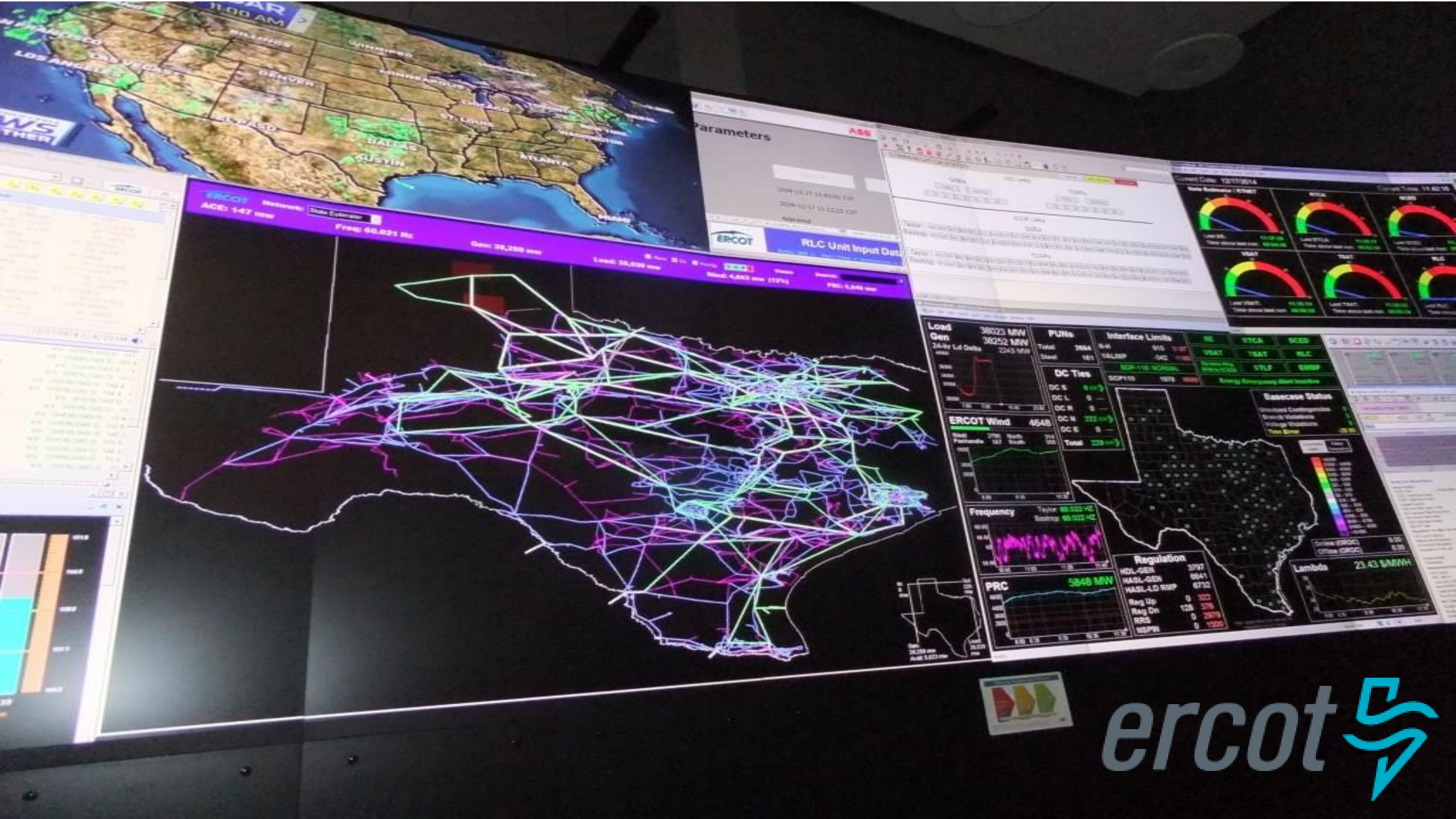
PREDICTING TEXAS ELECTRICITY SPOT PRICES

PAIGE BAILEY
CS 5310 – DATA MINING
APRIL 29 2017


LOS ANGELES
DENVER
DALLAS
AUSTIN
ATLANTA
MIAMI
Parameters
ERCOT
ACE: 147 mw
Freq: 60.021 Hz
Load
Gen
PUNs
Interface Limits
DC Ties
ERCOT Wind
Frequency
PRC
5848 MW
Regulation
Basecase Status
Lambda
23.43 $/MWH
ercot

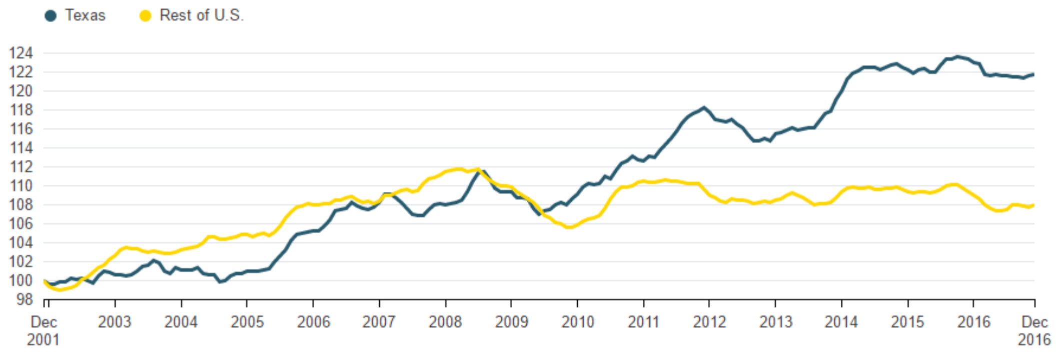
MOTIVATION

BRIEF

ERCOT: Narrow reserves boost real-time prices above $700/MWh

Lone Star State

Texas' electricity demand kept growing after 2008, unlike the U.S. overall



Source: Energy Information Administration

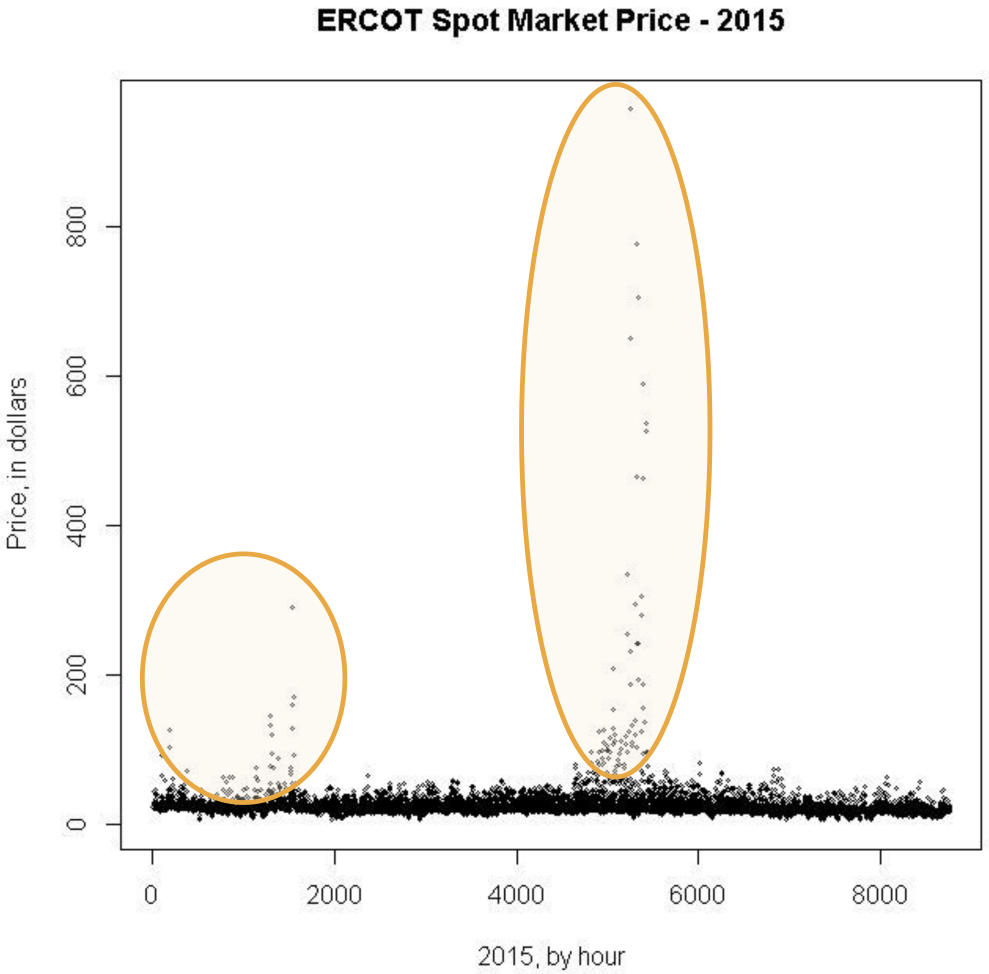
Note: Trailing 12-month retail sales of electricity, indexed to 100 in December 2001.

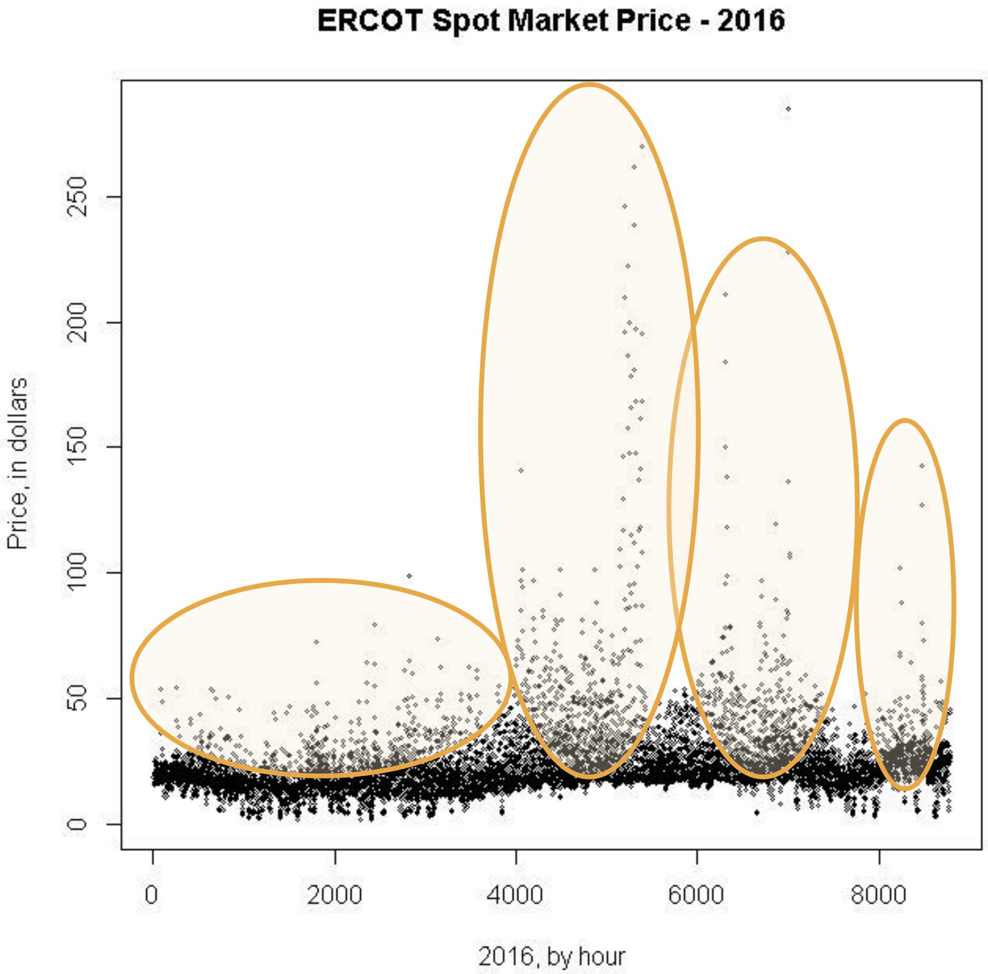
from the Houston Chronicle:

By late afternoon on February 22, 2017, temperatures around [Texas] had climbed into the 90s, and the demand for power surged. **In Houston, where midday wholesale power prices typically hover around $25, the price spiked to $4,000 per megawatt-hour** - a sign that the market was short of supply.

Houston has a capacity problem, and energy demand is only anticipated to increase.

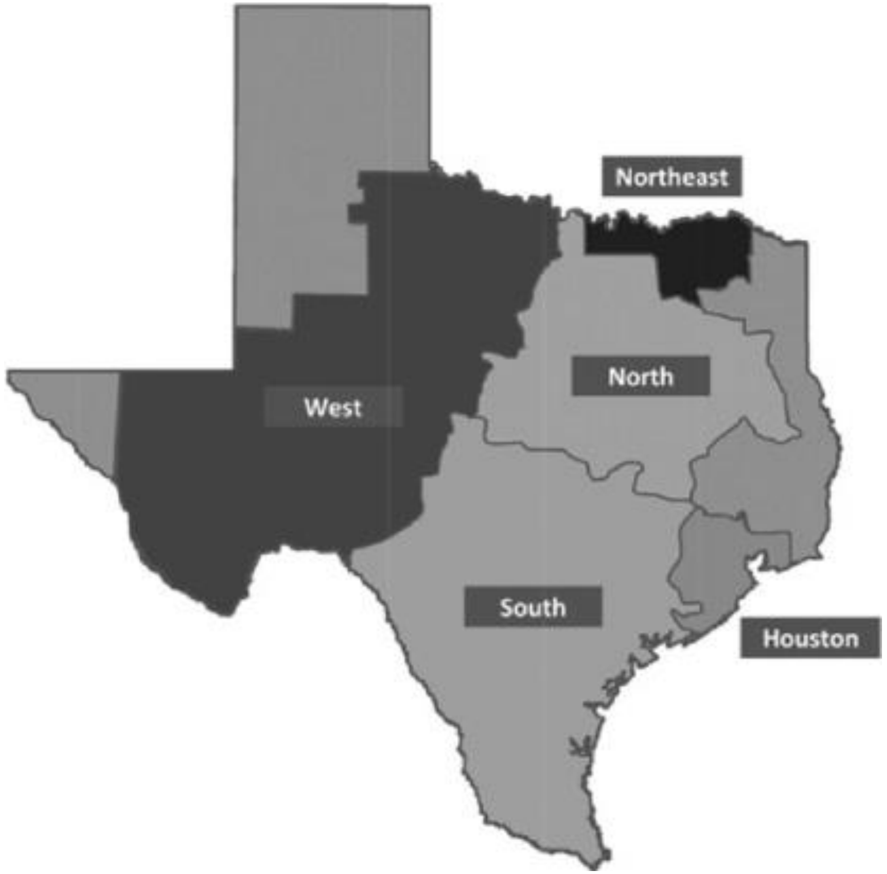
VARIABILITY IN ELECTRICITY PRICES





RESEARCH QUESTIONS

- How do **weather patterns** in Houston (humidity, precipitation, temperature) impact price of electricity for the Houston hub?
- What other **factors or combination of factors** impact electricity prices in the Houston hub?
- What are the **optimal times** to purchase electricity at the Houston hub? To sell?
- How could you **operationalize the purchase** of electricity for a business, on a short-term and long-term basis?



DATA SOURCES

ERCOT

Real-Time Settlement Point Prices Display

*SPP values **include** the 15 Minute Online Reserve Price Adders and the Real-Time Online Reliability Deployment Price Adders.*

Operating Day: 03/03/2017 ▾

Oper Day	Interval Ending	HB_BUSAVG	HB_HOUSTON	HB_HUBAVG	HB_NORTH	HB_SOUTH	HB_WEST	LZ_AEN
03/03/2017	0015	14.64	14.64	14.64	14.64	14.64	14.64	14.64
03/03/2017	0030	14.53	14.53	14.53	14.53	14.53	14.53	14.53
03/03/2017	0045	14.49	14.49	14.49	14.49	14.49	14.49	14.49
03/03/2017	0100	14.33	14.33	14.33	14.33	14.33	14.33	14.33
03/03/2017	0115	14.11	14.11	14.11	14.11	14.11	14.11	14.11
03/03/2017	0130	14.02	14.02	14.02	14.02	14.02	14.02	14.02
03/03/2017	0145	13.96	13.96	13.96	13.96	13.96	13.96	13.96
03/03/2017	0200	14.08	14.08	14.08	14.08	14.08	14.08	14.08
03/03/2017	0215	14.00	14.00	14.00	14.00	14.00	14.00	14.00
03/03/2017	0230	13.90	13.90	13.90	13.90	13.90	13.90	13.90
03/03/2017	0245	14.13	14.13	14.13	14.13	14.13	14.13	14.13
03/03/2017	0300	14.44	14.44	14.44	14.44	14.44	14.44	14.44
03/03/2017	0315	14.35	14.35	14.35	14.35	14.35	14.35	14.35
03/03/2017	0330	14.34	14.34	14.34	14.34	14.34	14.34	14.34
03/03/2017	0345	14.07	14.07	14.07	14.07	14.07	14.07	14.07
03/03/2017	0400	14.07	14.07	14.07	14.07	14.07	14.07	14.07
03/03/2017	0415	14.63	14.63	14.63	14.63	14.63	14.63	14.63
03/03/2017	0430	14.82	14.82	14.82	14.82	14.82	14.82	14.82
03/03/2017	0445	15.01	15.01	15.01	15.01	15.01	15.01	15.01
03/03/2017	0500	15.06	15.06	15.06	15.06	15.06	15.06	15.06
03/03/2017	0515	15.51	15.51	15.51	15.51	15.51	15.51	15.51
03/03/2017	0530	15.72	15.72	15.72	15.72	15.72	15.72	15.72
03/03/2017	0545	16.16	16.16	16.16	16.16	16.16	16.16	16.16
03/03/2017	0600	16.71	16.71	16.71	16.71	16.71	16.71	16.71
03/03/2017	0615	17.73	17.73	17.73	17.73	17.73	17.73	17.73
03/03/2017	0630	18.98	18.98	18.98	18.98	18.98	18.98	18.98
03/03/2017	0645	20.20	20.20	20.20	20.20	20.20	20.20	20.20

MRCC CLI-MATE DATA

Hourly Data Between Two Dates
HOUSTON HOBBY AP (TX)
12918
Lat/Lon/Elev: 29.6381/-95.2819/44 ft

Date	Time	Temp (F)	RH (%)	Dewpt (F)	Wind Spd (mph)	Wind Direction (deg)	Peak Wind Gust(mph)	Atm Press (hPa)	Precip (in)	Wind Chill (F)	Heat Index (F)
2011-01-01	00:00	63	35	35	16	20	M	1009.03	M	NC	NC
2010-01-01	01:00	49	77	42	20	330	24	1024.27	M	NC	NC
2010-01-01	02:00	47	80	41	20	340	29	1024.94	M	40	NC

Time Range: 2010 – present
Sampling Rate: every 15 minutes / hourly
Weather Prediction: darksky (R package)

DATA PREPARATION

- Consistent date/time formats
- Derived column for z-scores
- Derive “class” column based on z-score
  - SELL: z-score greater than or equal to 1
  - HOLD: z-score between -1 and 1
  - BUY: z-score greater than or equal to -1
- Average monthly pricing
- Determine appropriate segmentation for weather
  - By month? By season?
  - By time of day, or hour?
  - Some combination?

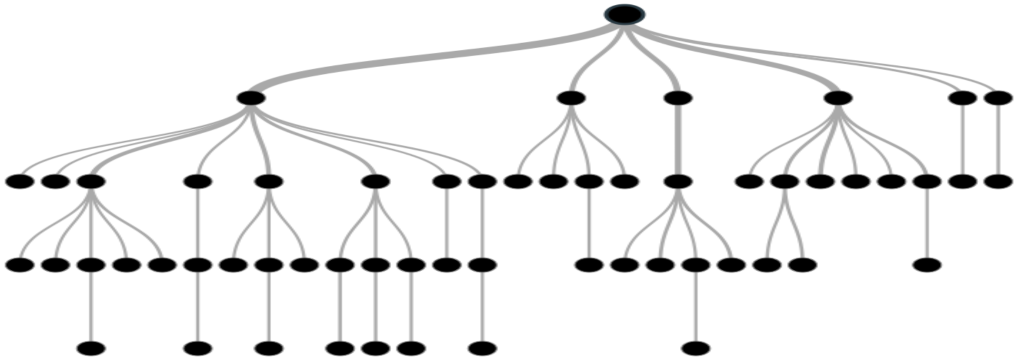
DATA ANALYSIS

Data Acquisition

- Obtain developer credentials for **DarkSky**
- Install **darksky** package in R
- Download data from **ERCOT website** (2010-2017)
- Download data from **MRCC website** (2010-2017)

ERCOT: approximately 75,000 observations per month, over a span of 77 months (~5.775MM)

MRCC: hourly observations, since 2010 (54,191)



Algorithms to Test

- Both numeric and categorical features
- Focus on usability for end user, as well as confidence for ability to predict
- Algorithms Tested
  - Decision Tree (6 and 16 features)
  - Linear Regression for price prediction (unsuccessful)
  - Three-pronged ("BUY", "SELL", "STAY"; tested alternatives)

RATIONALE FOR FEATURE SELECTION

Focus on usability and confidence to predict

- PCA was performed to see which variables would impact model to the greatest extent
  - Day, Month, Time
  - Forecasted temperature; historic average temperature
  - Dewpoint, precipitation, Atmospheric Pressure, Relative Humidity
  - Wind direction, wind chill, wind speed
- Three variables could potentially serve as good predictors; but are not capable of being accurately forecasted
  - Relative humidity
  - Precipitation and dewpoint

SELECTED VARIABLES
Month
Day
Time
Forecasted Temperature
Atmospheric Pressure
Difference of forecasted temperature from historic average temperature

DATA ANALYSIS

Trained on historical data (2010 – 2016)

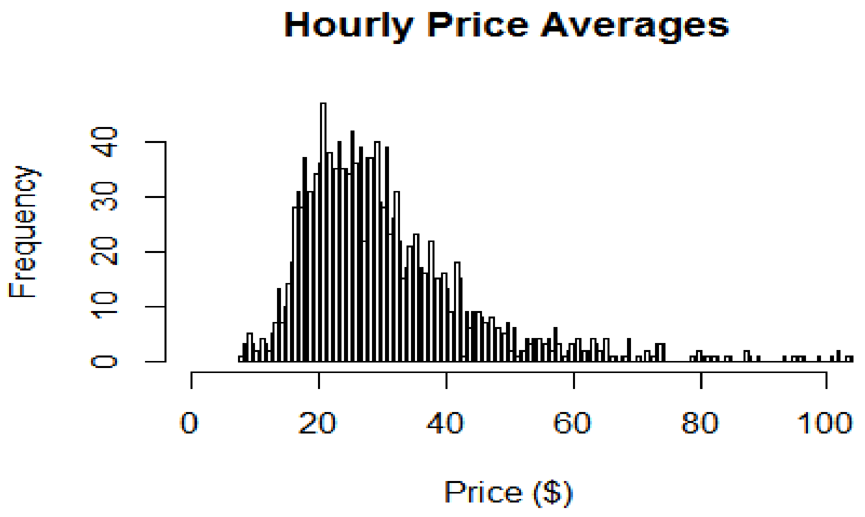
- Test predictions from January 2017 onward

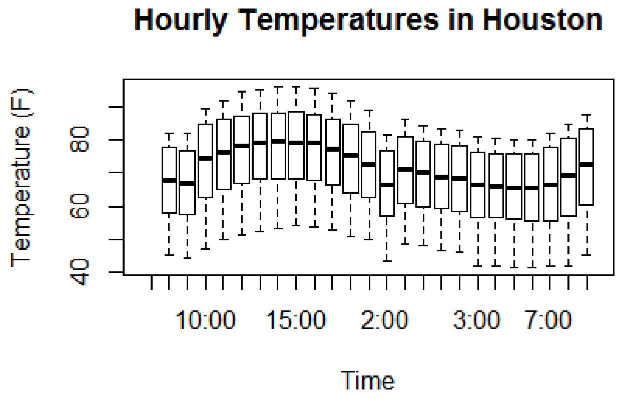
ERCOT: approximately 75,000 observations per month, over a span of 77 months (~5.775MM)

MRCC: hourly observations, since 2010 (54,191)

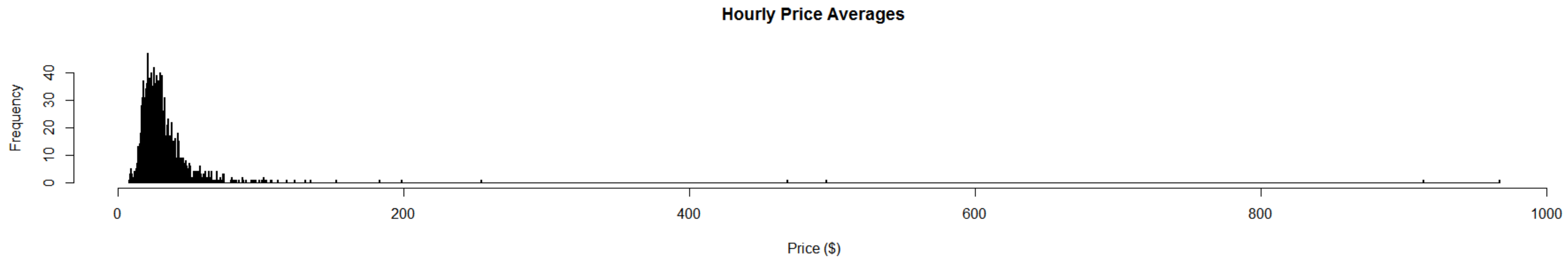
R Packages used:

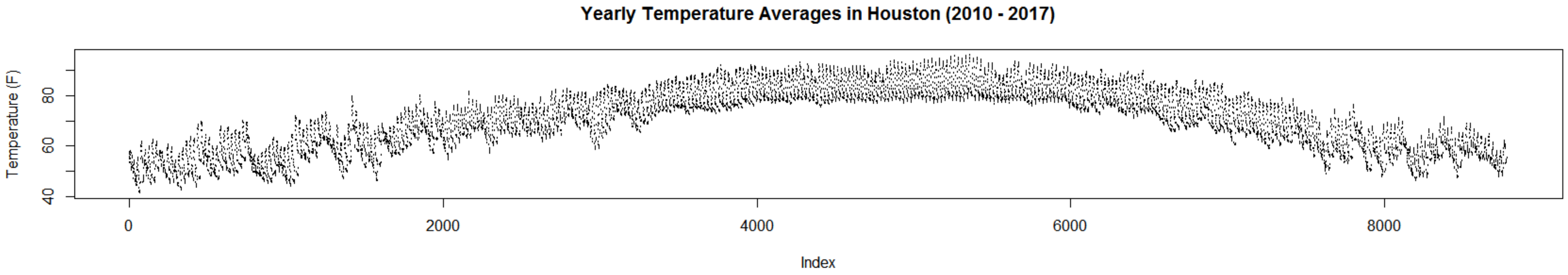
- **lubridate**: to make manipulating dates / times simpler
- **darksky**: temperature forecasting
- **dplyr**: data manipulation
- **ggplot2**: visualization
- **C50**: decision tree algorithm
- **gmodels**: visualizing decision tree results with CrossTable()



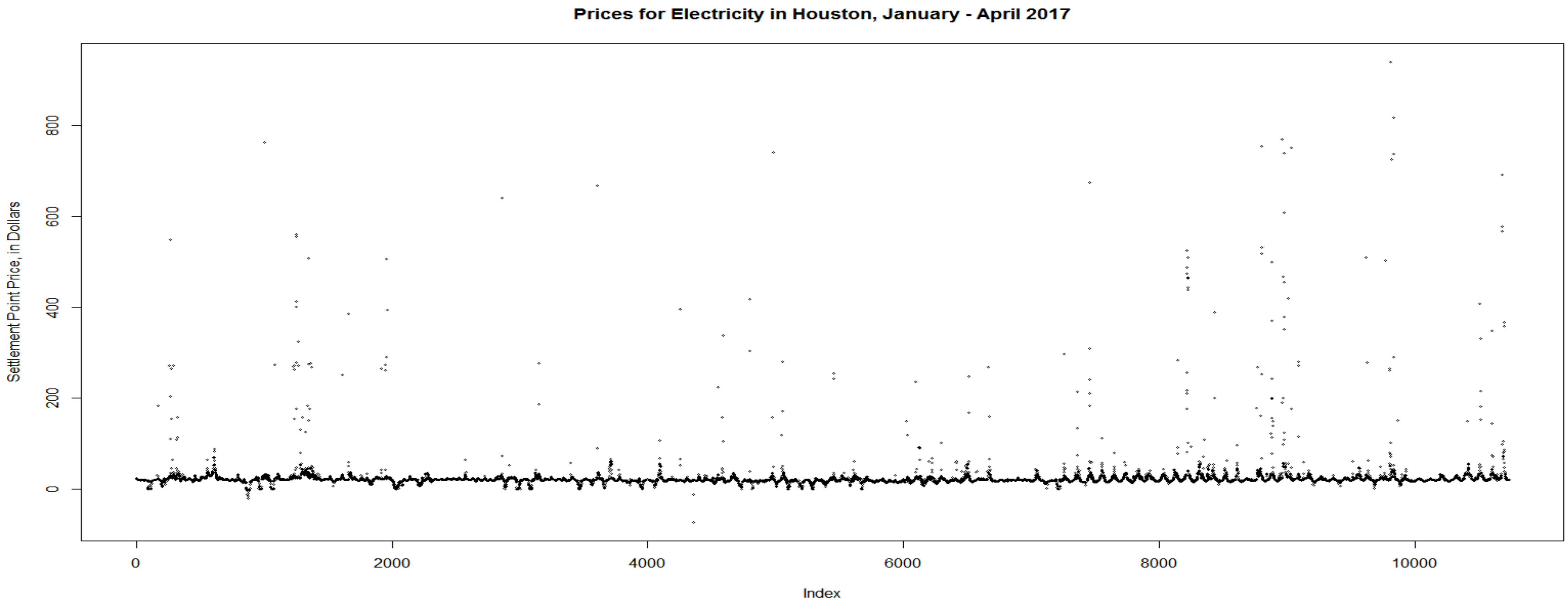


REMINDER: PRICE DISTRIBUTION IS EXTREMELY RIGHT-SKEWED!





ELECTRICITY PRICES IN HOUSTON (JANUARY 2017 – APRIL 2017)

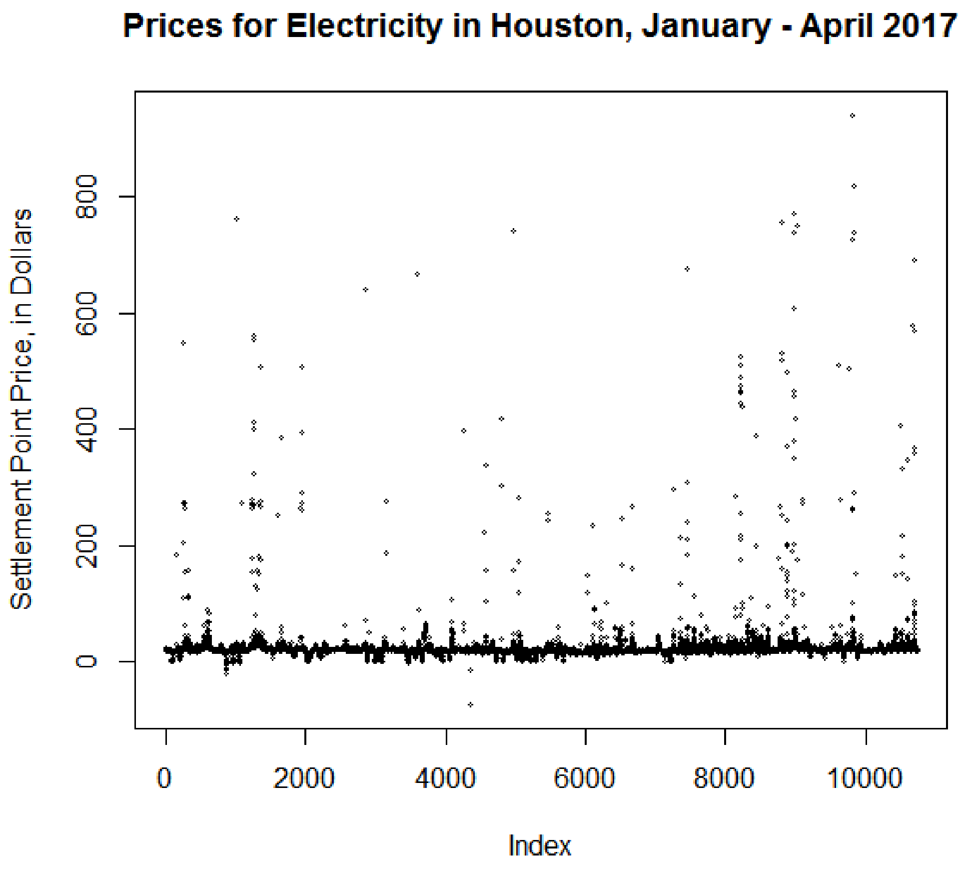


M1: ±0.5 STANDARD DEVIATIONS

```
Evaluation on training data (52774 cases):

Trial       Decision Tree
-----     ----------------
   Size        Errors

   0    1610 13472(25.5%)
   1     640 15688(29.7%)
   2     600 16824(31.9%)
   3     517 18222(34.5%)
   4     325 19730(37.4%)
   5     492 18261(34.6%)
   6     243 20999(39.8%)
   7     637 17201(32.6%)
   8     685 16383(31.0%)
   9     706 15180(28.8%)
boost        10979(20.8%)   <<
```

```
   (a)   (b)    <-classified as
  ----  ----
 24555  4329    (a): class BUY
  6650 17240    (b): class SELL


 Attribute usage:

 100.00% Temp..F.
 100.00% Atm.Press..hPa.
  99.63% Day
  99.51% mean_temps
  99.37% Month
  98.57% Time


Time: 2.7 secs
```

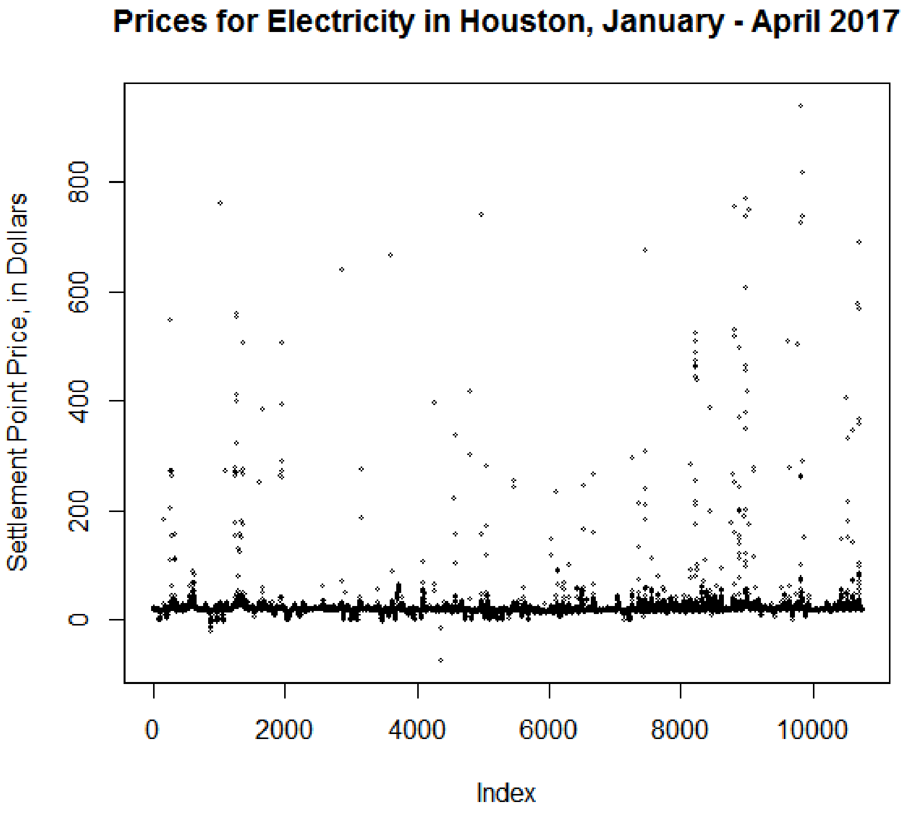
```
Call:
C5.0.default(x = final_prod_2[1:6], y = final_prod_2$key, trials = 10)

Classification Tree
Number of samples: 52774
Number of predictors: 6

Number of boosting iterations: 10
Average tree size: 645.5
```



M2: THREE-PRONGED APPROACH FOR ±1 STANDARD DEVIATION

```
    Size      Errors

    0      693 12274(23.3%)
    1      258 14021(26.6%)
    2      497 14715(27.9%)
    3      521 15479(29.3%)
    4      571 16546(31.4%)
    5      579 16097(30.5%)
    6      609 17034(32.3%)
    7      513 14483(27.4%)
    8      465 13716(26.0%)
    9      386 13542(25.7%)
 boost        11514(21.8%)   <<


    (a)   (b)   (c)    <-classified as
   ----  ----  ----
   1523  5171    29    (a): class BUY
    224 37556   350    (b): class HOLD
     28  5712  2181    (c): class SELL
```

```
   (a)   (b)   (c)    <-classified as
  ----  ----  ----
  1523  5171    29    (a): class BUY
   224 37556   350    (b): class HOLD
    28  5712  2181    (c): class SELL


Attribute usage:

100.00% Day
100.00% mean_temps
100.00% Temp..F.
100.00% Atm.Press..hPa.
 99.91% Month
 92.15% Time
```



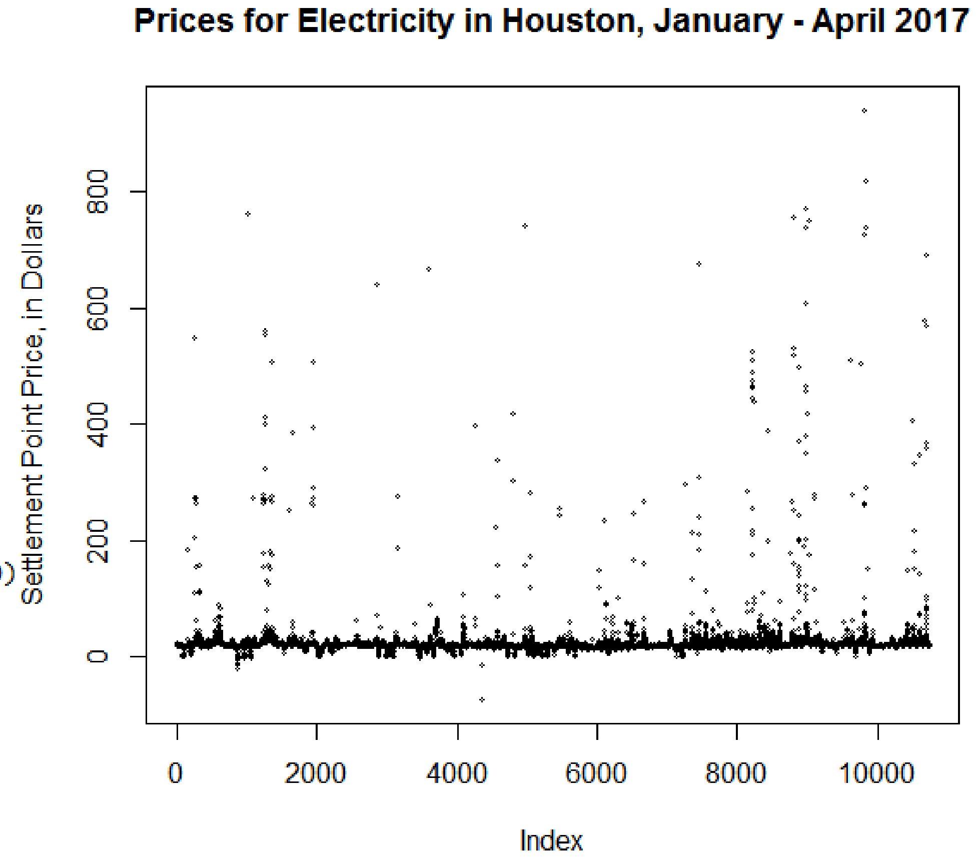

```
Call:
C5.0.default(x = final_prod_2[1:6], y = as.factor(final_product$key2), trials = 10)

Classification Tree
Number of samples: 52774
Number of predictors: 6

Number of boosting iterations: 10
Average tree size: 509.2
```

M3: ±2 STANDARD DEVIATIONS

```
Evaluation on training data (52774 cases):

Trial       Decision Tree
-----     -----------------
   Size        Errors

   0      99 1641( 3.1%)
   1      56 4628( 8.8%)
   2      76 4240( 8.0%)
   3     122 5120( 9.7%)
   4     102 5963(11.3%)
   5     136 5541(10.5%)
   6     108 3771( 7.1%)
   7      75 2870( 5.4%)
   8      69 2566( 4.9%)
   9      42 2084( 3.9%)
boost          1671( 3.2%)   <<
```

```
   (a)   (b)    <-classified as
  ----  ----
 50583    23    (a): class BUY
  1648   520    (b): class SELL


Attribute usage:

100.00% Month
100.00% mean_temps
100.00% Temp..F.
100.00% Atm.Press..hPa.
 99.96% Time
 99.45% Day


Time: 2.5 secs
```

```
Call:
C5.0.default(x = final_prod_2[1:6], y = as.factor(final_product$key3), trials = 10)

Classification Tree
Number of samples: 52774
Number of predictors: 6

Number of boosting iterations: 10
Average tree size: 88.5

Non-standard options: attempt to group attributes
```

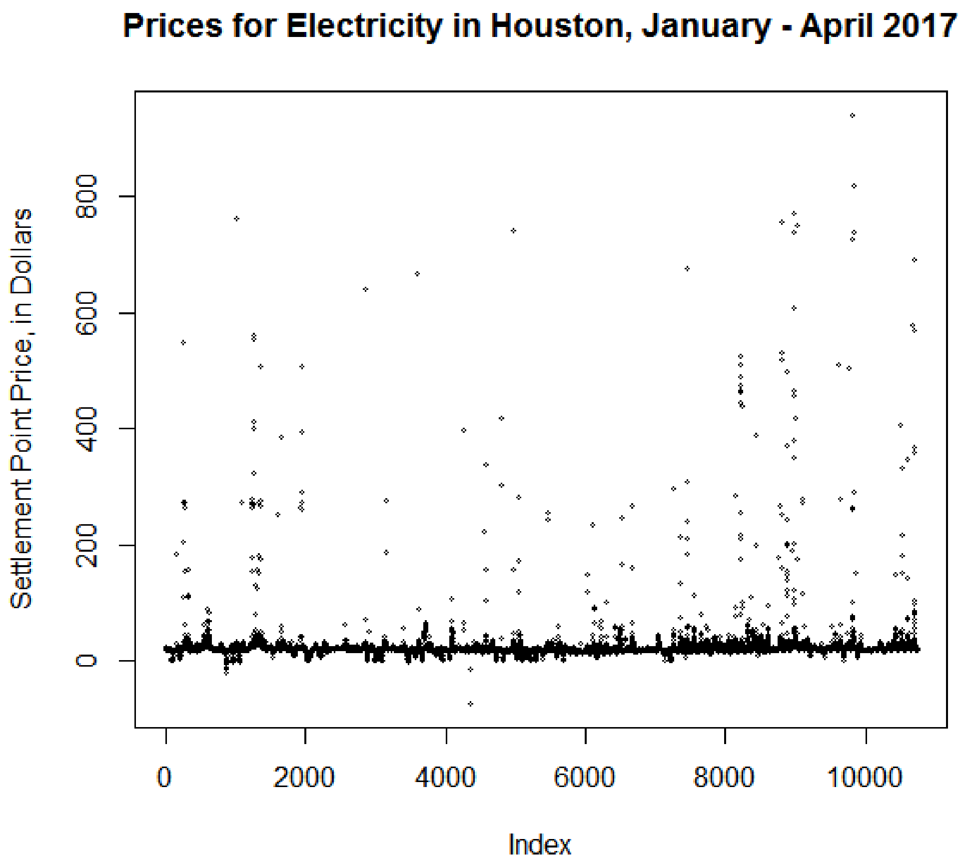


M4: ±1.5 STANDARD DEVIATIONS

```
Trial       Decision Tree
-----    ----------------
   Size        Errors

   0     119 3435( 6.5%)
   1      60 6598(12.5%)
   2     101 7280(13.8%)
   3      97 6404(12.1%)
   4     169 8779(16.6%)
   5     144 8161(15.5%)
   6     108 5490(10.4%)
   7      92 5068( 9.6%)
   8      71 4214( 8.0%)
   9      59 4081( 7.7%)
boost         3392( 6.4%)   <<
```

```
   (a)   (b)    <-classified as
  ----  ----
 48588   104    (a): class BUY
  3288   794    (b): class SELL


Attribute usage:

100.00% Month
100.00% mean_temps
100.00% Temp..F.
100.00% Atm.Press..hPa.
 99.65% Time
 95.95% Day
```

```
Call:
C5.0.default(x = final_prod_2[1:6], y = as.factor(final_product$key4), trials = 10)

Classification Tree
Number of samples: 52774
Number of predictors: 6

Number of boosting iterations: 10
Average tree size: 102
```

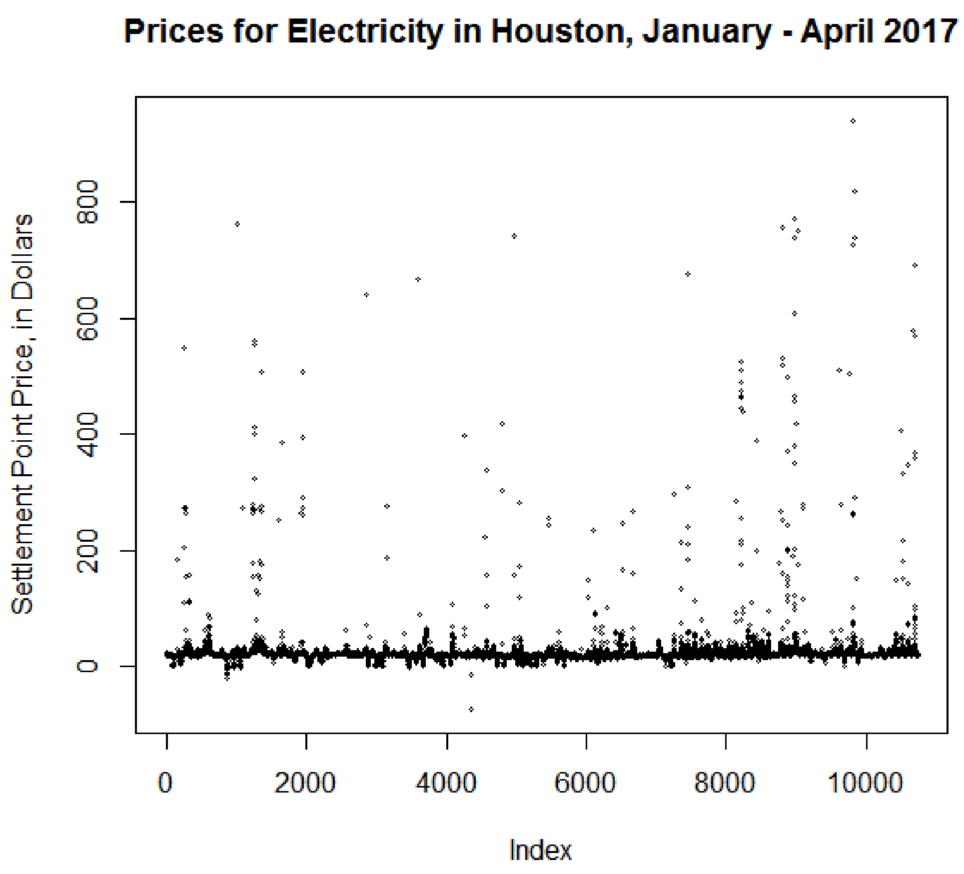


M5: ±1.25 STANDARD DEVIATIONS

```
Evaluation on training data (52774 cases):

Trial       Decision Tree
-----     ----------------
   Size      Errors

   0     158 4956( 9.4%)
   1      78 7027(13.3%)
   2     115 8151(15.4%)
   3     143 10746(20.4%)
   4     163 10541(20.0%)
   5     176 10538(20.0%)
   6     154 7414(14.0%)
   7     119 6891(13.1%)
   8     116 5706(10.8%)
   9      81 5687(10.8%)
boost          4711( 8.9%)   <<
```

```
   (a)    (b)    <-classified as
  ----   ----
 46827    186    (a): class BUY
  4525   1236    (b): class SELL


Attribute usage:

100.00% Month
100.00% mean_temps
100.00% Temp..F.
100.00% Atm.Press..hPa.
 99.67% Day
 93.31% Time


Time: 2.5 secs
```

```
Call:
C5.0.default(x = final_prod_2[1:6], y = as.factor(final_product$key5), trials = 10)

Classification Tree
Number of samples: 52774
Number of predictors: 6

Number of boosting iterations: 10
Average tree size: 130.3
```

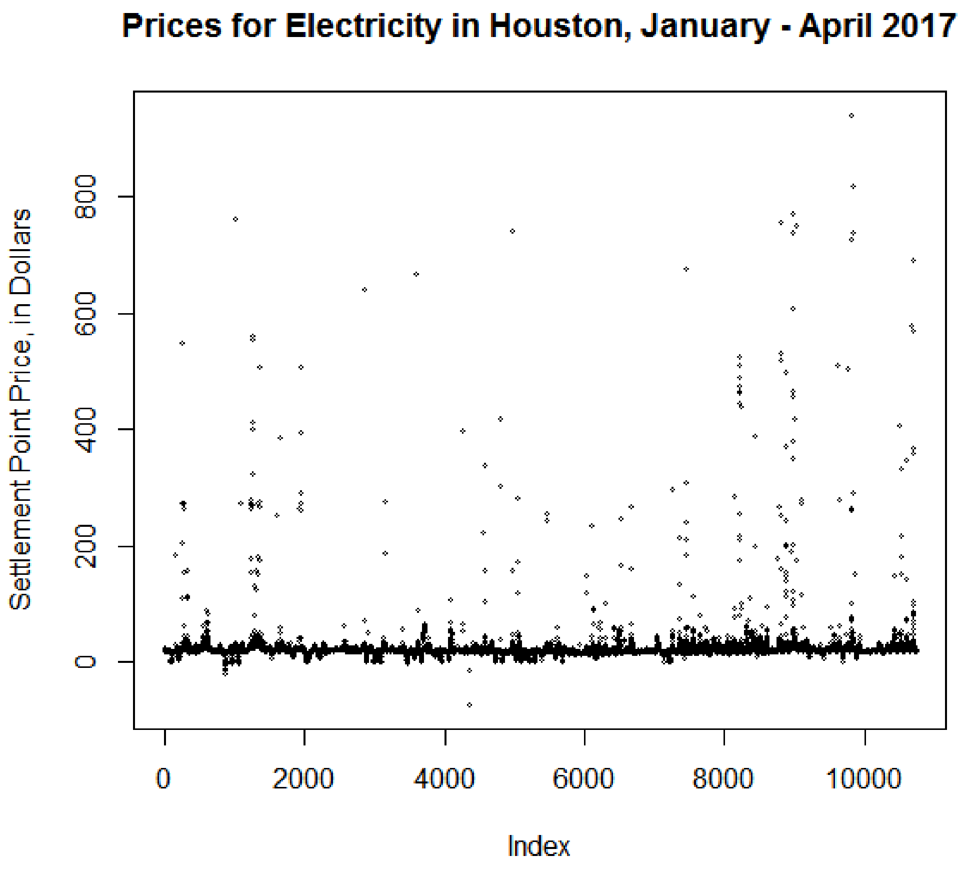


M6: ±1 STANDARD DEVIATIONS

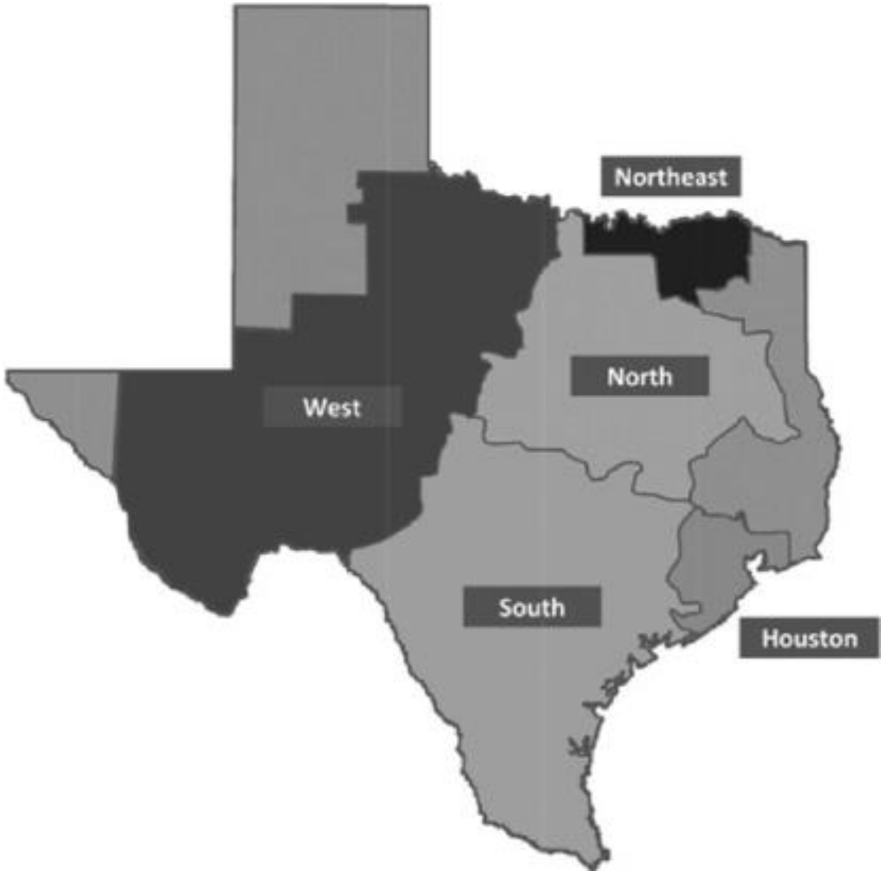
```
Evaluation on training data (52774 cases):

Trial       Decision Tree
-----   ----------------
  Size       Errors

  0    1612 13100(24.8%)
  1     592 16611(31.5%)
  2     465 17848(33.8%)
  3     438 19085(36.2%)
  4     537 17878(33.9%)
  5     263 21121(40.0%)
  6     303 20198(38.3%)
  7     294 19296(36.6%)
  8     687 15773(29.9%)
  9     785 14729(27.9%)
boost        11043(20.9%)   <<
```

```
   (a)   (b)    <-classified as
  ----  ----
 24319  4289    (a): class BUY
  6754 17412    (b): class SELL


Attribute usage:

100.00% Temp..F.
100.00% Atm.Press..hPa.
 99.80% Month
 99.55% mean_temps
 99.46% Day
 96.53% Time


Time: 2.7 secs
```

```
Call:
C5.0.default(x = final_prod_2[1:6], y = as.factor(final_product$key6), trials = 10)

Classification Tree
Number of samples: 52774
Number of predictors: 6

Number of boosting iterations: 10
Average tree size: 597.6
```



- How do weather patterns (humidity, precipitation, temperature) impact price of electricity for each of the hubs?
- What other factors or combination of factors impact electricity prices in other hubs? Are they similar, or are there regional variations?
  - Example: wind speed in West Texas
- What are the optimal times to purchase electricity at each of the hubs? To sell?
- How do other factors – for example, capacity – impact supply of electricity in Houston? How would that change, if new infrastructure is built to support the capacity?



FUTURE WORK / NEXT STEPS

QUESTIONS / SUGGESTIONS?