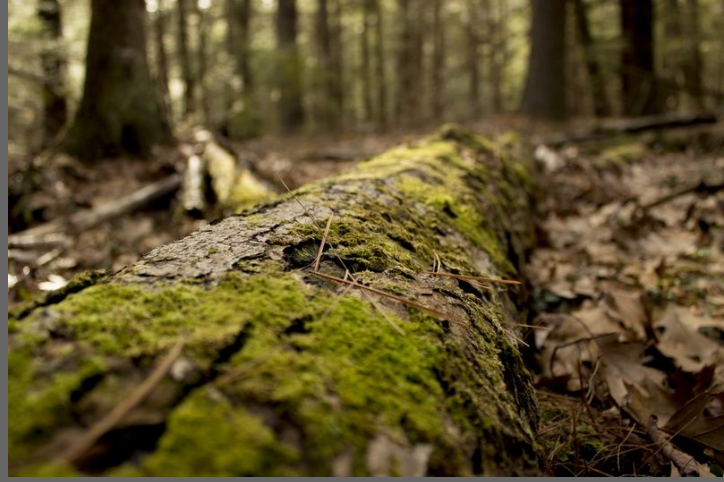


Final Capstone - Harvard and Duke Forests environmental growing conditions

Project goals:

1. Combine 3 datasets into 1
2. Predict difference between Duke and Harvard forests
3. Model the relationship between air temperature and PAR
4. Use classification models to predict treatment changes
5. Use Random Forest to predict Tree Species with treatments ('Only for Harvard forest')
6. PCA model of the dataset - how is it related or different
7. Time series modeling of Air temperature and photosynthetically active radiation

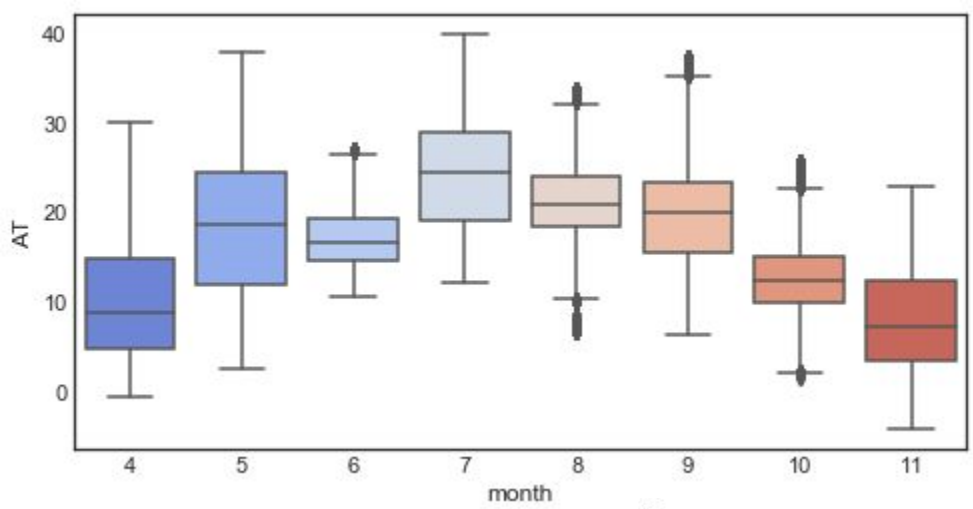
Features Used

year: year
month: month
day: day of month
time: hour of day
chamber: chamber number (1-12)
treatment: light treatment
G: chamber in open gap
S: chamber under closed canopy

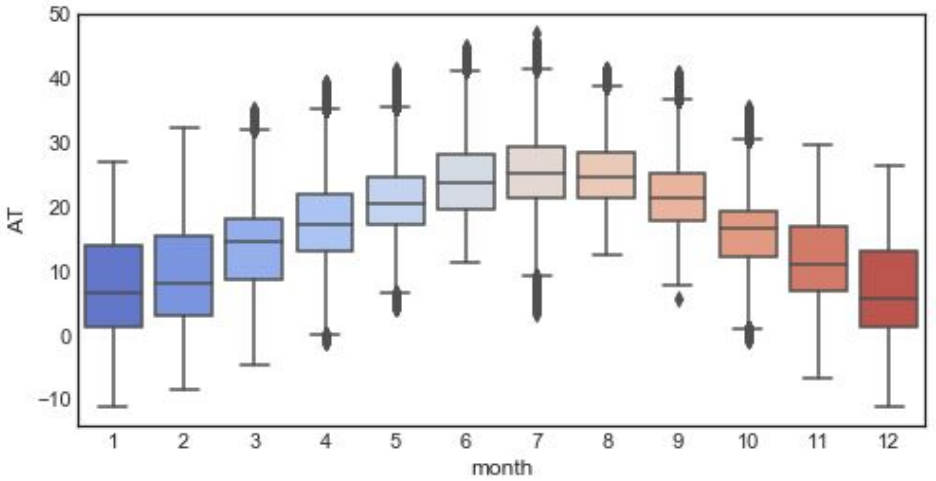
warming: warming treatment
3: 3 degrees C
5: 5 degrees C
A: ambient
C: control

AT: air temperature (unit: celsius / missing value: NA)
Q: photosynthetically active radiation (unit: micromolePerMeterSquaredPerSecond / missing value: NA)
Rh: relative humidity (%) (unit: dimensionless / missing value: NA)
SM: volumetric water content (fractional) (unit: dimensionless / missing value: NA)
ST: soil temperature at 5cm depth (unit: celsius / missing value: NA)

Exploratory data analysis

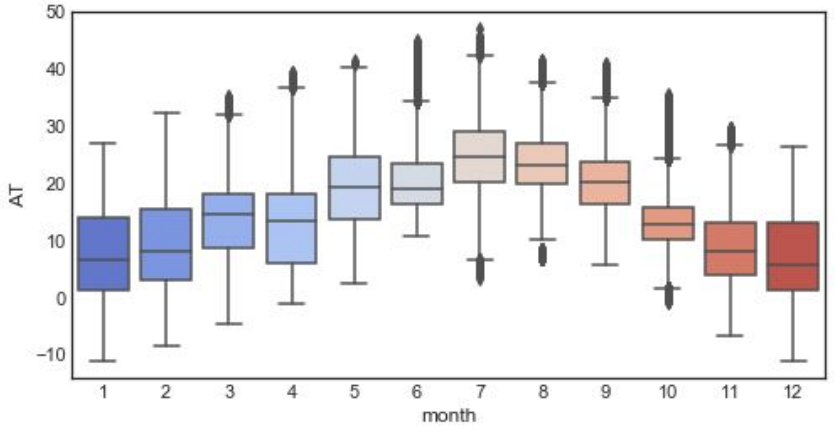


HF

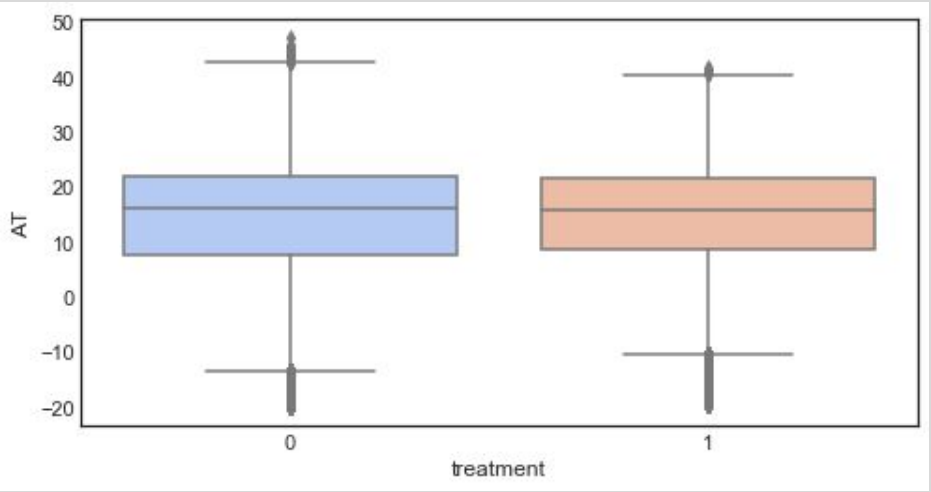


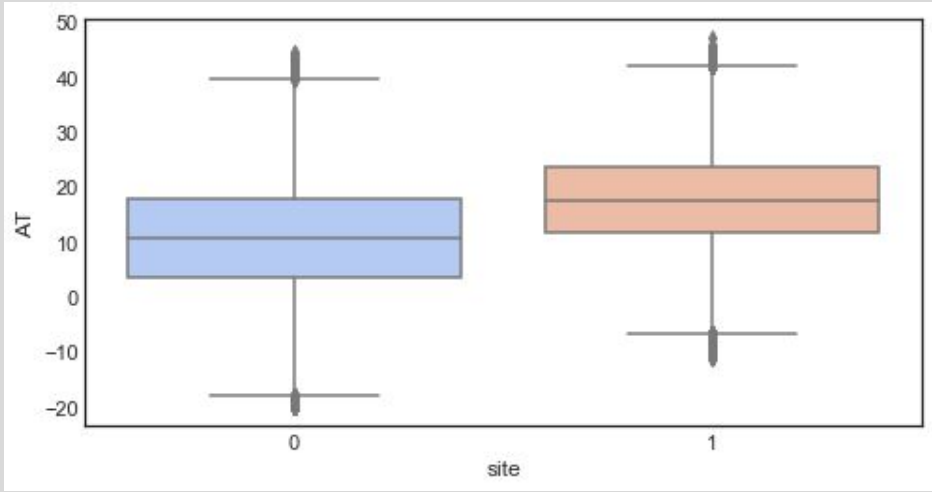
DF

HFDF

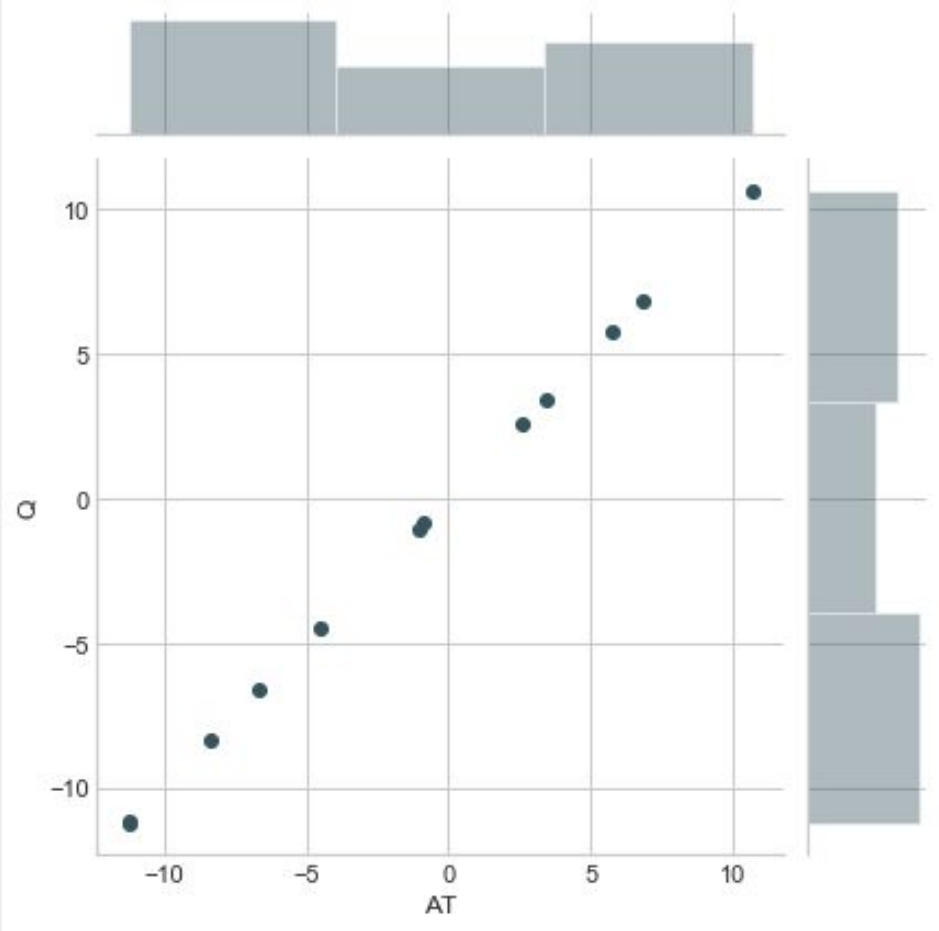


Exploratory data analysis

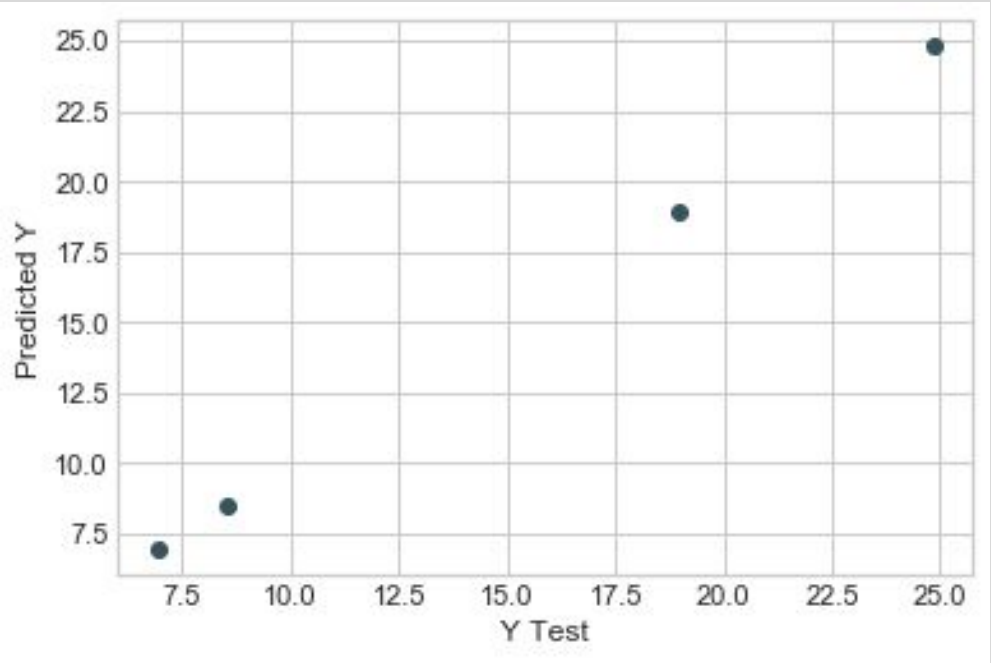




Relationship of Q (PAR) and AT (air temperature) use mean



Linear regression result for predicting Q (PAR)



MAE: 9.414691248821327e-14
MSE: 1.059597967892863e-26
RMSE: 1.0293677515314257e-13

	Coeffecient
AT	-0.210
Rh	1.210

Does the light treatment have an effect?

```
X = HfDf.ix[:, 'chamber':'ST'].drop(columns=['treatment'])
y = HfDf['treatment']
```

Logistic regression

	precision	recall	f1-score	support
0	0.53	0.84	0.65	370200
1	0.51	0.18	0.27	341922
micro avg	0.52	0.52	0.52	712122
macro avg	0.52	0.51	0.46	712122
weighted avg	0.52	0.52	0.47	712122

Random forest

Training set score: 0.9754998326596813

Test set score: 0.838022196992516

	precision	recall	f1-score	support
0	0.83	0.87	0.85	112174
1	0.85	0.81	0.83	103621
micro avg	0.84	0.84	0.84	215795
macro avg	0.84	0.84	0.84	215795
weighted avg	0.84	0.84	0.84	215795

Training set score: 0.44117259224341254

Test set score: 0.42928410011512386

	precision	recall	f1-score	support
acba	0.00	0.00	0.00	680
acru	0.53	0.86	0.66	56089
acsa	0.30	0.01	0.01	1438
acun	0.00	0.00	0.00	154
beal	0.20	0.20	0.20	3090
bele	0.17	0.05	0.08	6651
bepa	0.18	0.20	0.19	7270
beun	0.29	0.61	0.39	10244
cagl	0.00	0.00	0.00	6
fagr	0.00	0.00	0.00	633
fram	0.20	0.02	0.04	1171
ilvo	0.00	0.00	0.00	32
list	0.27	0.02	0.04	3102
litu	0.00	0.00	0.00	4840
magr	0.00	0.00	0.00	1218
maun	0.00	0.00	0.00	64
mavi	0.00	0.00	0.00	385
nysy	0.00	0.00	0.00	1587
pipa	0.00	0.00	0.00	214
pire	0.34	0.01	0.03	4206
pist	0.40	0.11	0.17	13903
pita	0.00	0.00	0.00	1051
piun	0.31	0.15	0.20	3816
prpe	0.00	0.00	0.00	296
prse	0.08	0.01	0.01	268
qual	0.15	0.14	0.15	8629
qufa	0.00	0.00	0.00	215
quni	0.00	0.00	0.00	544
quph	0.00	0.00	0.00	118
quru	0.31	0.05	0.08	7326
quun	0.00	0.00	0.00	775
quve	0.00	0.00	0.00	338
ulam	0.00	0.00	0.00	6
unkn	0.00	0.00	0.00	359
micro avg	0.43	0.43	0.43	140718
macro avg	0.11	0.07	0.07	140718
weighted avg	0.35	0.43	0.35	140718

Predictive species at Harvard forest

```
X = df.ix[:, 'year':'ST'].drop(columns=['treatment', 'warming'])
Y = df['Species']
```

acru: Acer rubrum, red maple - 0.66

beal: Betula alleghaniensis, yellow birch - 0.20

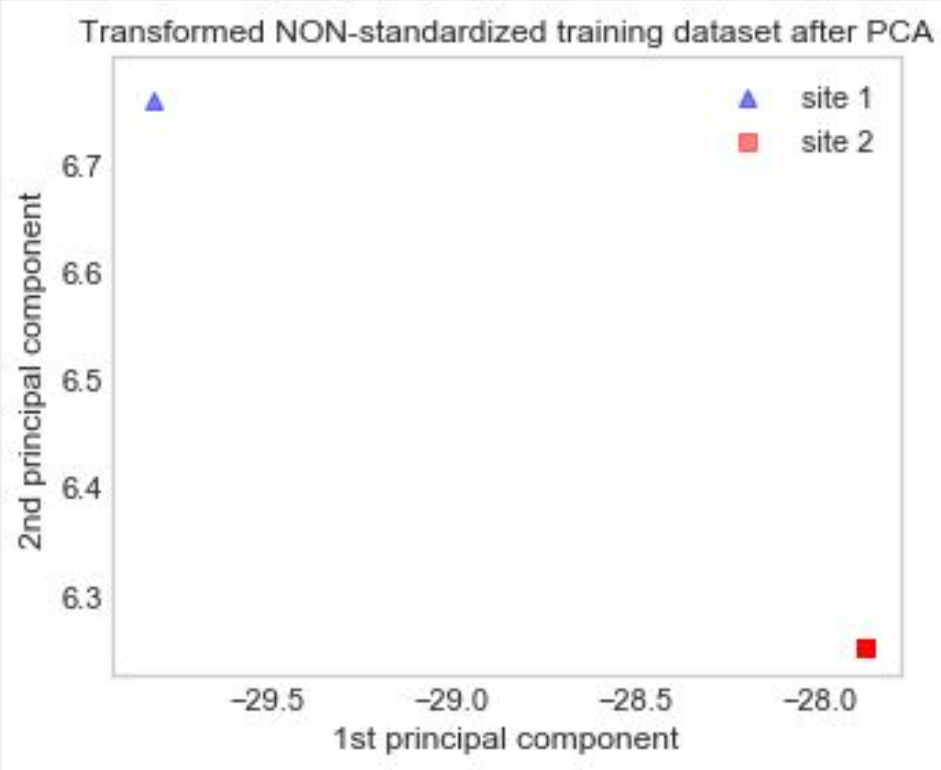
bepa: Betula papyrifera, paper birch - 0.19

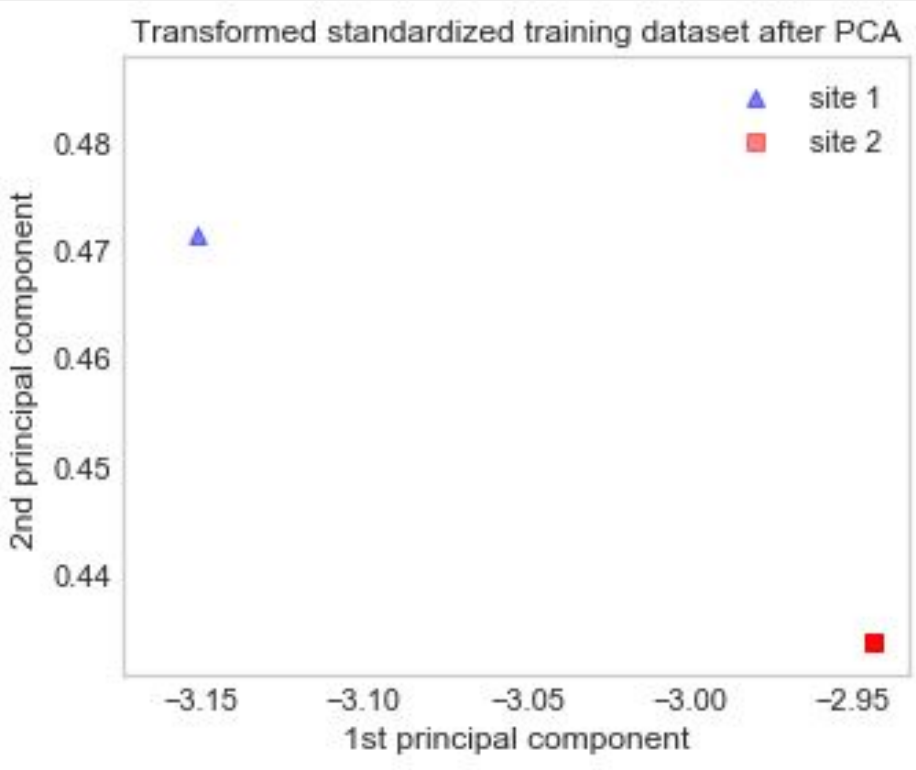
pist: Pinus strobus, white pine - 0.17

piun: Pinus spp, ambiguous pine - 0.20

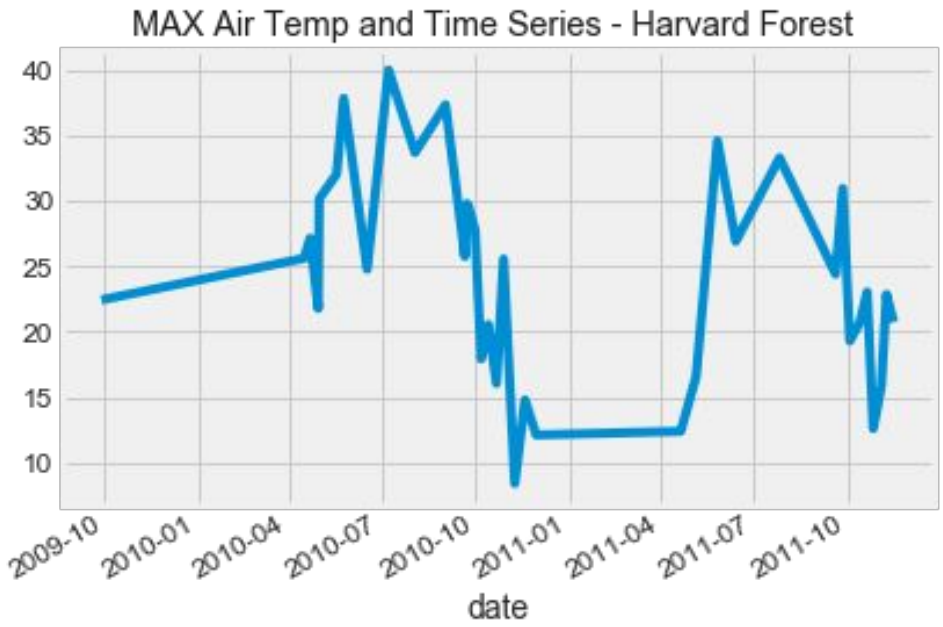
qual: Quercus alba, white oak - 0.15

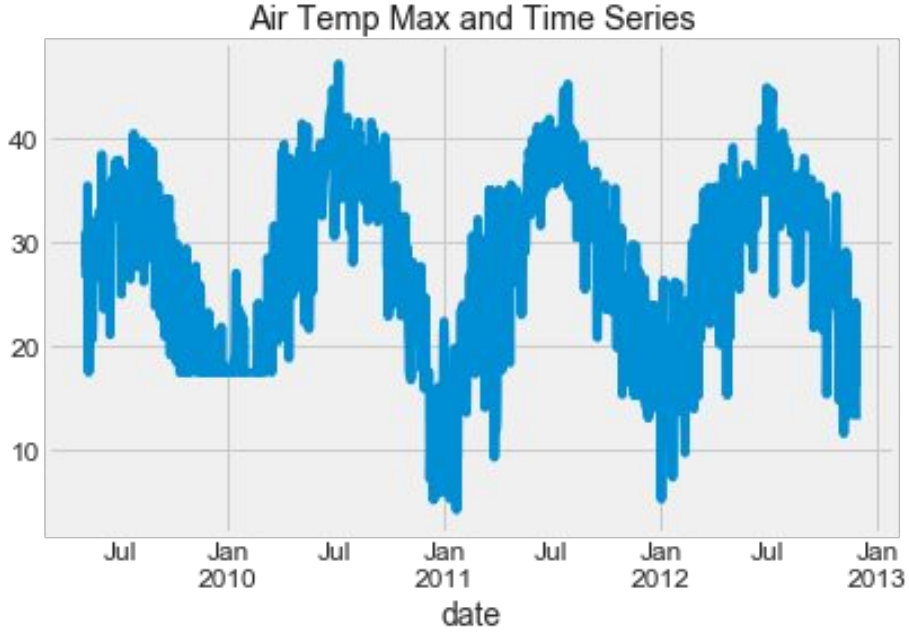
Unsupervised - PCA





Time series modeling





ARIMA Model Results

Dep. Variable:	D.AT	No. Observations:	1407176
Model:	ARIMA(0, 1, 1)	Log Likelihood	-2818612.143
Method:	css-mle	S.D. of innovations	1.793
Date:	Sat, 22 Jun 2019	AIC	5637230.285
Time:	18:53:15	BIC	5637266.757
Sample:	1	HQIC	5637240.187

	coef	std err	z	P>|z|	[0.025	0.975]
const	-8.499e-06	0.002	-0.005	0.996	-0.003	0.003
ma.L1.D.AT	0.1052	0.001	139.441	0.000	0.104	0.107

Roots

	Real	Imaginary	Modulus	Frequency
MA.1	-9.5067	+0.0000j	9.5067	0.5000

Residuals Description

count	1407176.000
mean	-0.000
std	1.793
min	-26.401
25%	-0.613
50%	-0.043
75%	0.502
max	30.875

dtype: float64

ARMA Model Results

Dep. Variable:	AT	No. Observations:	750768
Model:	ARMA(0, 1)	Log Likelihood	-2375560.272
Method:	css-mle	S.D. of innovations	5.727
Date:	Sat, 22 Jun 2019	AIC	4751126.544
Time:	20:20:23	BIC	4751161.130
Sample:	0	HQIC	4751136.173

	coef	std err	z	P>|z|	[0.025	0.975]
const	17.3824	0.012	1495.864	0.000	17.360	17.405
ma.L1.AT	0.7580	0.000	1550.674	0.000	0.757	0.759

Roots

	Real	Imaginary	Modulus	Frequency
MA.1	-1.3192	+0.0000j	1.3192	0.5000

Residuals Description

count	750768.000
mean	-0.000
std	5.727
min	-33.442
25%	-3.506
50%	0.393
75%	3.774
max	31.440

dtype: float64

Conclusion

- There is not enough difference between closed and open canopy
- Linear relationship between photosynthesis and air temperature
  - Insights into timing of blooming
  - Phenology
  - How plants may respond to warming conditions in the future

- Random forest - Harvard forest tree species
  - Give insights into how these species may respond to increased warming
  - Changes in forest resstructing
  - Ranges in ecological ranges of tree species

- Not enough time to make in depth prediction from time series
  - Max temp is much higher in DF than HF
  - Sample size
  - Future work (will continue studies)