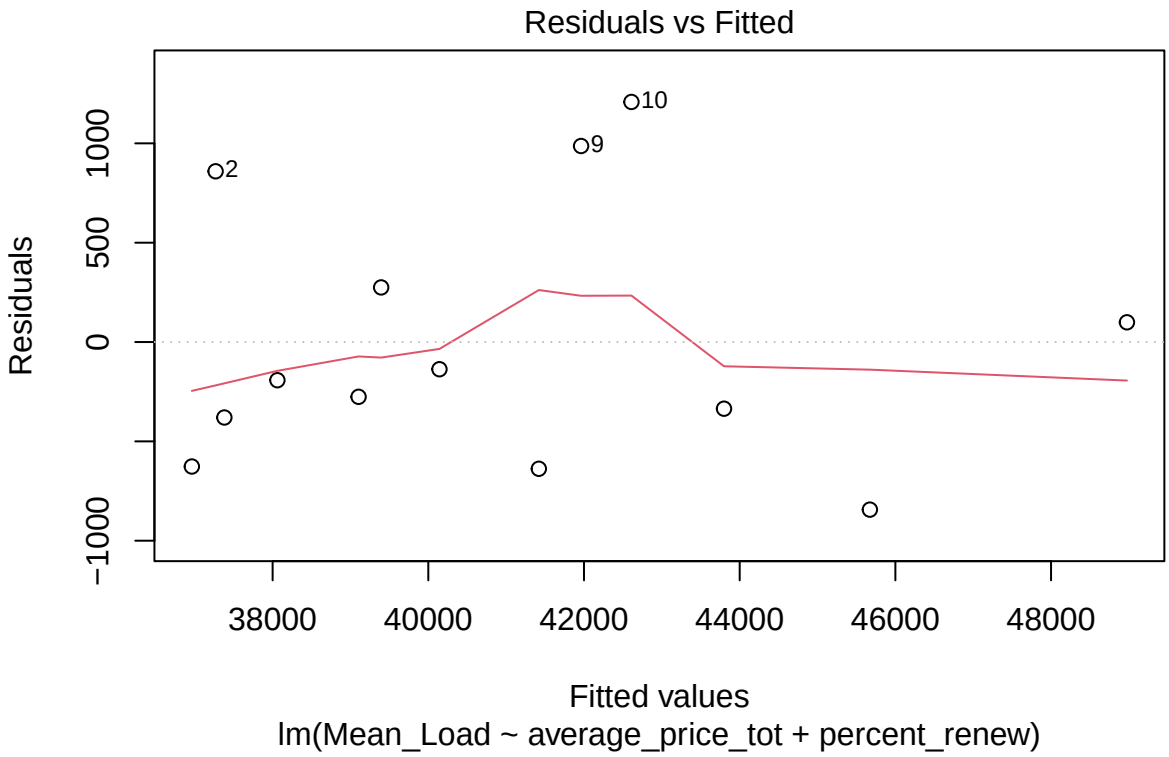
```
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)        8.518e+03  3.809e+04   0.224   0.8295
## average_price_tot  2.154e+03  8.338e+02   2.583   0.0363 *
## population         3.029e-04  8.163e-04   0.371   0.7216
## percent_renew      2.605e+02  1.987e+02   1.311   0.2313
## med_income         7.059e-02  1.438e-01   0.491   0.6386
## percent_minority  -1.237e+02  8.480e+02  -0.146   0.8881
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 807.6 on 7 degrees of freedom
## Multiple R-squared:  0.9716, Adjusted R-squared:  0.9514
## F-statistic: 47.97 on 5 and 7 DF,  p-value: 2.882e-05


## Start:  AIC=178
## Mean_Load ~ average_price_tot + population + percent_renew +
##     med_income + percent_minority
##
##                      Df Sum of Sq     RSS    AIC
## - percent_minority   1     13877 4579225 176.04
## - population         1     89803 4655151 176.25
## - med_income         1    157074 4722422 176.44
## <none>                           4565348 178.00
## - percent_renew      1   1120406 5685754 178.85
## - average_price_tot  1   4352280 8917628 184.70
##
## Step:  AIC=176.04
## Mean_Load ~ average_price_tot + population + percent_renew +
##     med_income
##
##                      Df Sum of Sq     RSS    AIC
## - population         1     77664 4656889 174.26
## - med_income         1    143291 4722516 174.44
## <none>                           4579225 176.04
## - percent_renew      1   1775016 6354241 178.30
## - average_price_tot  1   5080903 9660128 183.74
##
## Step:  AIC=174.26
## Mean_Load ~ average_price_tot + percent_renew + med_income
##
##                      Df Sum of Sq      RSS    AIC
## - med_income         1    500318  5157207 173.58
## <none>                           4656889 174.26
## - percent_renew      1   3026135  7683024 178.76
## - average_price_tot  1   5888552 10545441 182.88
##
## Step:  AIC=173.58
## Mean_Load ~ average_price_tot + percent_renew
##
##                      Df Sum of Sq      RSS    AIC
## <none>                           5157207 173.58
## - average_price_tot  1   5555548 10712755 181.09
## - percent_renew      1  18640994 23798201 191.46
```

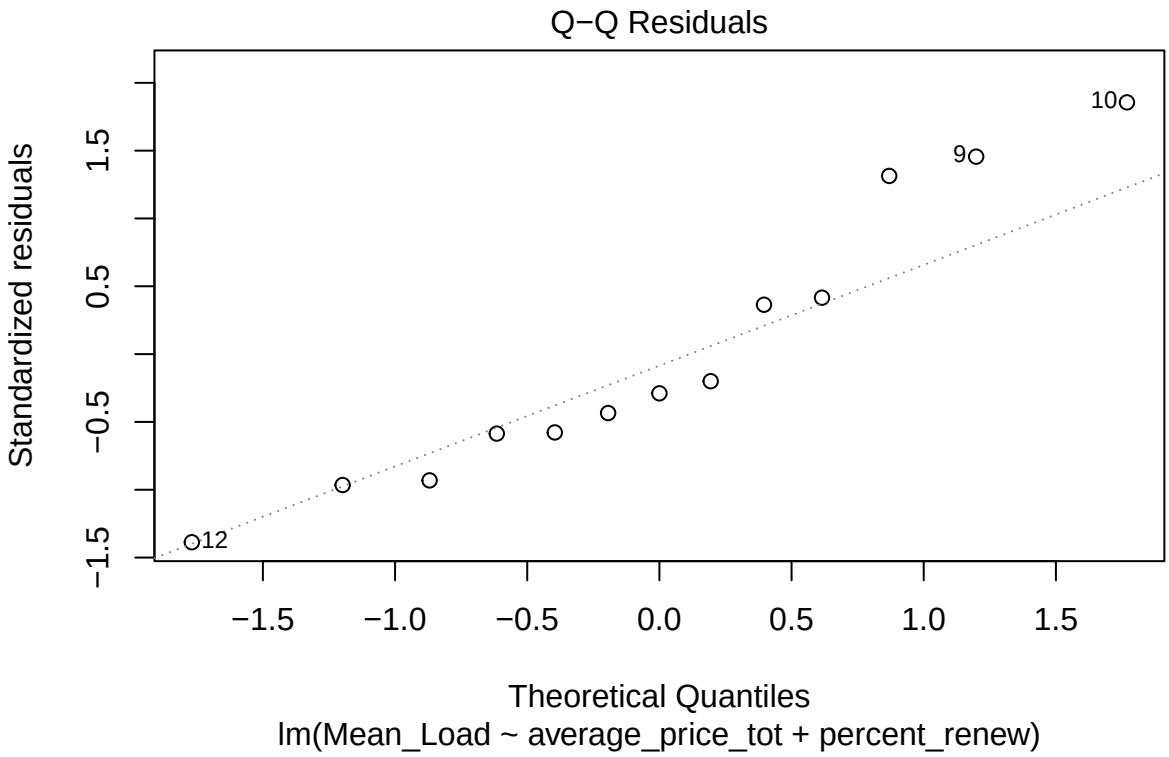
```
##
## Call:
## lm(formula = Mean_Load ~ average_price_tot + percent_renew, data = ercot_linear_model_data_totprice)
##
## Coefficients:
##       (Intercept)  average_price_tot      percent_renew
##           15132.4             1976.8              358.8
```

Variables	AIC Value
Mean_Load + population + percent_renew + med_income + percent_minority	178
average_price_tot + population + percent_renew + med_income	176.04
average_price_tot + percent_renew + med_income	174.26
average_price_tot + percent_renew	173.58

```
##
## Call:
## lm(formula = Mean_Load ~ average_price_tot + percent_renew, data = ercot_linear_model_data_totprice)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -843.6 -379.8 -192.2  274.8 1208.3
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       15132.37    5610.11   2.697  0.02242 *
## average_price_tot  1976.77     602.28   3.282  0.00826 **
## percent_renew       358.75      59.67   6.012  0.00013 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 718.1 on 10 degrees of freedom
## Multiple R-squared:  0.968,  Adjusted R-squared:  0.9616
## F-statistic: 151.1 on 2 and 10 DF,  p-value: 3.373e-08
```

Variable	Coefficients	p-value
Intercept	15,132.37	0.0224
average_price	1,976.77	0.0083
percent_renew	358.75	0.0001


Residuals vs Fitted
Residuals
1000
500
0
−1000
10
9
2
38000
40000
42000
44000
46000
48000
Fitted values
lm(Mean_Load ~ average_price_tot + percent_renew)


Q−Q Residuals
Standardized residuals
−1.5
−0.5
0.5
1.5
10
9
12
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
Theoretical Quantiles
lm(Mean_Load ~ average_price_tot + percent_renew)

Summary and Conclusions

References

<add references here if relevant, otherwise delete this section>