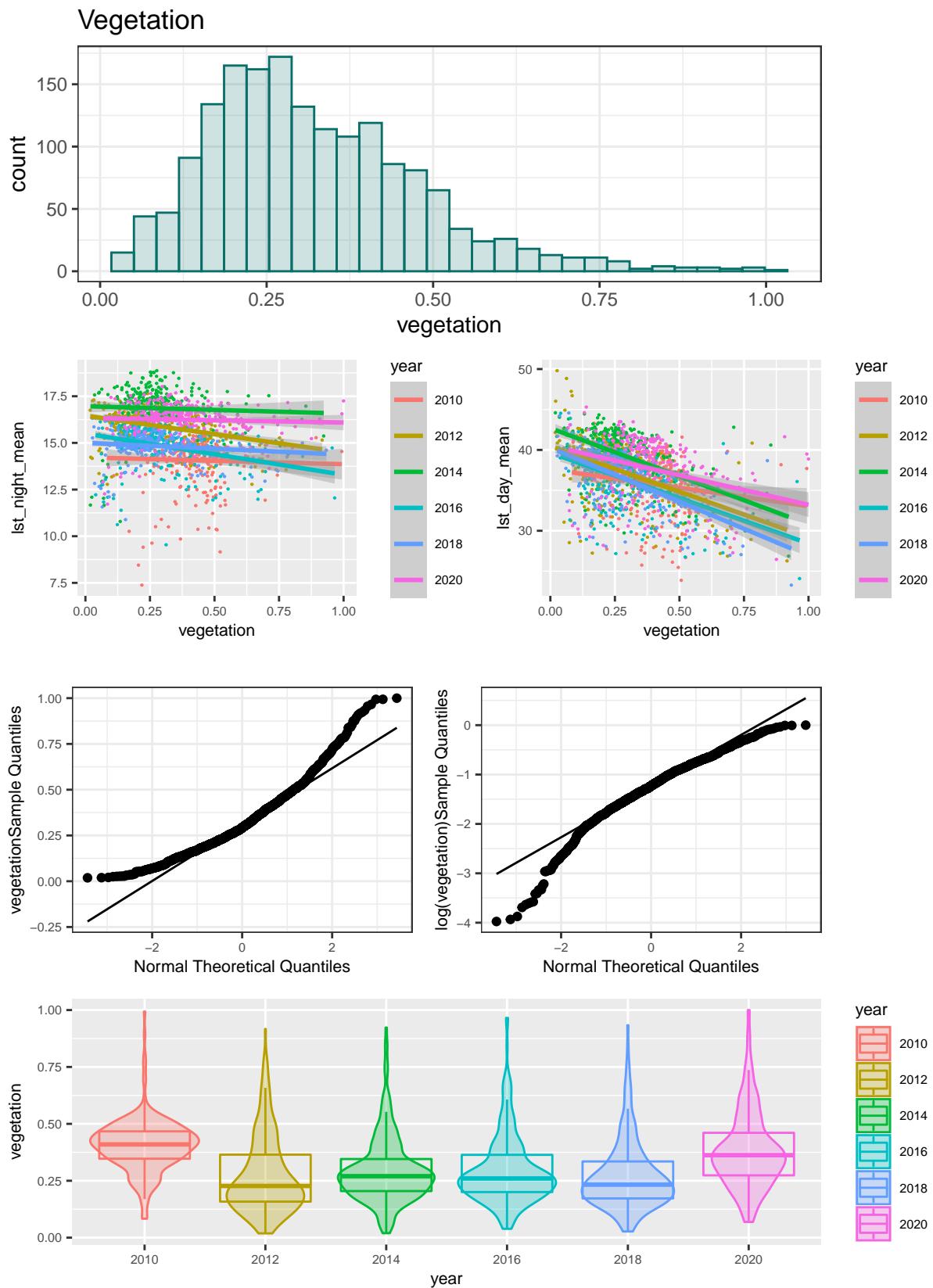


# LawnProject Panel Data Analysis / Random Forest Data

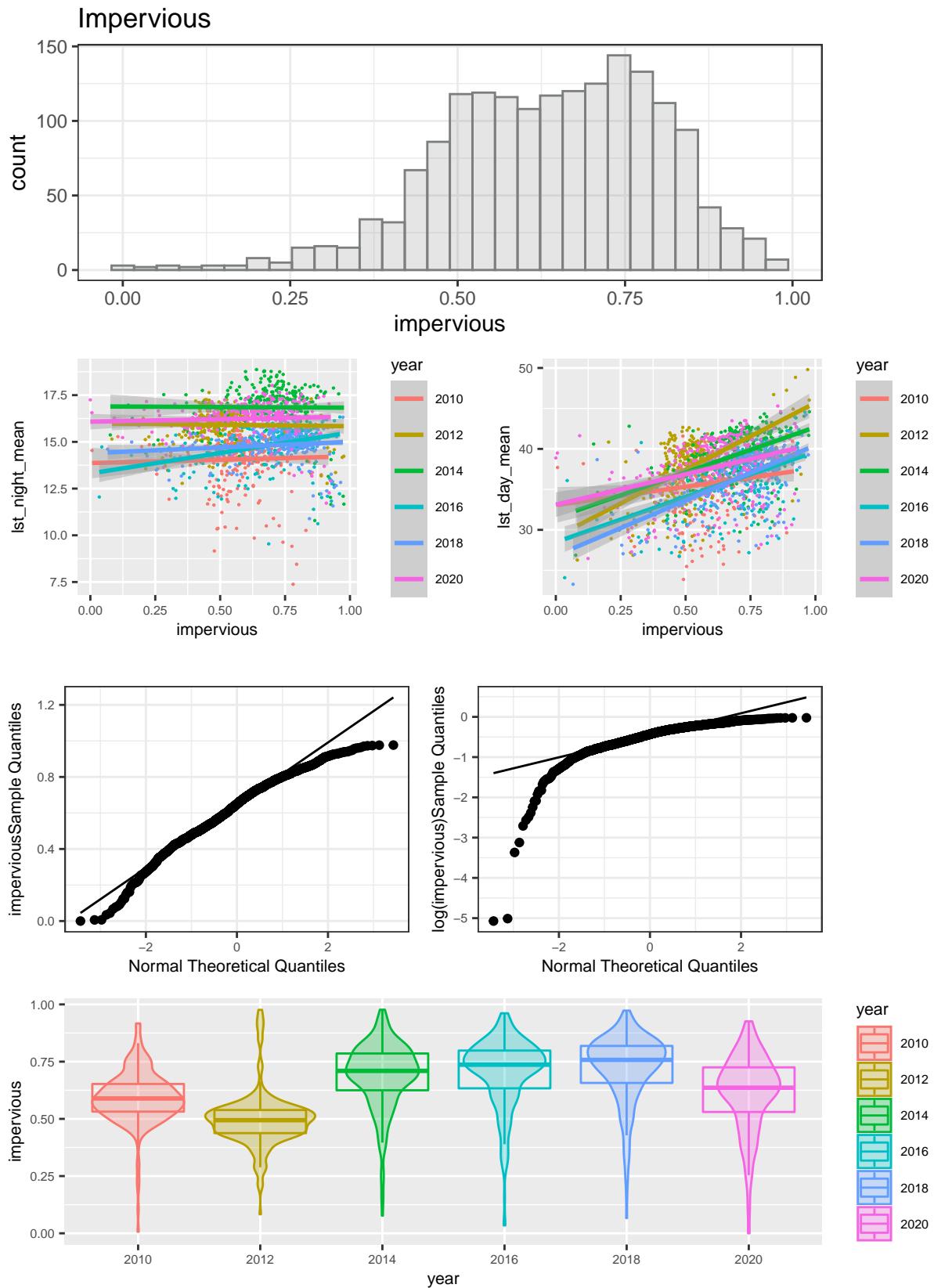
LawnProject Team

Intro

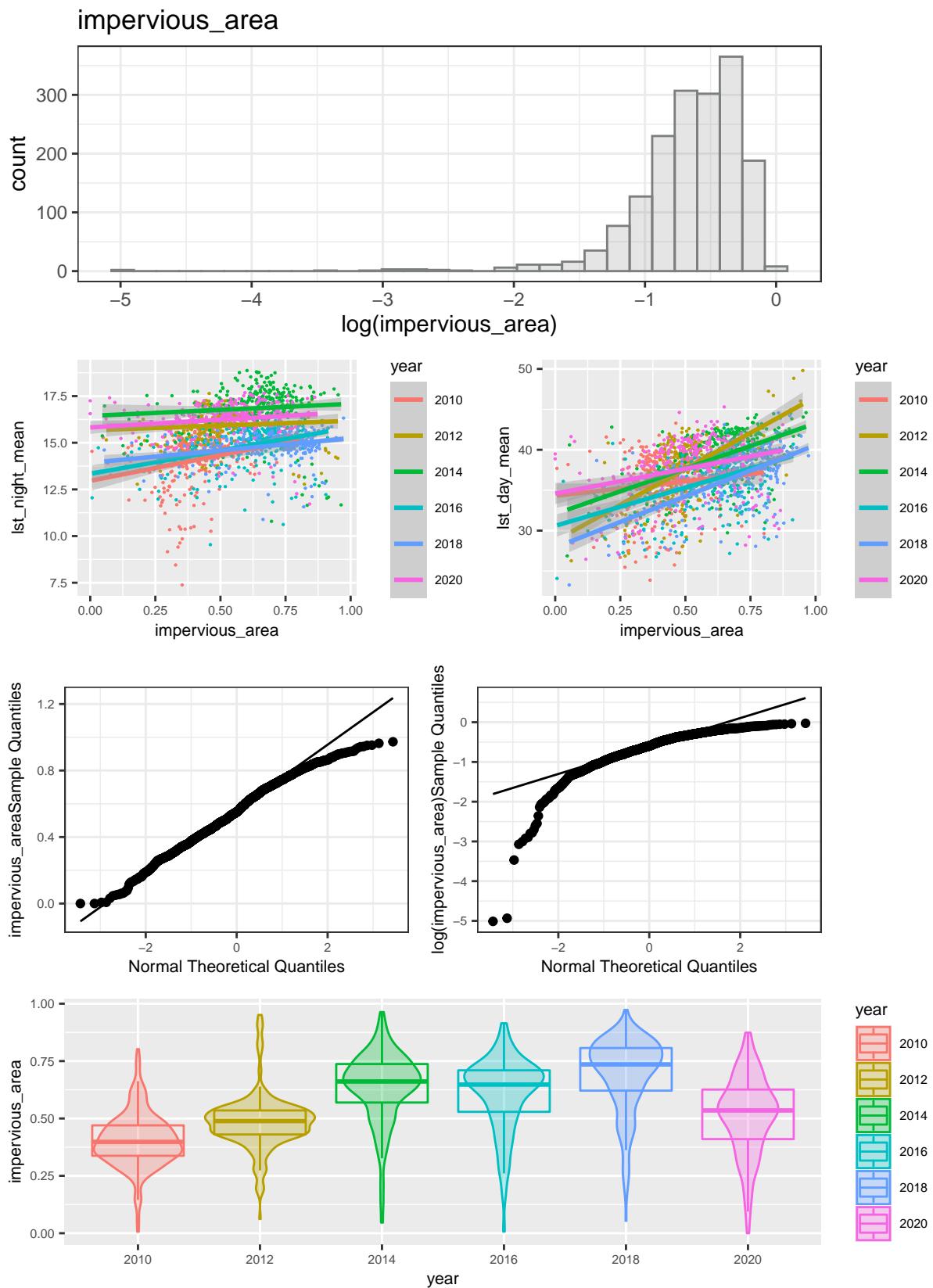
## Aggregate Vegetation Area (Trees + Lawn)



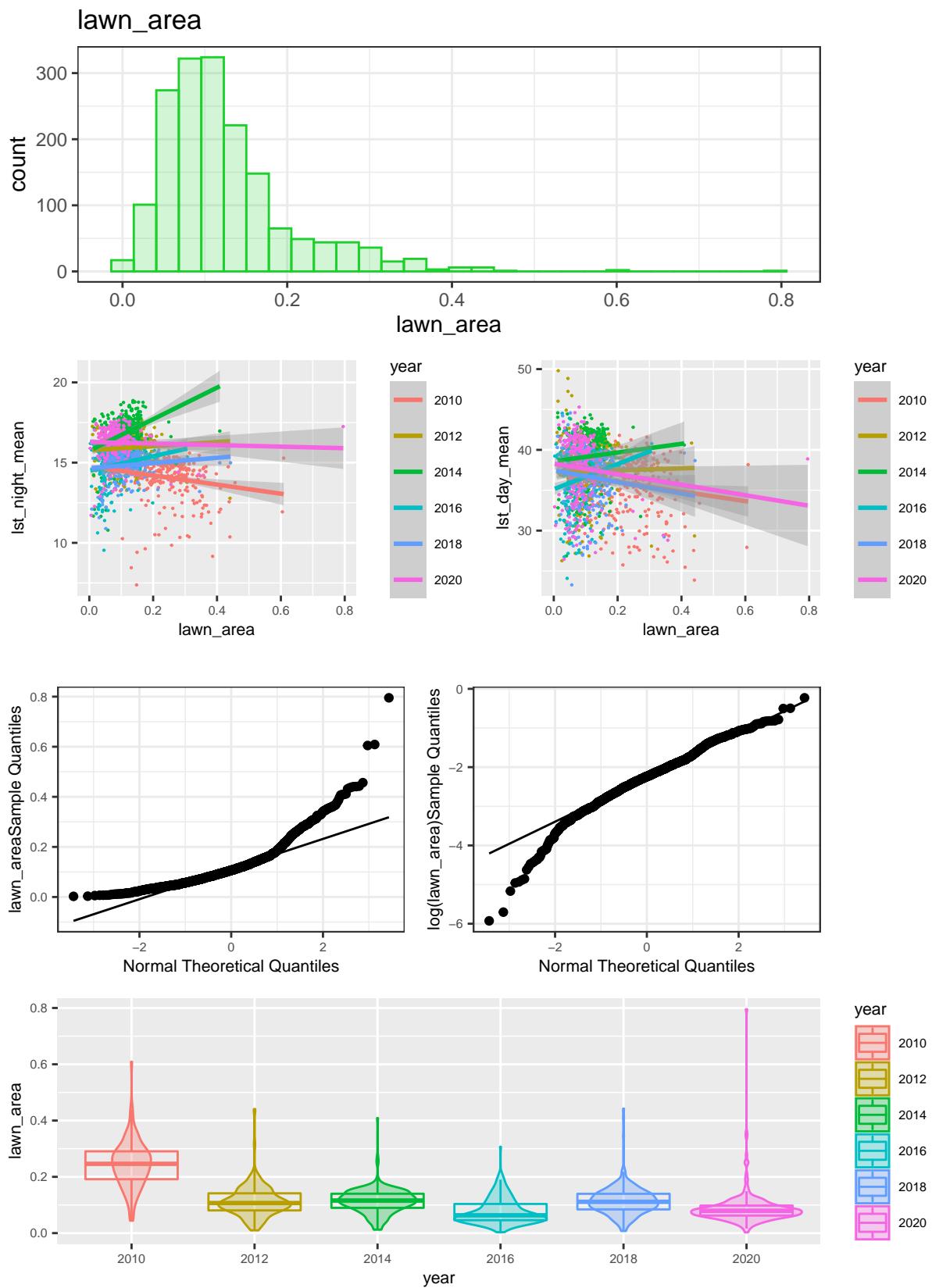
## Aggregate Impervious Area (Soil + Turf + Impervious)



## Impervious Area

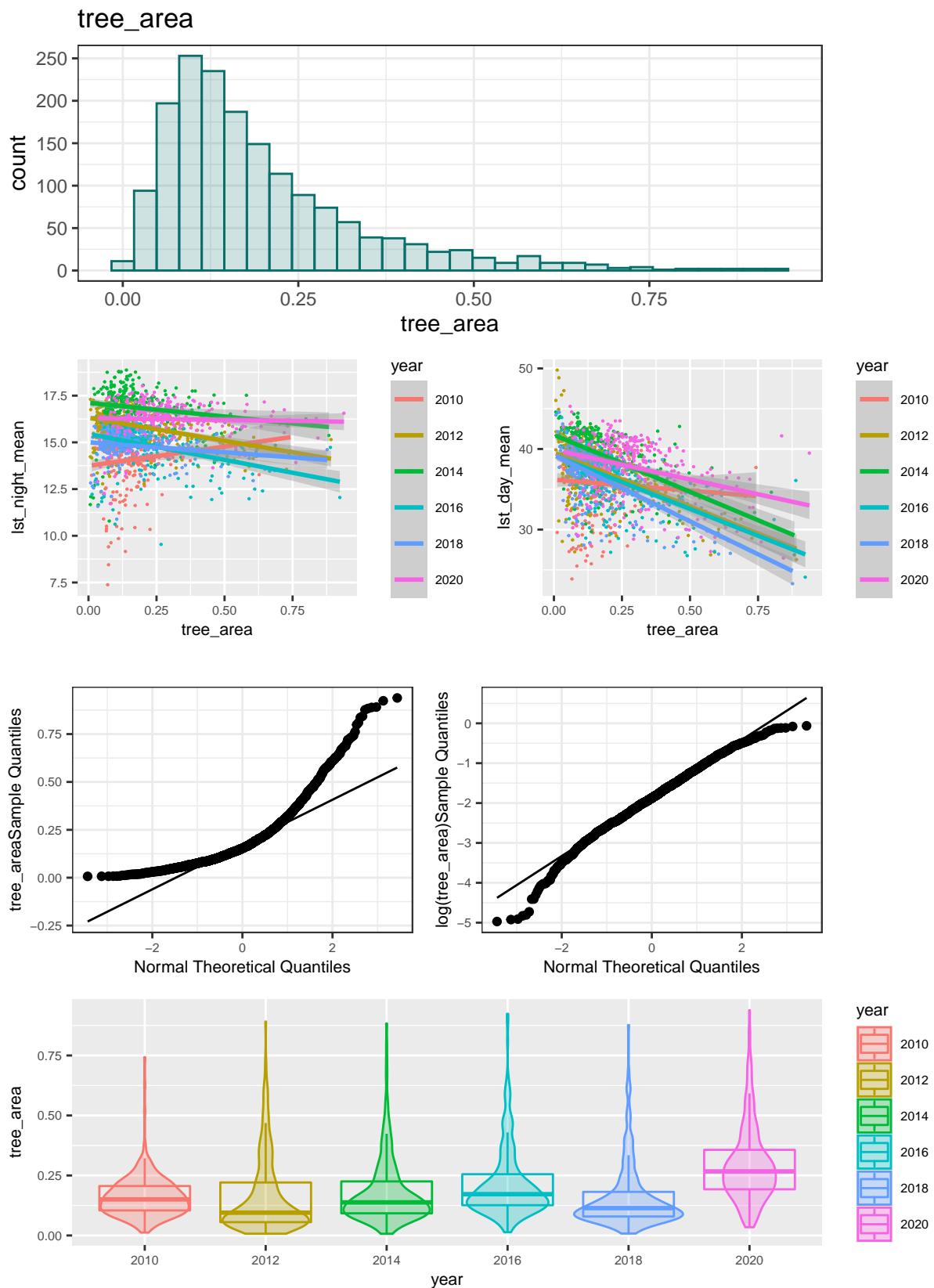


## Lawn Area

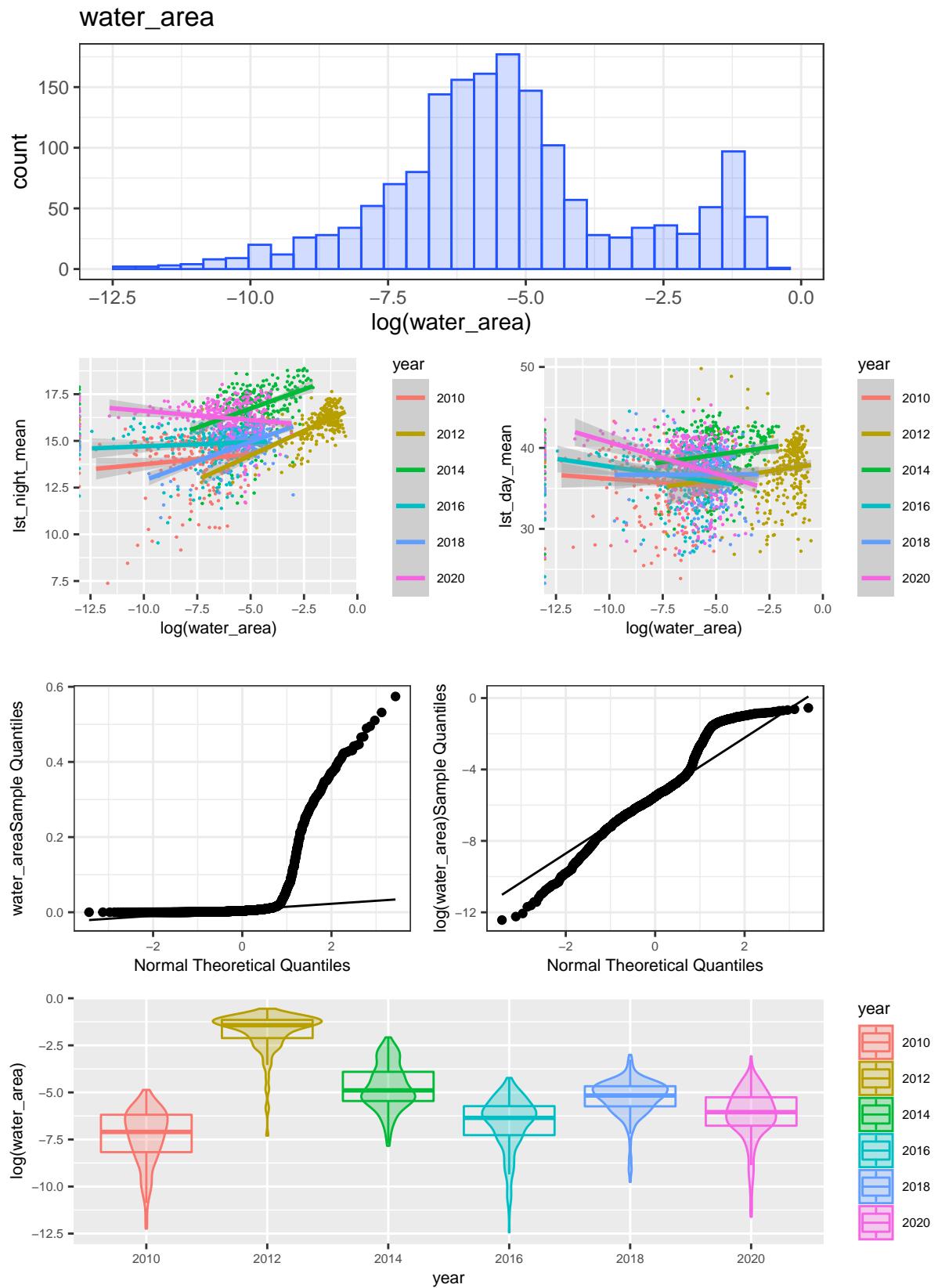


Google Earth: Random Forest Model

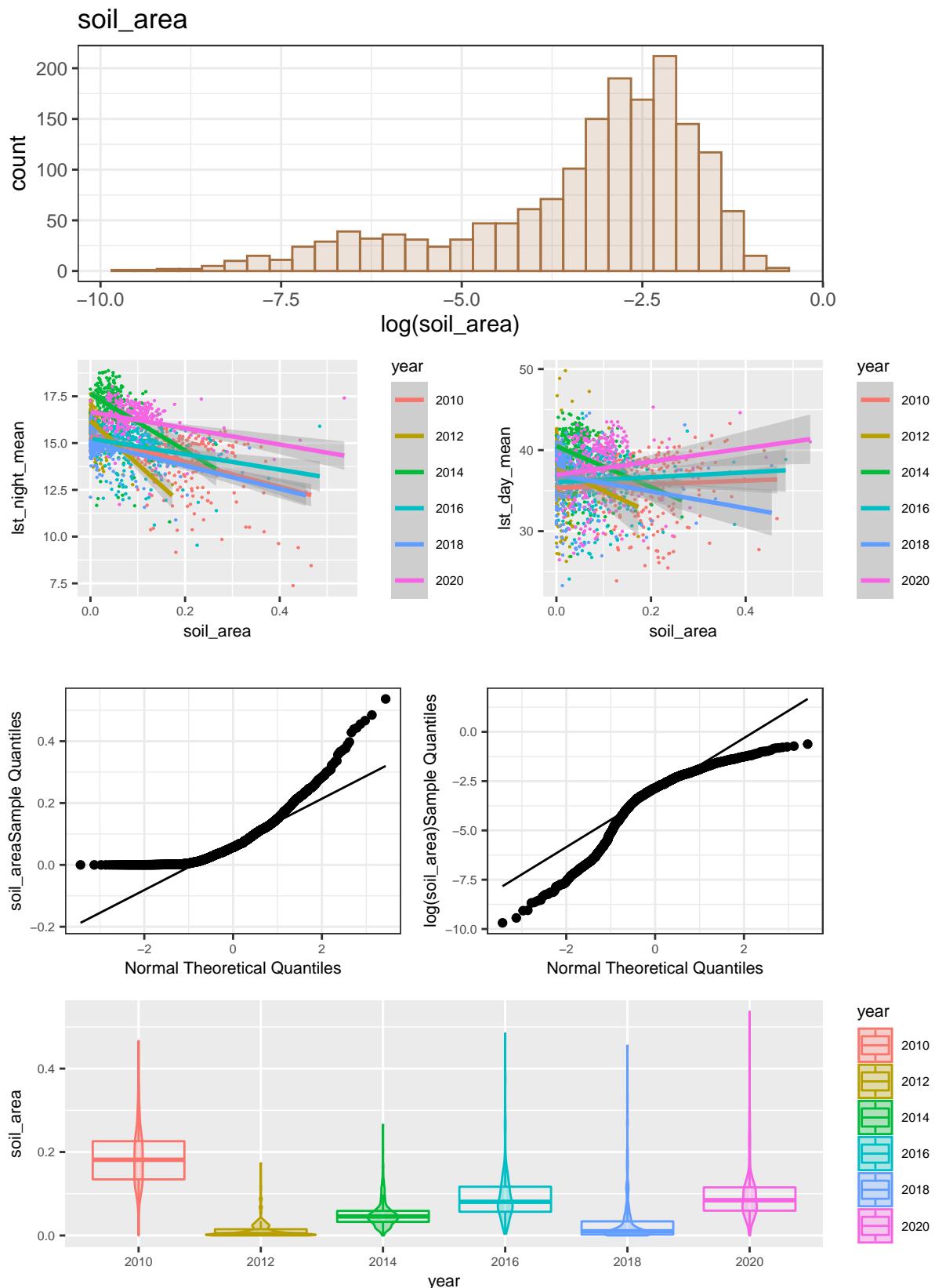
## Tree Area



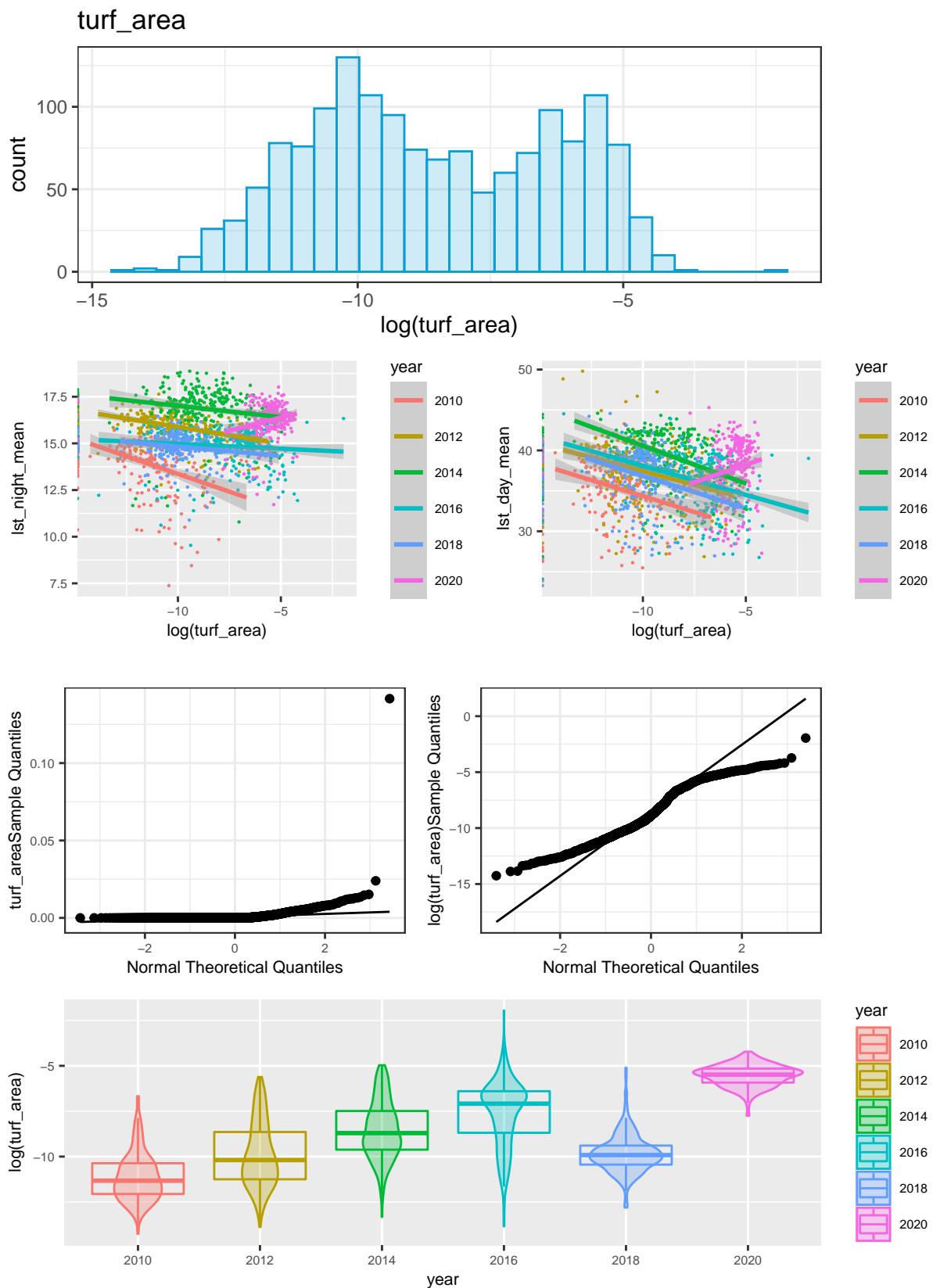
## Water Area



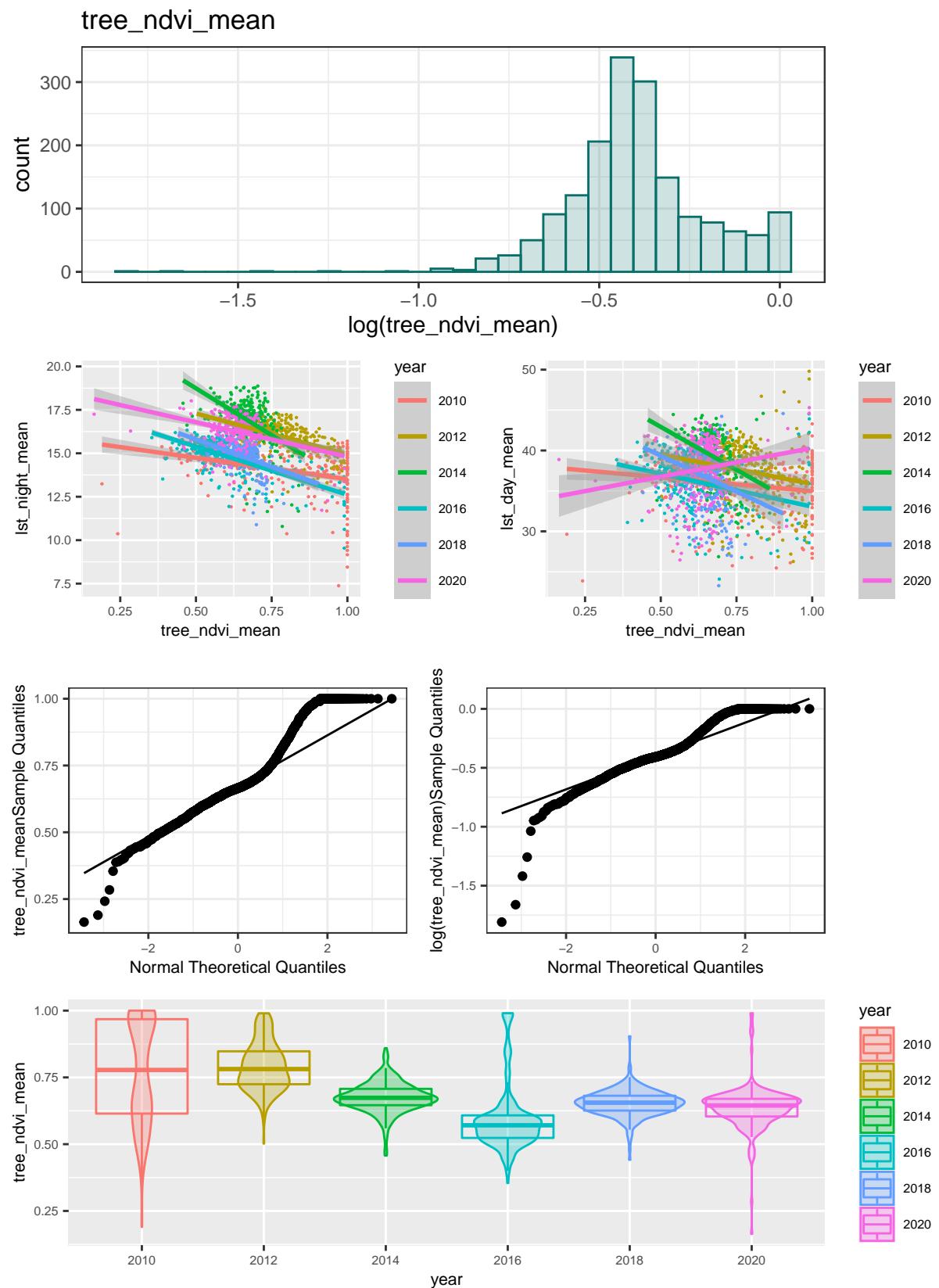
## Soil Area



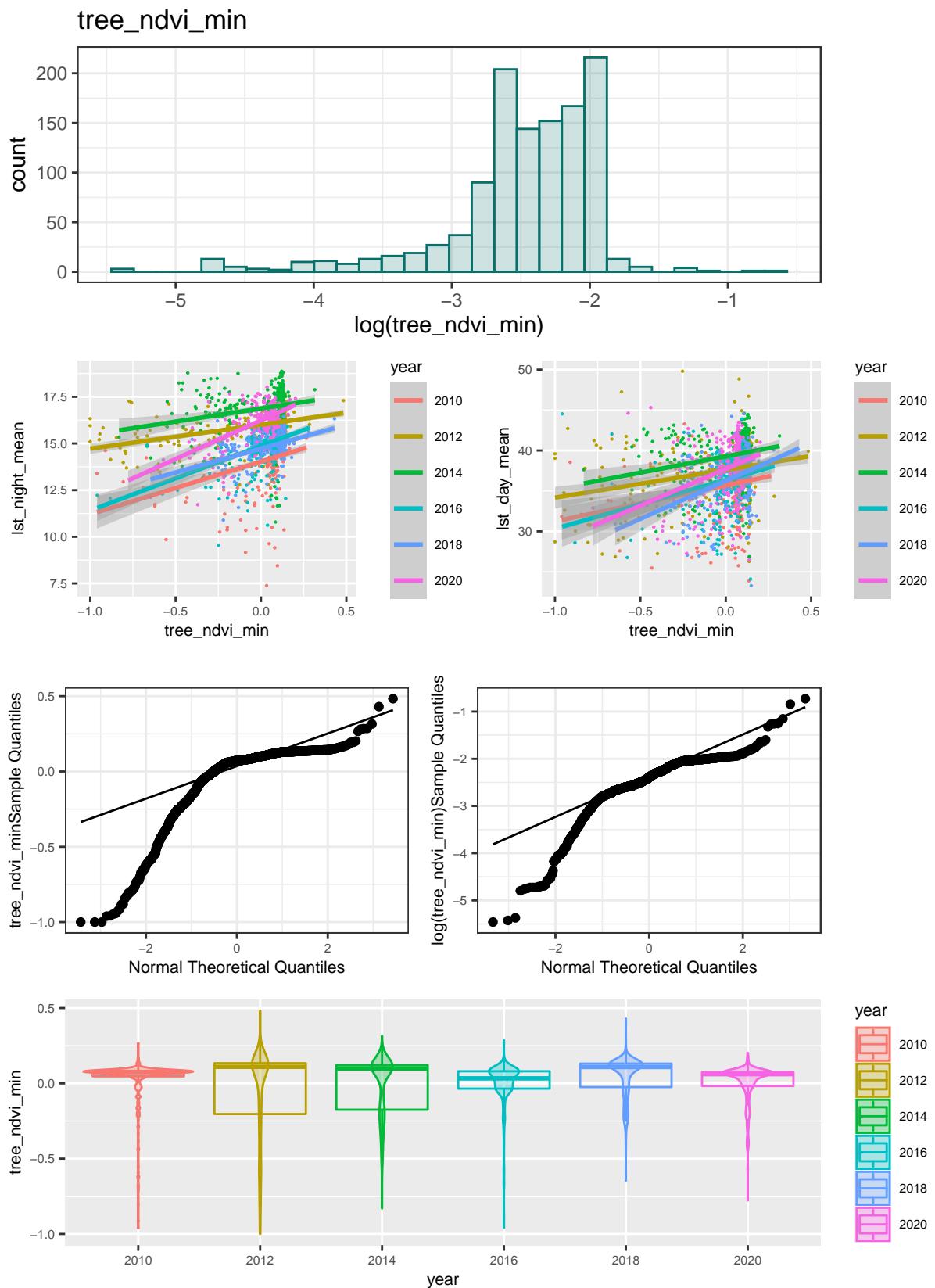
## Turf Area



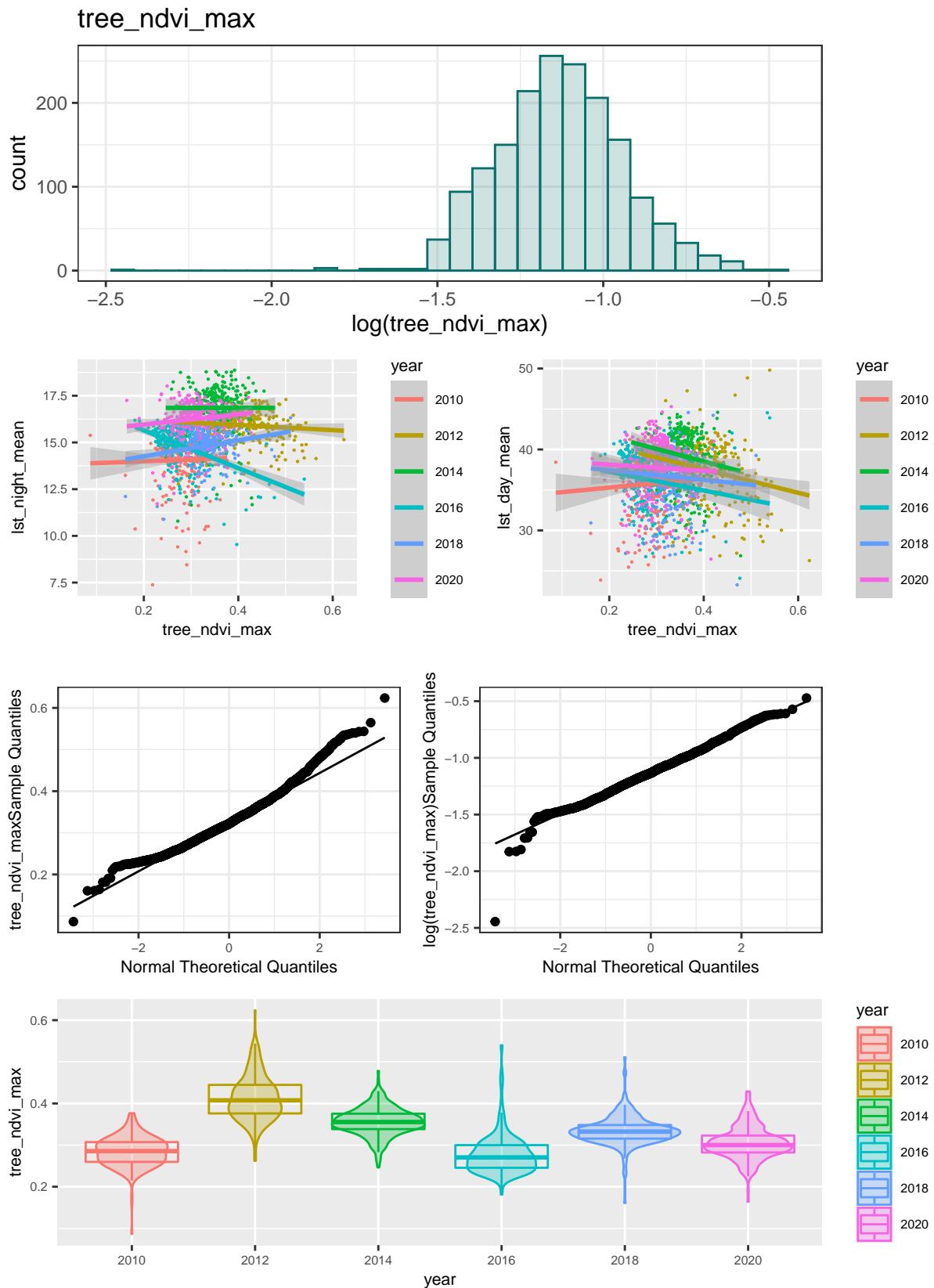
## Tree NDVI Mean



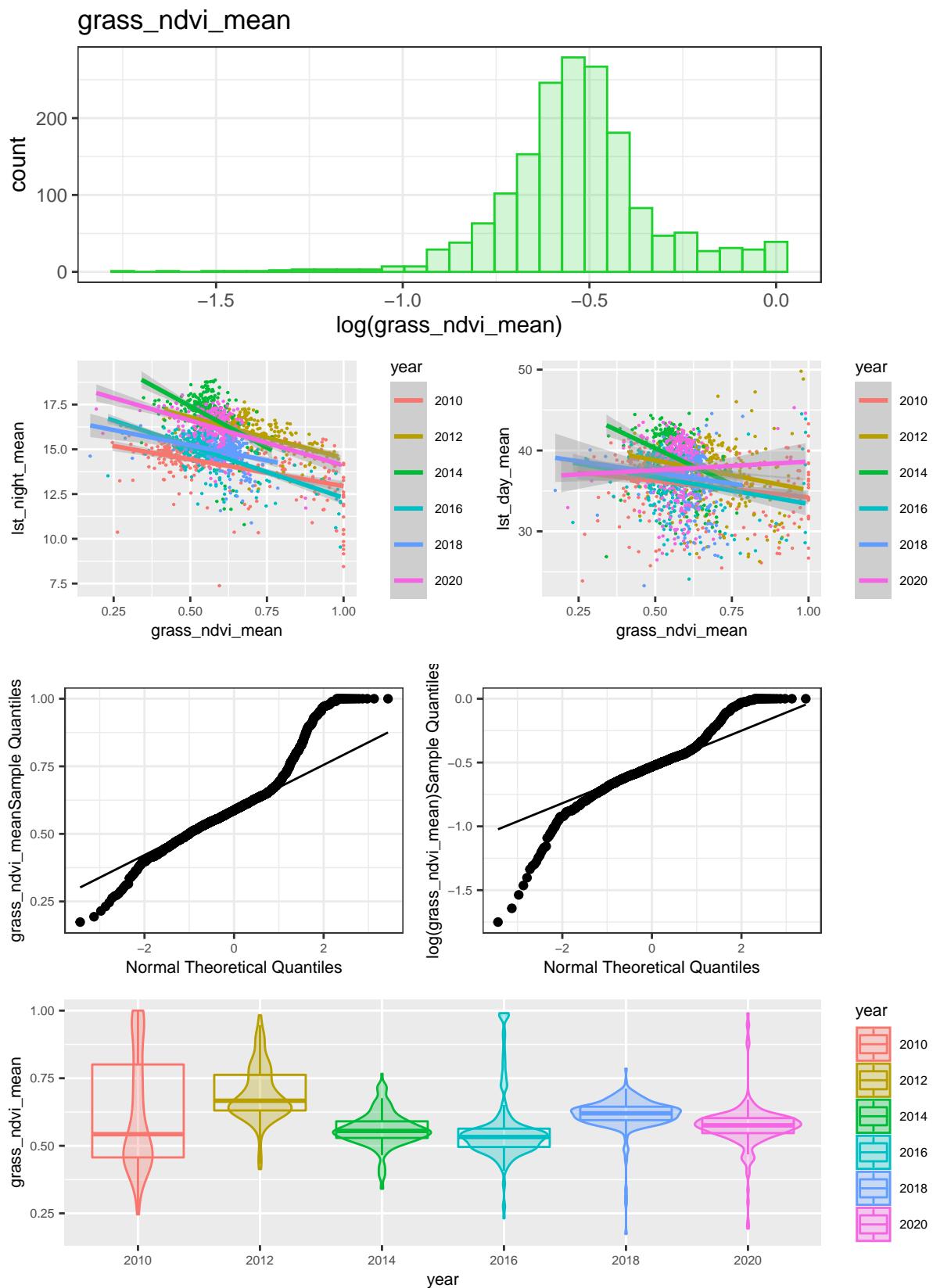
## Tree NDVI Min



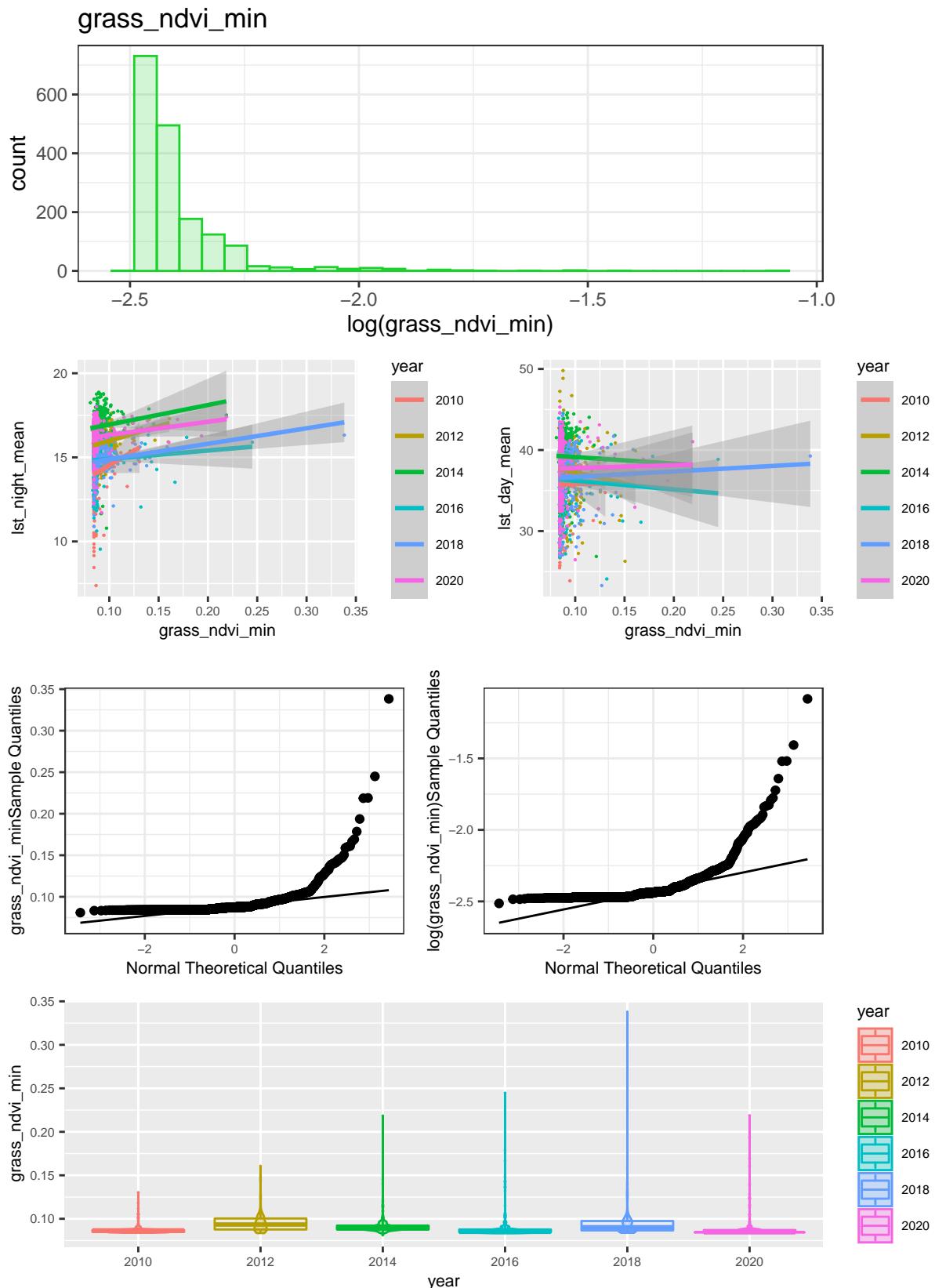
## Tree NDVI Max



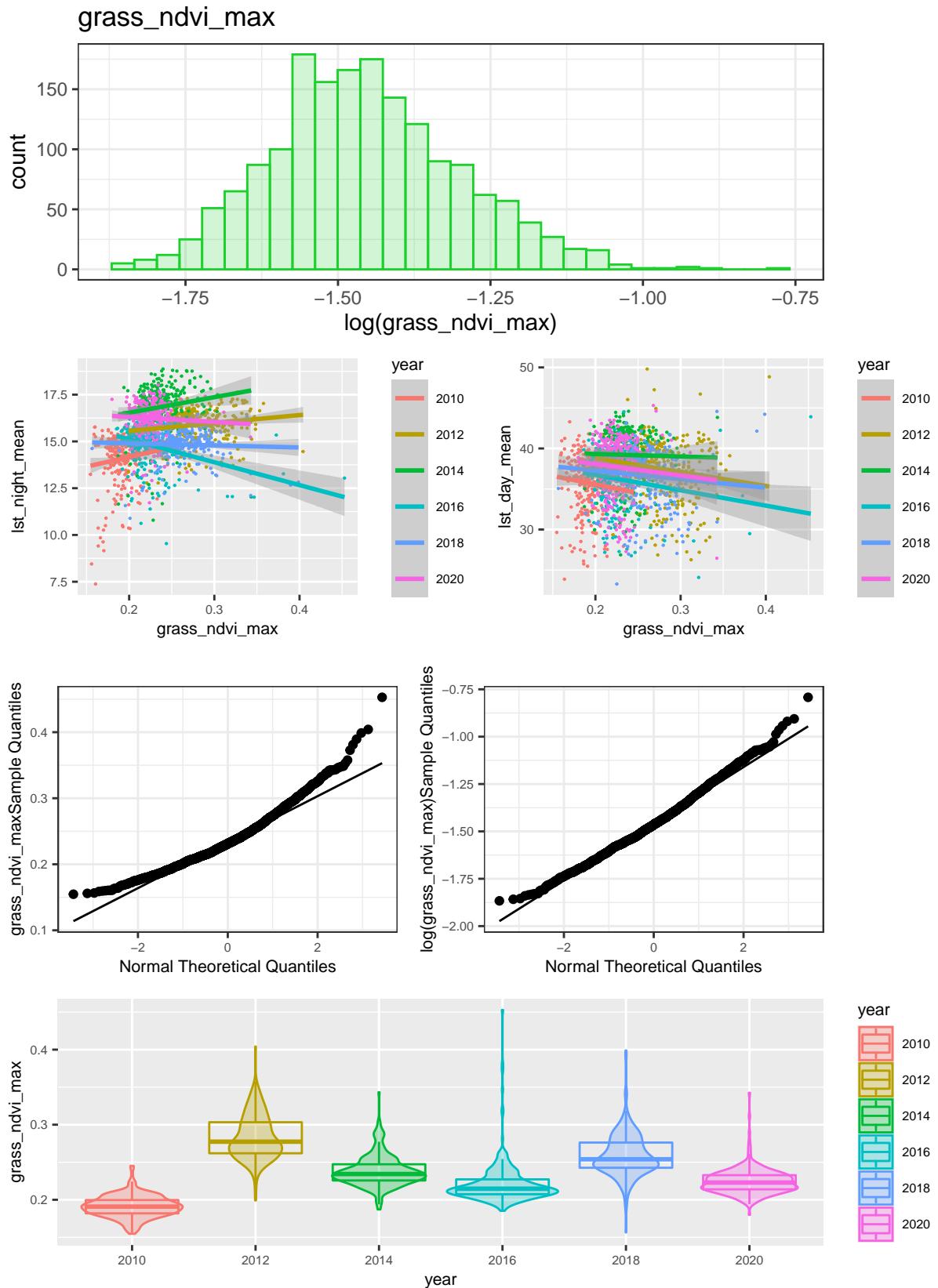
## Lawn NDVI Mean



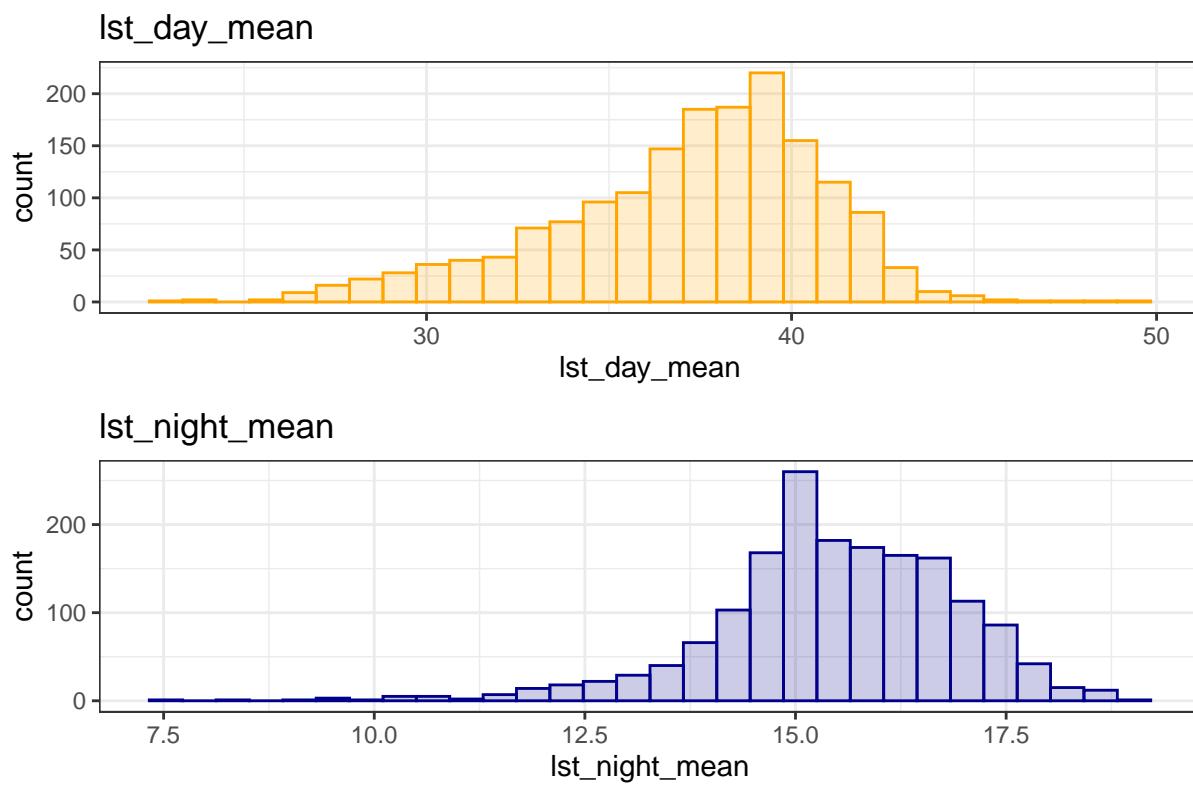
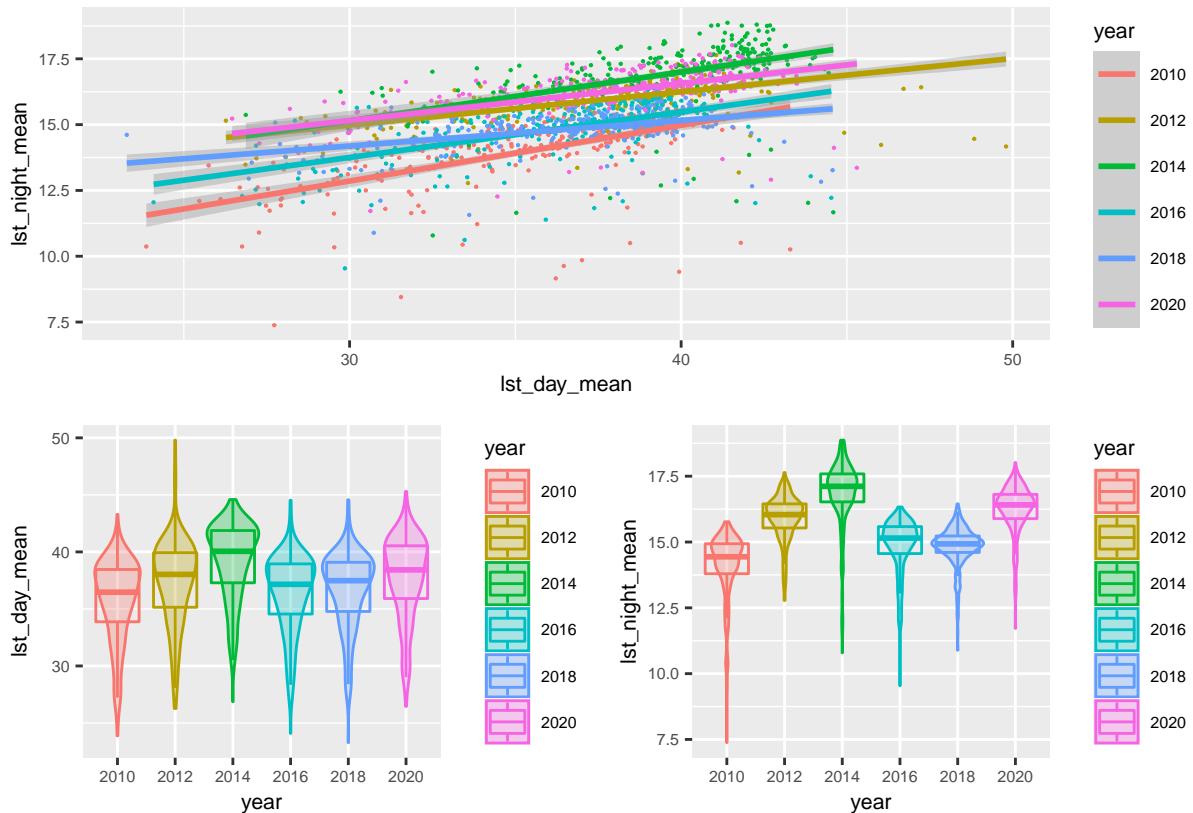
## Lawn NDVI Min



## Lawn NDVI Max



## Mean Temperature



## Create MicroClimate Panel Data Set

```
data <- subset(data, subset = year > 2012)

# Create panel dataframe object
microClimatePanel <- pdata.frame(data, index=c("zipcode", "year"))
```

## First Attempt with Traditional OLS Model

```
OLSM1 <- lm(
  log(lst_day_mean) ~ 0 + factor(year) + polygon_area + tree_area + lawn_area + water_area +
    turf_area + grass_ndvi_mean + tree_ndvi_mean,
  data = microClimatePanel)

OLSM2 <- lm(
  log(lst_night_mean) ~ 0 + factor(year) + polygon_area + tree_area + lawn_area + water_area +
    turf_area + grass_ndvi_mean + tree_ndvi_mean,
  data = microClimatePanel)

stargazer::stargazer(OLSM1, OLSM2, single.row = TRUE,
                      title = 'OLS Model',
                      column.labels = c("OLSM1 Day Temp",
                                        "OLSM2 Night Temp"))
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac  
at gmail.com % Date and time: Wed, Jul 27, 2022 - 7:23:34 AM

## Second Attempt with a Mixed Effects Model

```
feM1 <- plm(
  log(lst_day_mean) ~ 0 + factor(year) + tree_area + lawn_area + water_area + soil_area +
    turf_area + grass_ndvi_mean + tree_ndvi_mean,
  index = c(zipcode, year), data = microClimatePanel, model = 'within')

feM2 <- plm(
  log(lst_night_mean) ~ 0 + factor(year) + tree_area + lawn_area + water_area + soil_area +
    turf_area + grass_ndvi_mean + tree_ndvi_mean,
  index = c(zipcode, year), data = microClimatePanel, model = 'within')

stargazer::stargazer(feM1, feM2, single.row = TRUE,
                      title = 'Fixed Effects Model',
```

Table 1: OLS Model

	<i>Dependent variable:</i>	
	log(lst_day_mean) OLSModel Day Temp	log(lst_night_mean) OLSModel Night Temp
	(1)	(2)
factor(year)2014	3.766*** (0.024)	2.980*** (0.015)
factor(year)2016	3.708*** (0.021)	2.872*** (0.013)
factor(year)2018	3.701*** (0.023)	2.870*** (0.015)
factor(year)2020	3.773*** (0.022)	2.971*** (0.014)
polygon_area	-0.000*** (0.000)	-0.000*** (0.000)
tree_area	-0.347*** (0.018)	-0.054*** (0.012)
lawn_area	0.044 (0.050)	0.169*** (0.033)
water_area	0.066 (0.213)	0.838*** (0.137)
soil_area	0.043 (0.052)	-0.091*** (0.033)
turf_area	-0.430 (0.544)	-0.146 (0.352)
grass_ndvi_mean	-0.157** (0.063)	-0.213*** (0.041)
tree_ndvi_mean	0.069 (0.064)	-0.066 (0.041)
Observations	1,132	1,132
R <sup>2</sup>	0.999	1.000
Adjusted R <sup>2</sup>	0.999	1.000
Residual Std. Error (df = 1120)	0.082	0.053
F Statistic (df = 12; 1120)	181,679.300***	250,864.100***

Note:

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

```

column.labels = c("FixedEffects Day Temp",
                 "FixedEffects Night Temp"))

```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Wed, Jul 27, 2022 - 7:23:35 AM

Table 2: Fixed Effects Model

	<i>Dependent variable:</i>	
	log(lst_day_mean) FixedEffects Day Temp	log(lst_night_mean) FixedEffects Night Temp
	(1)	(2)
factor(year)2014	0.041*** (0.002)	0.038*** (0.003)
factor(year)2016	-0.040*** (0.002)	-0.091*** (0.003)
factor(year)2018	-0.032*** (0.002)	-0.079*** (0.003)
tree_area	-0.052*** (0.012)	-0.004 (0.018)
lawn_area	-0.065*** (0.020)	0.018 (0.030)
water_area	-0.168*** (0.052)	0.300*** (0.077)
soil_area	0.042*** (0.016)	0.169*** (0.023)
turf_area	0.096 (0.123)	0.020 (0.184)
grass_ndvi_mean	0.042** (0.016)	-0.047* (0.024)
tree_ndvi_mean	0.009 (0.015)	-0.005 (0.022)
Observations	1,132	1,132
R <sup>2</sup>	0.824	0.876
Adjusted R <sup>2</sup>	0.762	0.833
F Statistic (df = 10; 837)	391.058***	593.976***

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Compare Performance of OLS vs Mixed Effects Models

```

pFtest(feM1, OLSM1)

##
## F test for individual effects
##
## data: log(lst_day_mean) ~ 0 + factor(year) + tree_area + lawn_area + ...
## F = 97.99, df1 = 283, df2 = 837, p-value < 2.2e-16
## alternative hypothesis: significant effects

```

```
## Fixed effects is a better choice than OLS
```

Fixed Effects is a Better Choice, Discard OLS Model

Try a Random Effects Model

```
reM1 <- plm(
  log(lst_day_mean) ~ 0 + tree_area + lawn_area + water_area + soil_area +
    turf_area + grass_ndvi_mean + tree_ndvi_mean,
  index = c(zipcode, year), data = microClimatePanel, model = 'random')

reM2 <- plm(
  log(lst_night_mean) ~ 0 + tree_area + lawn_area + water_area + soil_area +
    turf_area + grass_ndvi_mean + tree_ndvi_mean,
  index = c(zipcode, year), data = microClimatePanel, model = 'random')

stargazer::stargazer(reM1,reM2,
                      single.row = TRUE,
                      title = 'Random Effects Model',
                      column.labels = c("RandomEffects Day Temp",
                                       "RandomEffects Night Temp"))
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Wed, Jul 27, 2022 - 7:23:36 AM

Compare Performance of Mixed Effects vs Random Effects Models

```
phtest(feM1, reM1)
```

```
##
## Hausman Test
##
## data: log(lst_day_mean) ~ 0 + factor(year) + tree_area + lawn_area + ...
## chisq = 452.91, df = 7, p-value < 2.2e-16
## alternative hypothesis: one model is inconsistent

## the p-value is significant so we choose fixed effects
## (since the unique errors are correlated with the regressors).
```

Table 3: Random Effects Model

	Dependent variable:		
	log(lst_day_mean)	log(lst_night_mean)	
	RandomEffects Day Temp	RandomEffects Night Temp	
	(1)	(2)	
tree_area	0.447*** (0.086)	0.922*** (0.095)	
lawn_area	0.981*** (0.162)	1.816*** (0.181)	
water_area	1.146*** (0.434)	2.443*** (0.505)	
soil_area	0.342*** (0.118)	0.746*** (0.136)	
turf_area	1.107 (1.093)	2.382* (1.272)	
grass_ndvi_mean	0.333*** (0.114)	0.524*** (0.129)	
tree_ndvi_mean	0.853*** (0.105)	1.413*** (0.120)	
Observations	1,132	1,132	
R <sup>2</sup>	0.107	0.114	
Adjusted R <sup>2</sup>	0.102	0.109	
F Statistic	450.366***	1,767.633***	

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

the p-value is significant so we choose fixed effects (since the unique errors are correlated with the regressors). There is omitted variable bias at the higher level that the RE model has not accounted for (but the FE model has).

## Try a Fixed Effects Model with Fixed Time

```
feM2FixedTime <- plm(
  log(lst_day_mean) ~ 0 + tree_area + lawn_area + water_area + soil_area +
    turf_area + grass_ndvi_mean + tree_ndvi_mean,
  index = c(zipcode, year), data = microClimatePanel, model = 'within')

stargazer::stargazer(feM2FixedTime, single.row = TRUE)
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Wed, Jul 27, 2022 - 7:23:36 AM

```
pFtest(feM1,feM2FixedTime)
```

```
## 
## F test for individual effects
## 
## data: log(lst_day_mean) ~ 0 + factor(year) + tree_area + lawn_area + ...
## F = 620.68, df1 = 3, df2 = 837, p-value < 2.2e-16
## alternative hypothesis: significant effects
```

Table 4:

<i>Dependent variable:</i>	
	log(lst_day_mean)
tree_area	0.064*** (0.017)
lawn_area	0.203*** (0.032)
water_area	0.685*** (0.085)
soil_area	-0.026 (0.023)
turf_area	0.190 (0.214)
grass_ndvi_mean	-0.368*** (0.023)
tree_ndvi_mean	0.386*** (0.021)
Observations	1,132
R <sup>2</sup>	0.431
Adjusted R <sup>2</sup>	0.235
F Statistic	91.079*** (df = 7; 840)

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

```
plmtest(feM1, effect="time", type="bp")
```

```
##
## Lagrange Multiplier Test - time effects (Breusch-Pagan) for unbalanced
## panels
##
## data: log(lst_day_mean) ~ 0 + factor(year) + tree_area + lawn_area + ...
## chisq = 2.0071, df = 1, p-value = 0.1566
## alternative hypothesis: significant effects
```

```
pbgtest(feM1)
```

```
##
## Breusch-Godfrey/Wooldridge test for serial correlation in panel models
##
## data: log(lst_day_mean) ~ 0 + factor(year) + tree_area + lawn_area + ...
## chisq = 55.205, df = 1, p-value = 1.086e-13
## alternative hypothesis: serial correlation in idiosyncratic errors
```

```
coeftest(feM1, vcovHC)
```

```
##
## t test of coefficients:
##
##             Estimate Std. Error   t value Pr(>|t|) 
## factor(year)2014  0.0409630  0.0023898  17.1407 < 2.2e-16 ***
##
```

```

## factor(year)2016 -0.0402422  0.0018599 -21.6370 < 2.2e-16 ***
## factor(year)2018 -0.0319612  0.0024204 -13.2048 < 2.2e-16 ***
## tree_area        -0.0523011  0.0148201 -3.5291 0.0004398 ***
## lawn_area        -0.0652285  0.0286661 -2.2755 0.0231304 *
## water_area       -0.1679960  0.0413353 -4.0642 5.273e-05 ***
## soil_area        0.0422364  0.0197583  2.1377 0.0328339 *
## turf_area        0.0956626  0.0533841  1.7920 0.0734991 .
## grass_ndvi_mean 0.0417854  0.0165033  2.5319 0.0115249 *
## tree_ndvi_mean   0.0090166  0.0162302  0.5555 0.5786732
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```
coeftest(feM1, vcovHC(feM1, method = "arellano"))
```

```

##
## t test of coefficients:
##
##                               Estimate Std. Error  t value Pr(>|t|)
## factor(year)2014  0.0409630  0.0023898 17.1407 < 2.2e-16 ***
## factor(year)2016 -0.0402422  0.0018599 -21.6370 < 2.2e-16 ***
## factor(year)2018 -0.0319612  0.0024204 -13.2048 < 2.2e-16 ***
## tree_area        -0.0523011  0.0148201 -3.5291 0.0004398 ***
## lawn_area        -0.0652285  0.0286661 -2.2755 0.0231304 *
## water_area       -0.1679960  0.0413353 -4.0642 5.273e-05 ***
## soil_area        0.0422364  0.0197583  2.1377 0.0328339 *
## turf_area        0.0956626  0.0533841  1.7920 0.0734991 .
## grass_ndvi_mean 0.0417854  0.0165033  2.5319 0.0115249 *
## tree_ndvi_mean   0.0090166  0.0162302  0.5555 0.5786732
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```
t(sapply(c("HC0", "HC1", "HC2", "HC3", "HC4"), function(x) sqrt(diag(vcovHC(feM1,
```

	factor(year)2014	factor(year)2016	factor(year)2018	tree_area	lawn_area
## HC0	0.002389807	0.001859880	0.002420429	0.01482012	0.02866607
## HC1	0.002400433	0.001868150	0.002431191	0.01488601	0.02879354
## HC2	0.002415570	0.001877275	0.002443308	0.01507436	0.02923489
## HC3	0.002443217	0.001896080	0.002467702	0.01533764	0.02982467
## HC4	0.002489705	0.001926482	0.002505385	0.01578775	0.03091299
	water_area	soil_area	turf_area	grass_ndvi_mean	tree_ndvi_mean
## HC0	0.04133528	0.01975830	0.05338408	0.01650326	0.01623023
## HC1	0.04151907	0.01984615	0.05362145	0.01657664	0.01630240
## HC2	0.04171210	0.02009951	0.05666708	0.01668467	0.01640315
## HC3	0.04209604	0.02045095	0.06370743	0.01687048	0.01657960
## HC4	0.04238859	0.02103423	0.10820002	0.01708566	0.01677279

```
totalRobust <- coeftest(feM1, vcov = vcovHC(feM1, type = 'HCO'))
cInterval <- coefci(feM1, vcov. = vcovHC(feM1, type = 'HCO'))
```

```
print(totalRobust)
```

```

## 
## t test of coefficients:
## 

##                               Estimate Std. Error   t value Pr(>|t|)    
## factor(year)2014    0.0409630  0.0023898 17.1407 < 2.2e-16 ***
## factor(year)2016   -0.0402422  0.0018599 -21.6370 < 2.2e-16 ***
## factor(year)2018   -0.0319612  0.0024204 -13.2048 < 2.2e-16 ***
## tree_area        -0.0523011  0.0148201 -3.5291 0.0004398 ***
## lawn_area         -0.0652285  0.0286661 -2.2755 0.0231304 *  
## water_area        -0.1679960  0.0413353 -4.0642 5.273e-05 ***
## soil_area          0.0422364  0.0197583  2.1377 0.0328339 *  
## turf_area          0.0956626  0.0533841  1.7920 0.0734991 .  
## grass_ndvi_mean   0.0417854  0.0165033  2.5319 0.0115249 *  
## tree_ndvi_mean    0.0090166  0.0162302  0.5555 0.5786732  
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```
print(cInterval)
```

```

##                                     2.5 %    97.5 %
## factor(year)2014  0.036272319  0.045653756
## factor(year)2016 -0.043892796 -0.036591641
## factor(year)2018 -0.036712032 -0.027210386
## tree_area          -0.081390018 -0.023212103
## lawn_area          -0.121494335 -0.008962668
## water_area         -0.249128985 -0.086863039
## soil_area          0.003454781  0.081018044
## turf_area          -0.009119798  0.200445000
## grass_ndvi_mean   0.009392787  0.074178055
## tree_ndvi_mean    -0.022840181  0.040873298

```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Wed, Jul 27, 2022 - 7:23:38 AM

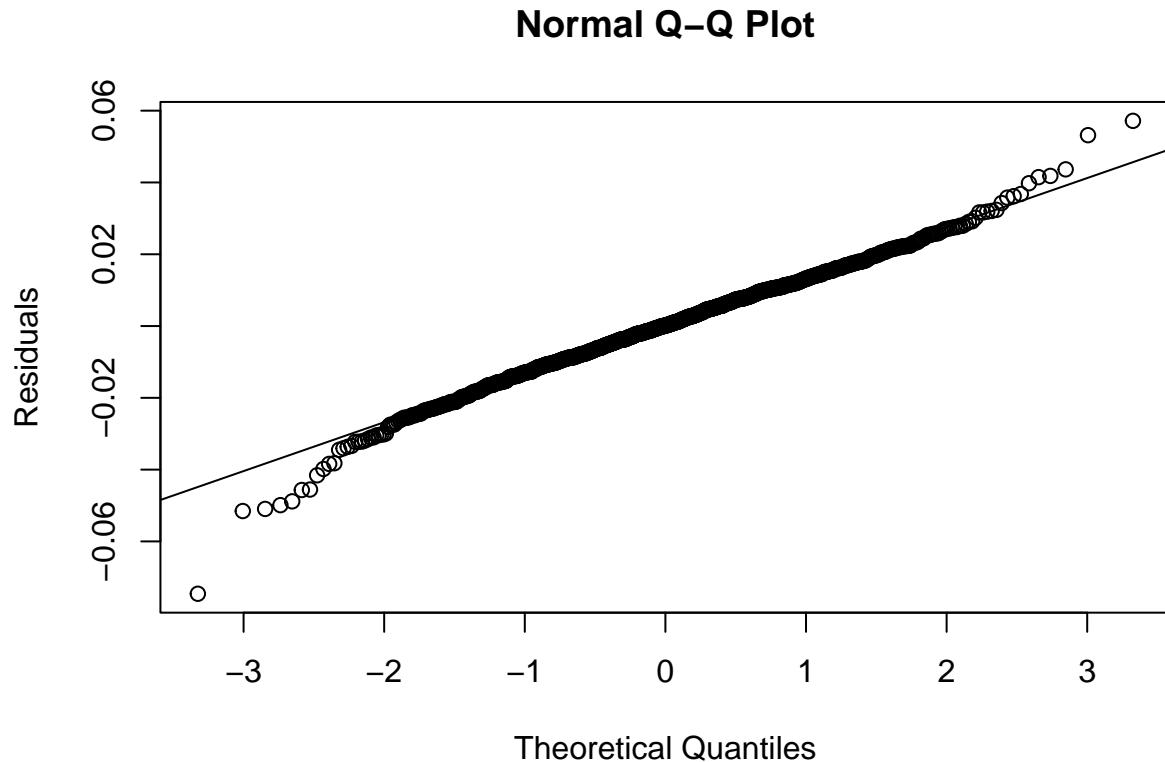
Table 5: Regression Models Sumamry

	Dependent variable: log(lst_day_mean)				
	OLS		panel linear		
	OLS	FixedEffects	RandomEffects	FixedEffects	FixedTime
	(1)	(2)	(3)	(4)	
factor(year)2014	3.766*** (0.024)	0.041*** (0.002)			
factor(year)2016	3.708*** (0.021)	-0.040*** (0.002)			
factor(year)2018	3.701*** (0.023)	-0.032*** (0.002)			
factor(year)2020	3.773*** (0.022)				
polygon_area	-0.000*** (0.000)				
tree_area	-0.347*** (0.018)	-0.052*** (0.012)	0.447*** (0.086)	0.064*** (0.017)	
lawn_area	0.044 (0.050)	-0.065*** (0.020)	0.981*** (0.162)	0.203*** (0.032)	
water_area	0.066 (0.213)	-0.168*** (0.052)	1.146*** (0.434)	0.685*** (0.085)	
soil_area	0.043 (0.052)	0.042*** (0.016)	0.342*** (0.118)	-0.026 (0.023)	
turf_area	-0.430 (0.544)	0.096 (0.123)	1.107 (1.093)	0.190 (0.214)	
grass_ndvi_mean	-0.157** (0.063)	0.042** (0.016)	0.333*** (0.114)	-0.368*** (0.023)	
tree_ndvi_mean	0.069 (0.064)	0.009 (0.015)	0.853*** (0.105)	0.386*** (0.021)	
Observations	1,132	1,132	1,132	1,132	
R <sup>2</sup>	0.999	0.824	0.107	0.431	
Adjusted R <sup>2</sup>	0.999	0.762	0.102	0.235	
Residual Std. Error	0.082 (df = 1120)				
F Statistic	181,679.300*** (df = 12; 1120)	391.058*** (df = 10; 837)	450.366***	91.079*** (df = 7; 840)	

Note:

\*p<0.1; \*\*p<0.05; \*\*\* p<0.01

```
qqnorm(residuals(feM1), ylab = 'Residuals')
qqline(residuals(feM1))
```



```
hist(residuals(feM1), xlab = 'Residuals')
```

### Histogram of residuals(feM1)

