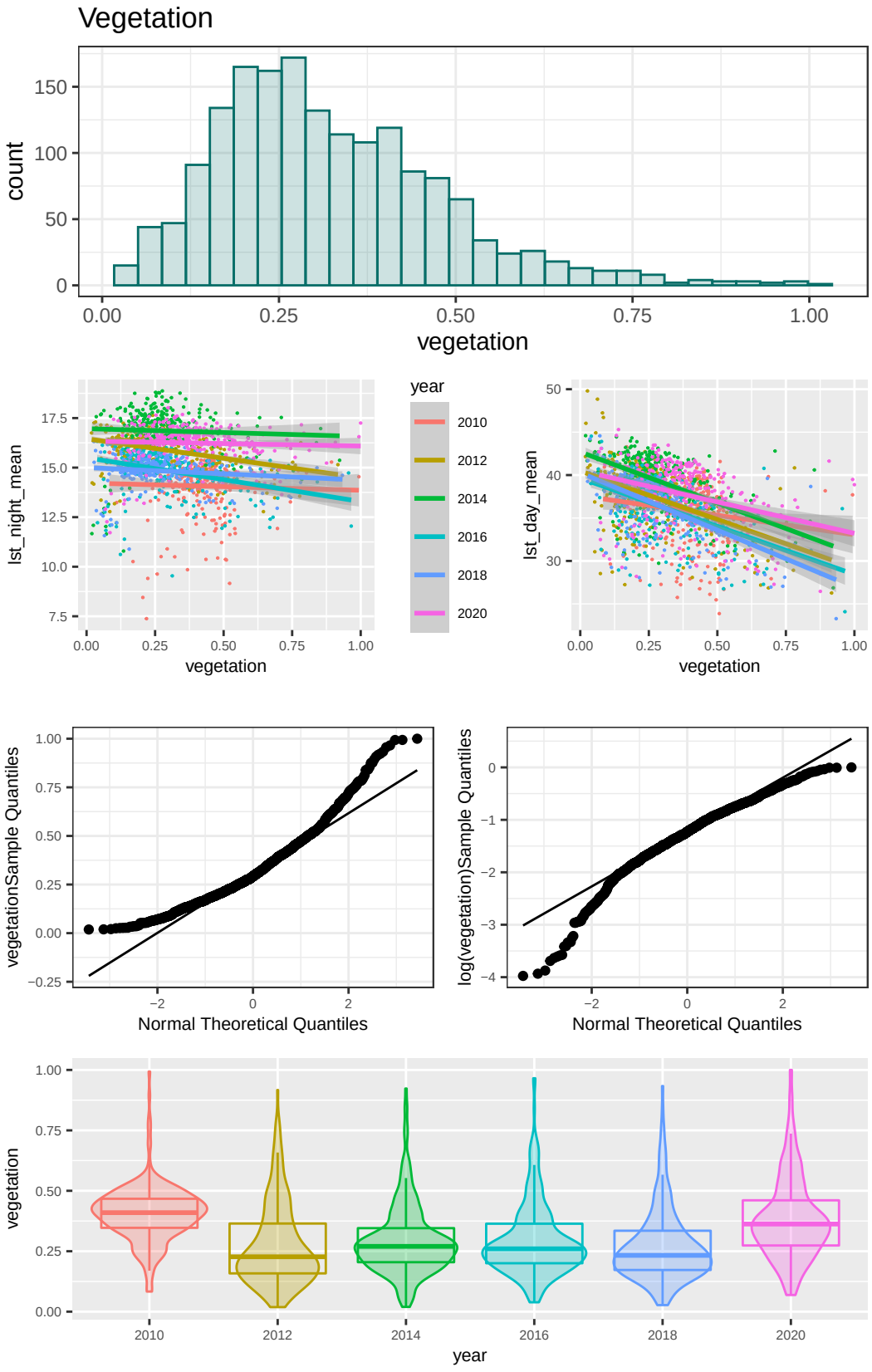
LawnProject Panel Data Analysis / Random Forest Data

LawnProject Team

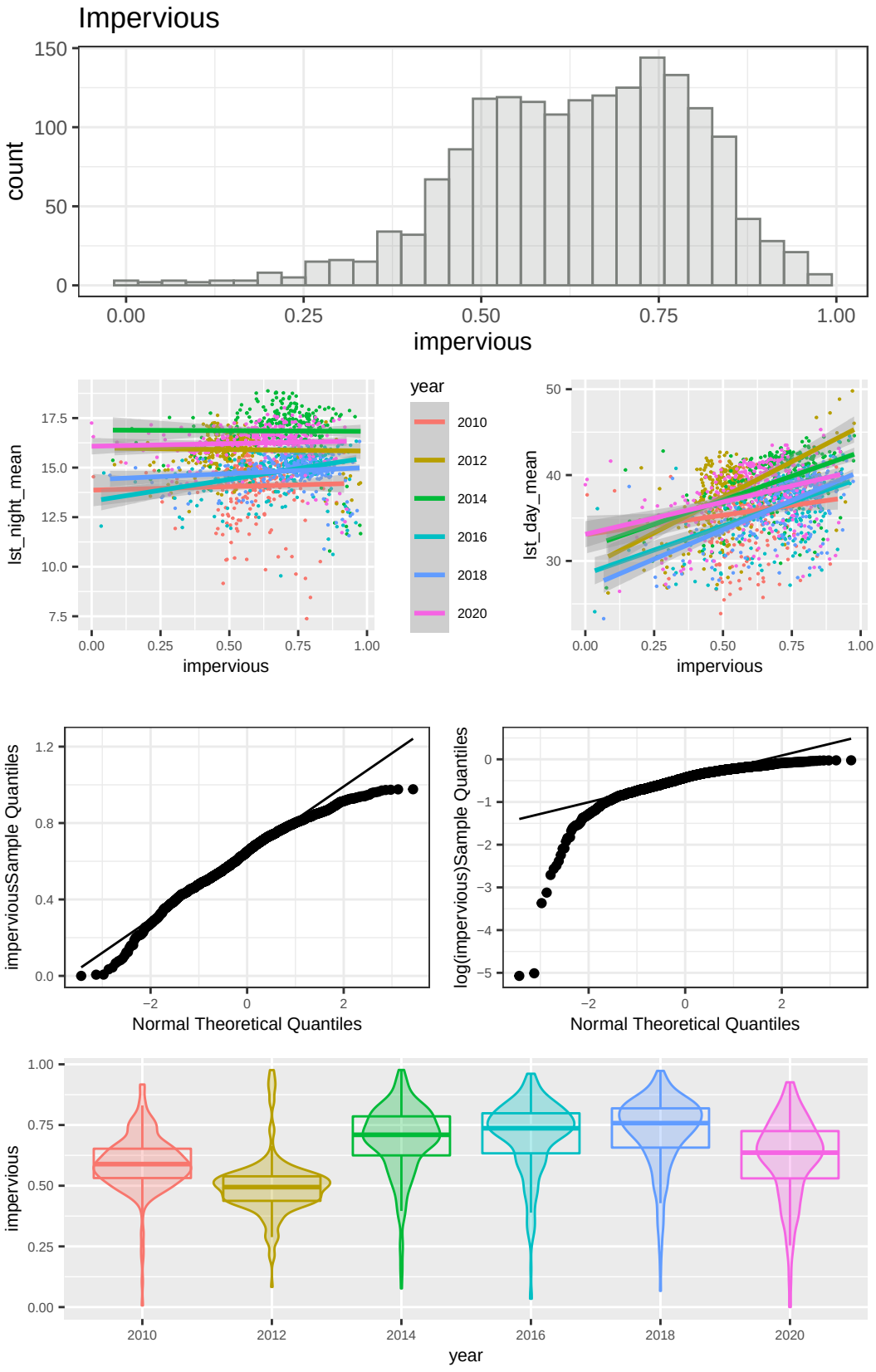
Intro

Aggregate Vegetation Area (Trees + Lawn)



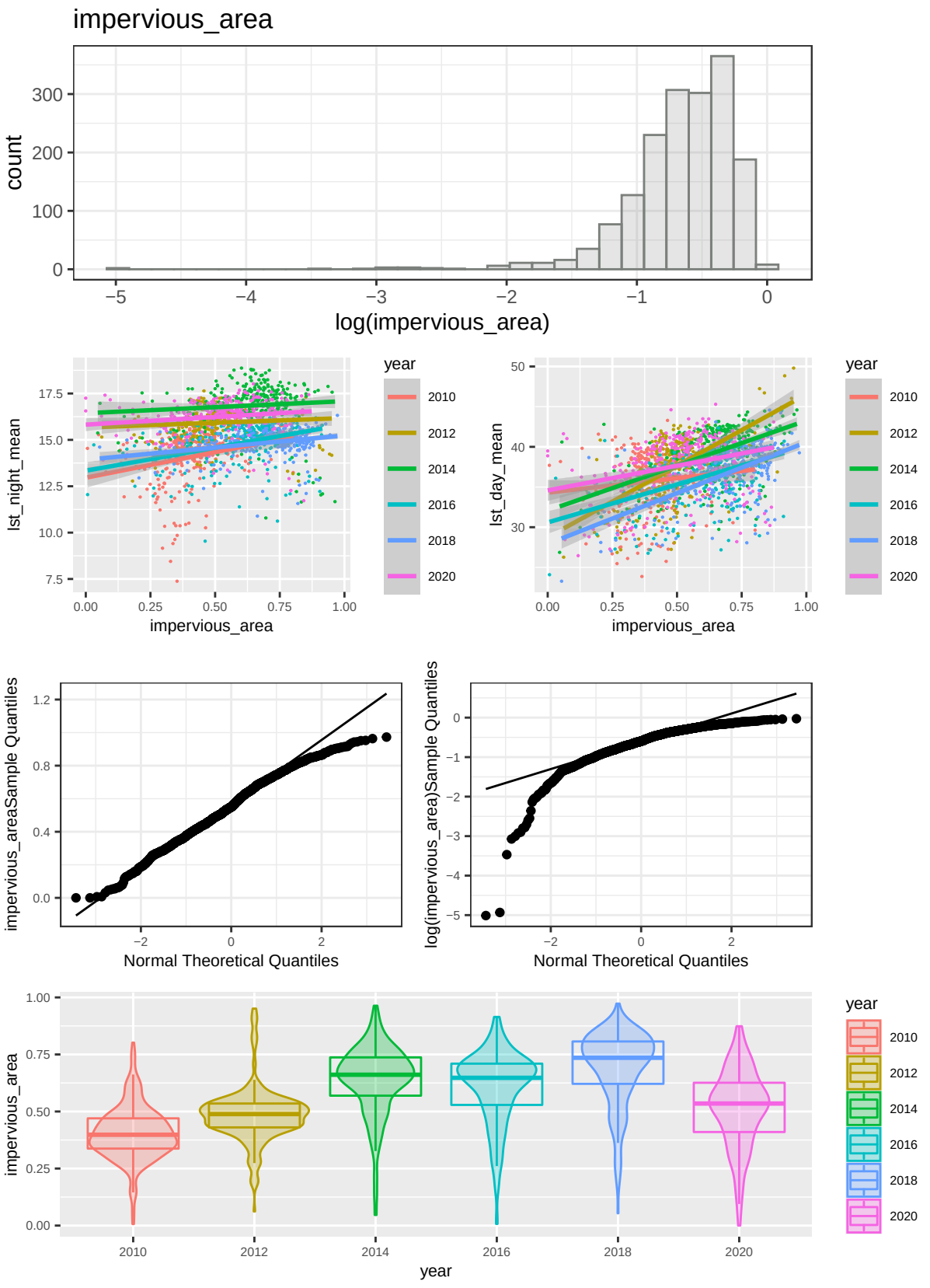
Google Earth: Random Forest Model

Aggregate Impervious Area (Soil + Turf + Impervious)



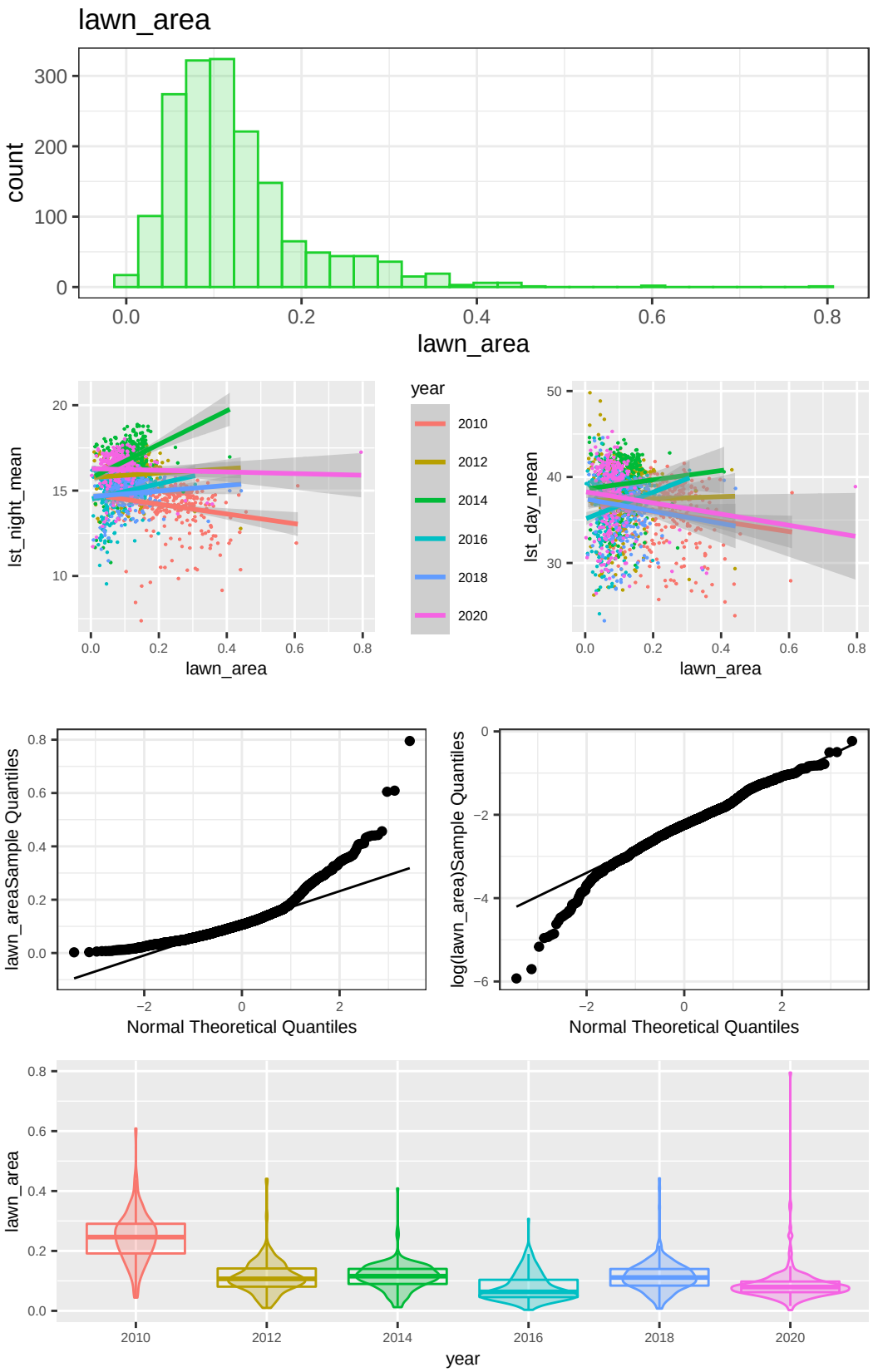
Google Earth: Random Forest Model

Impervious Area



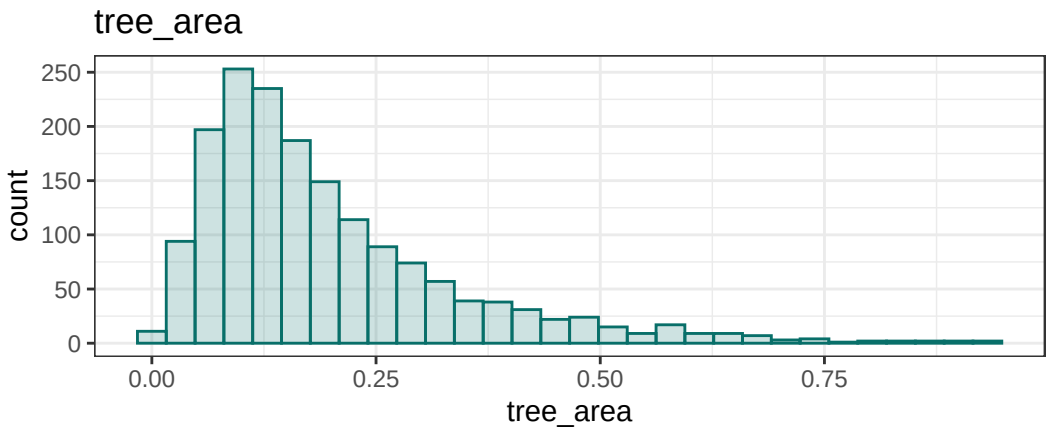
Google Earth: Random Forest Model

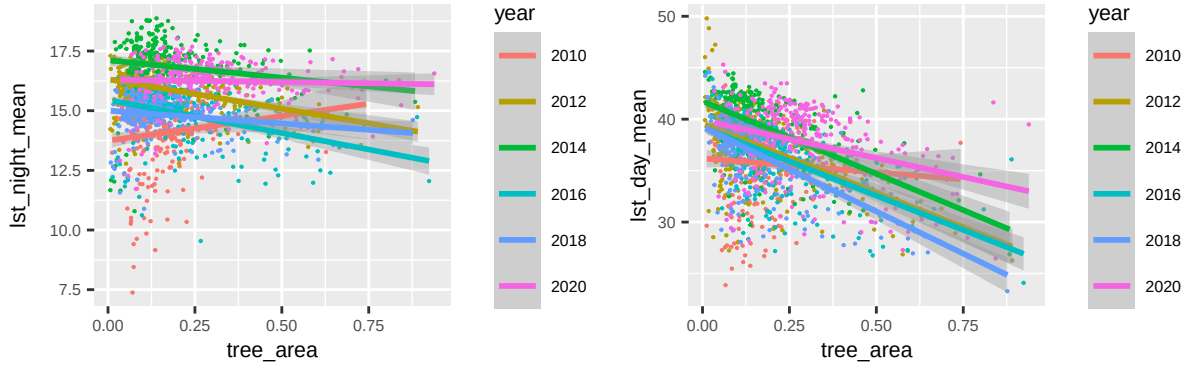
Lawn Area

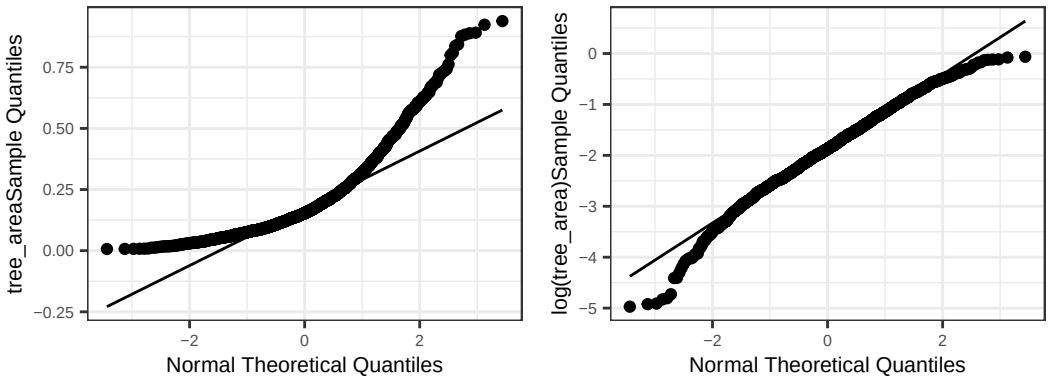


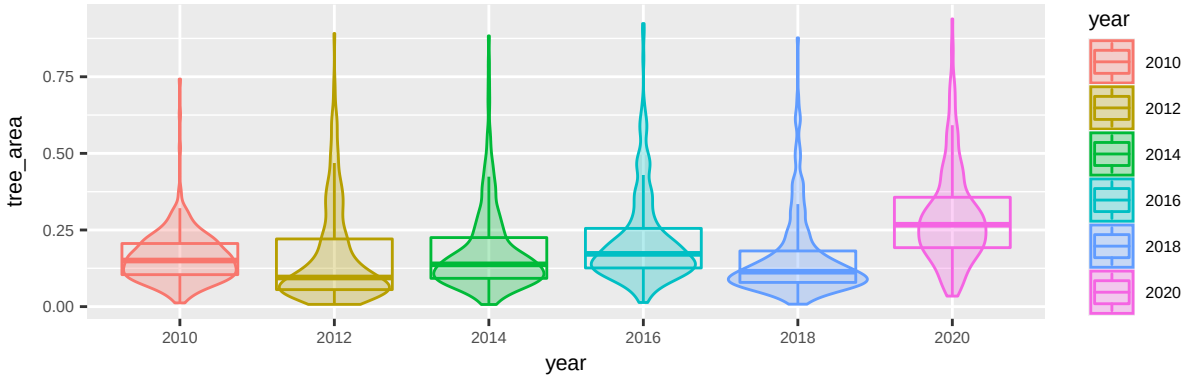
Google Earth: Random Forest Model

Tree Area



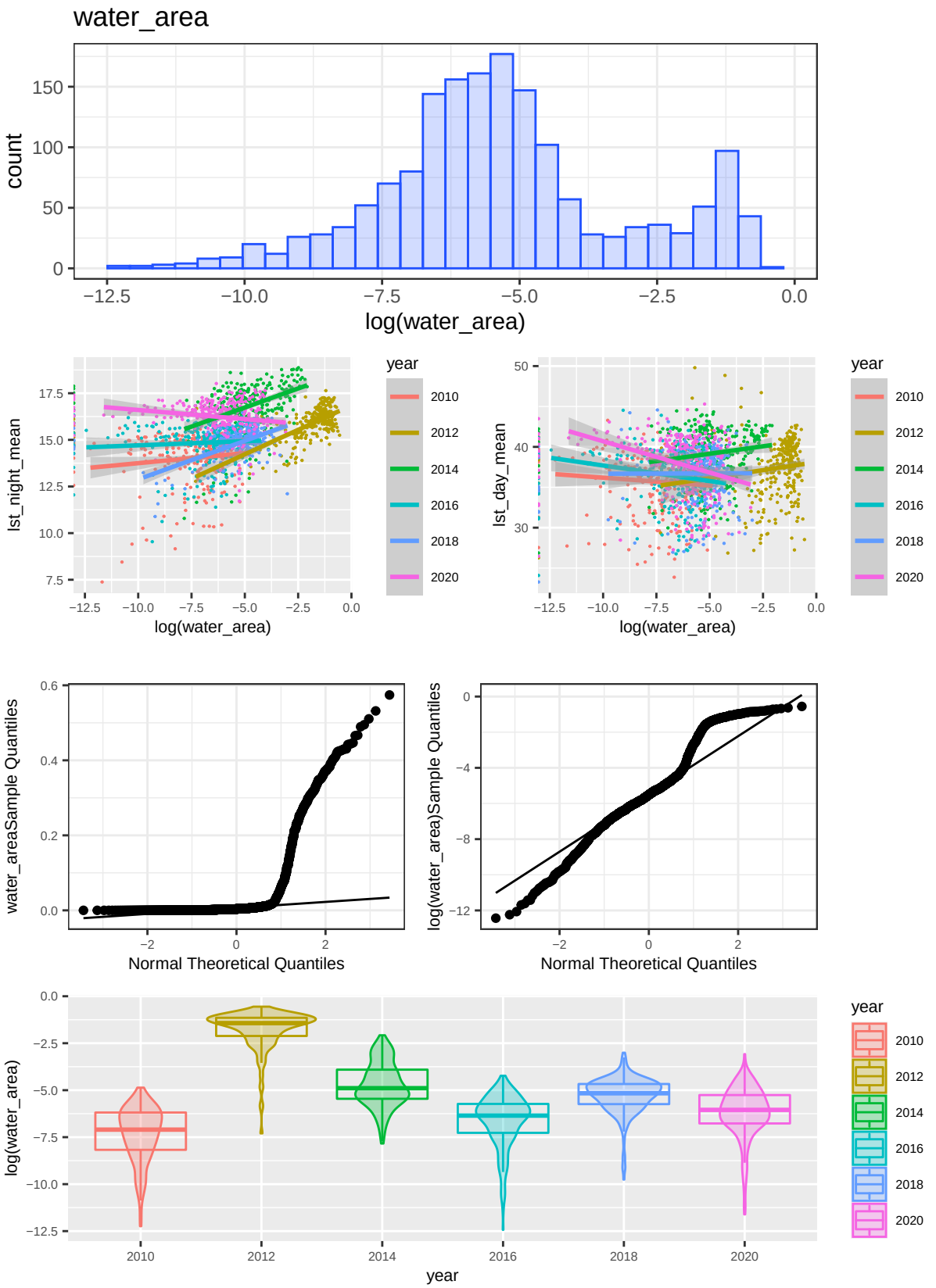






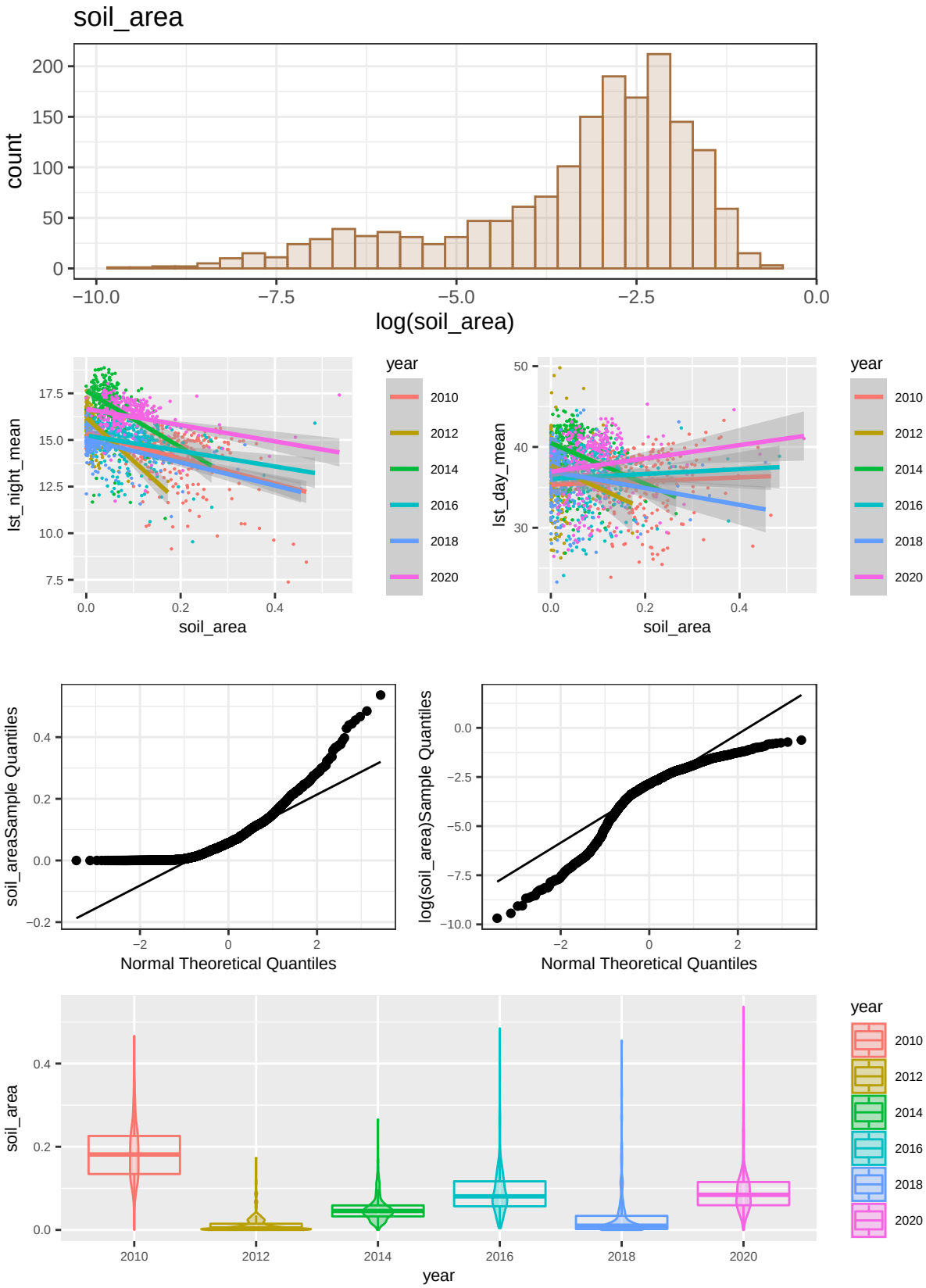
Google Earth: Random Forest Model

Water Area



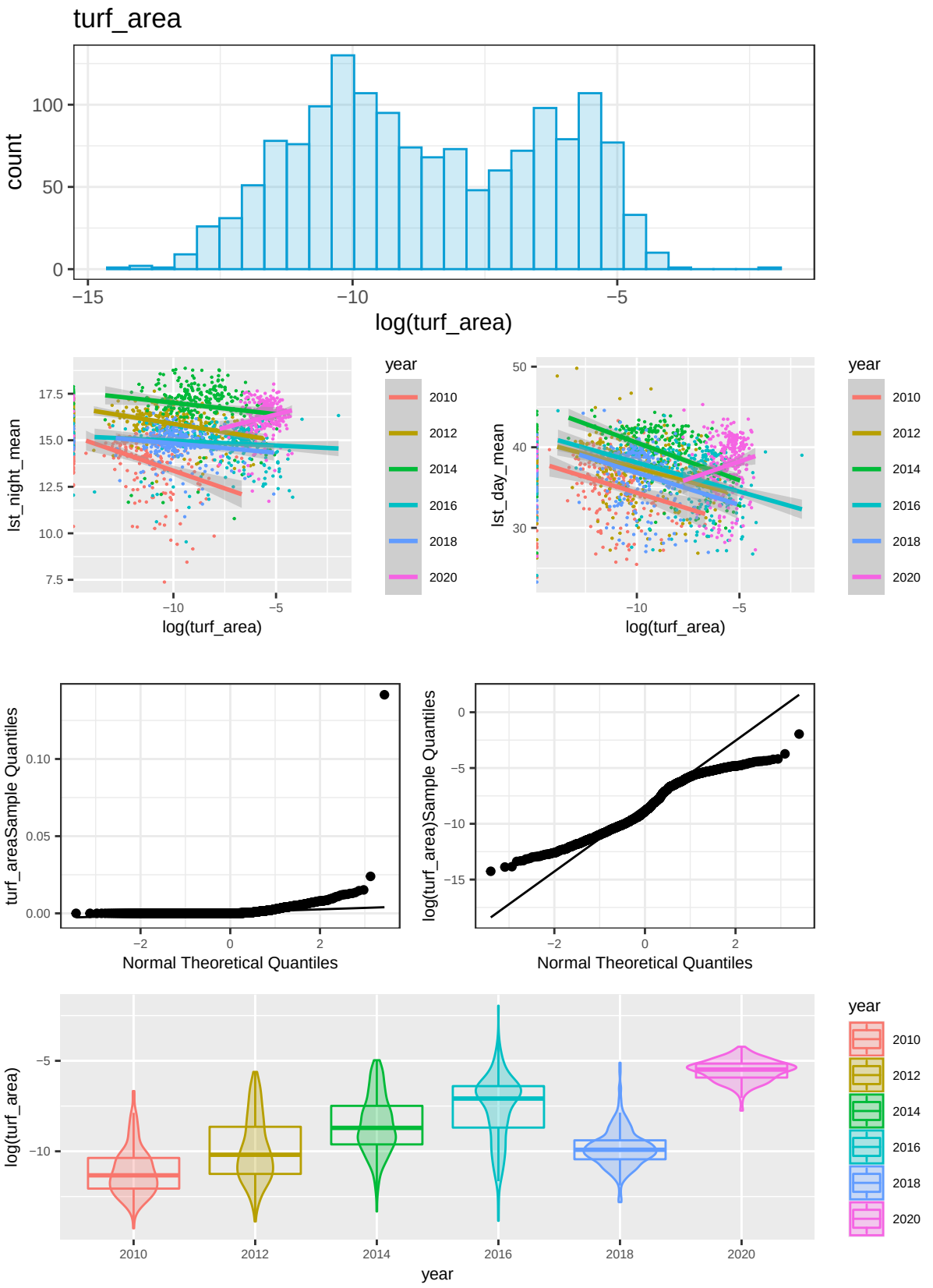
Google Earth: Random Forest Model

Soil Area



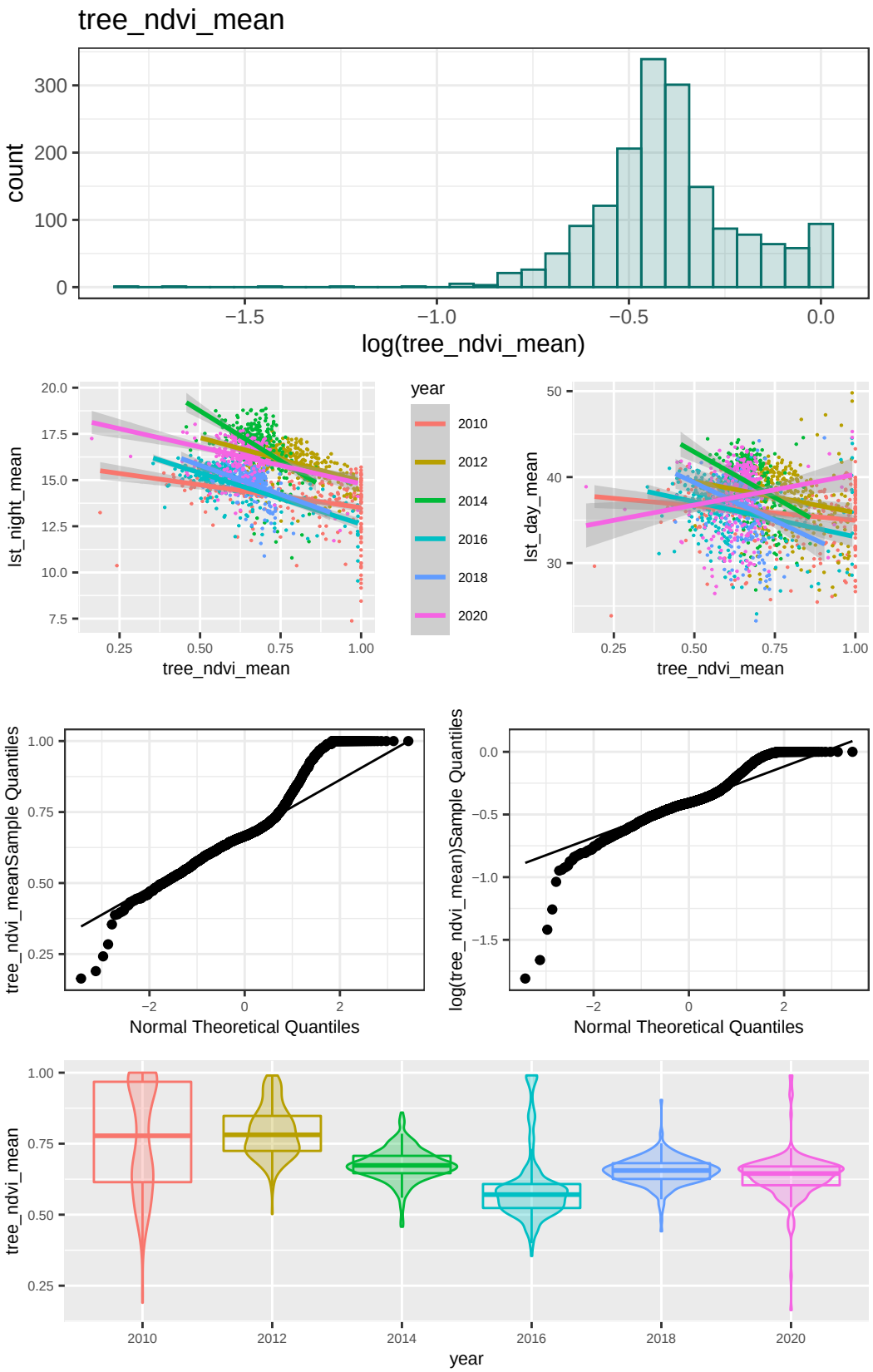
Google Earth: Random Forest Model

Turf Area



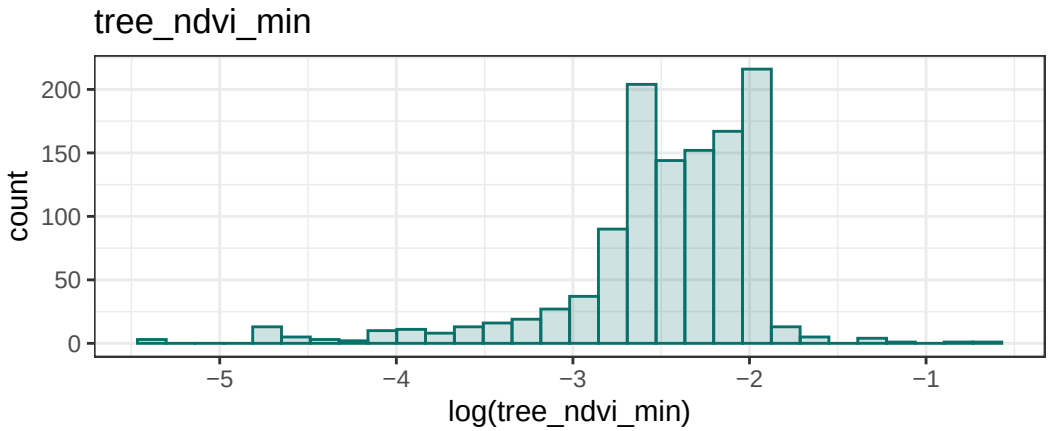
Google Earth: Random Forest Model

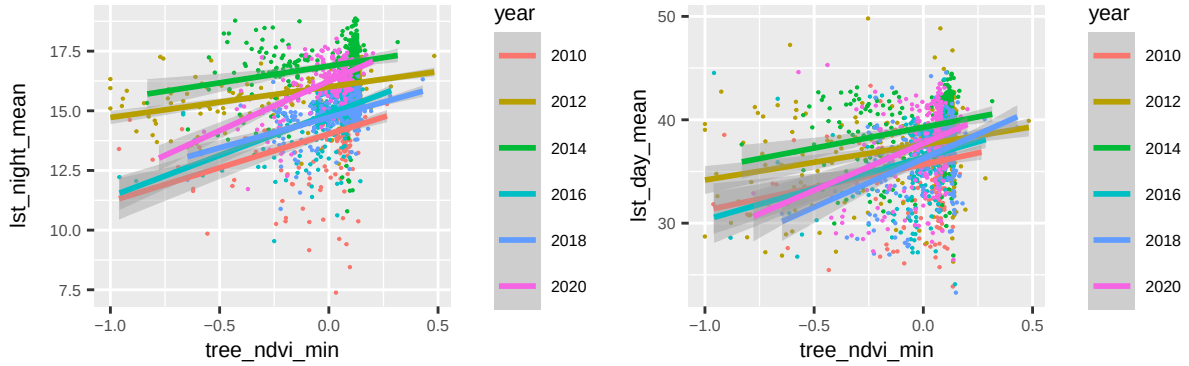
Tree NDVI Mean

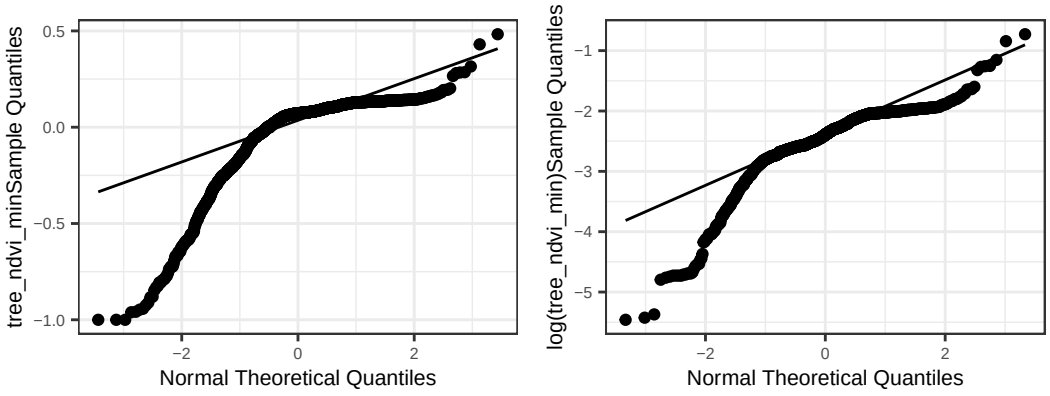


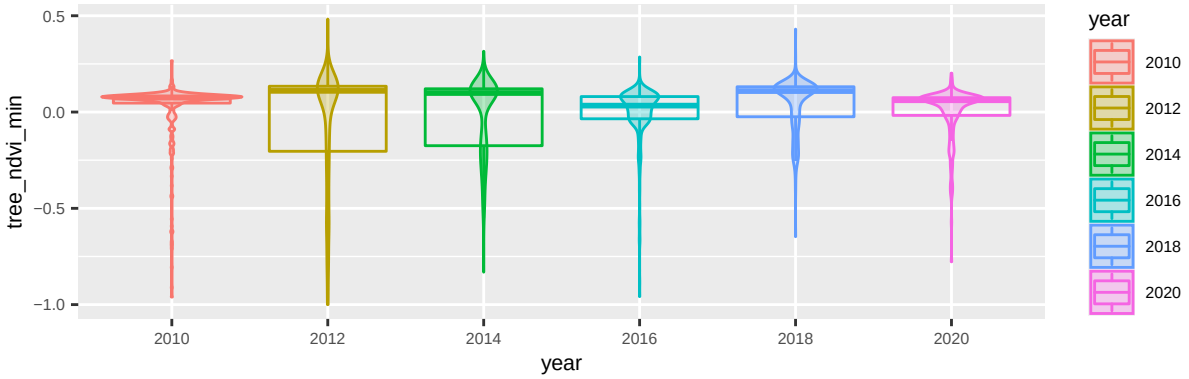
Google Earth: Random Forest Model

Tree NDVI Min



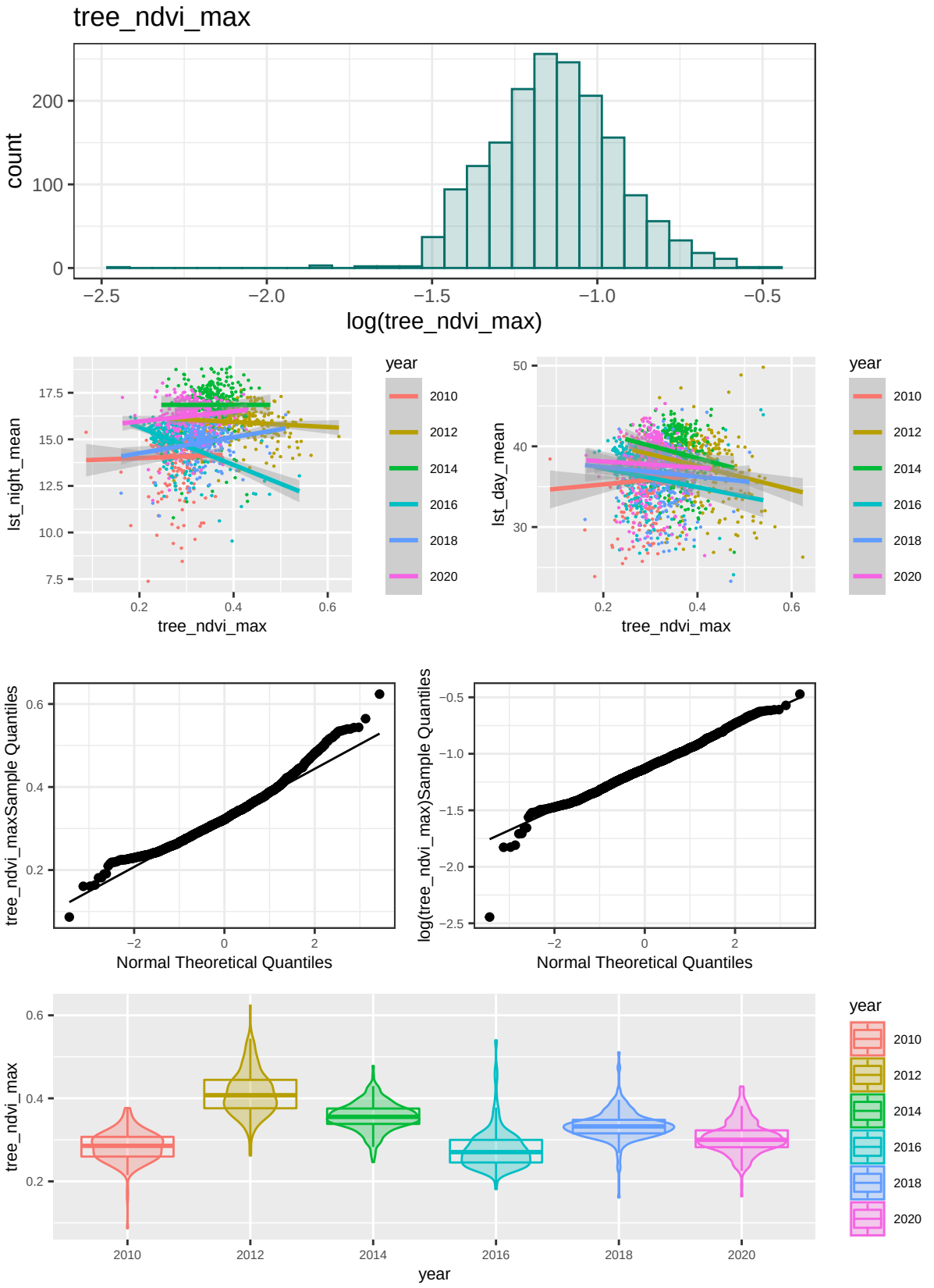






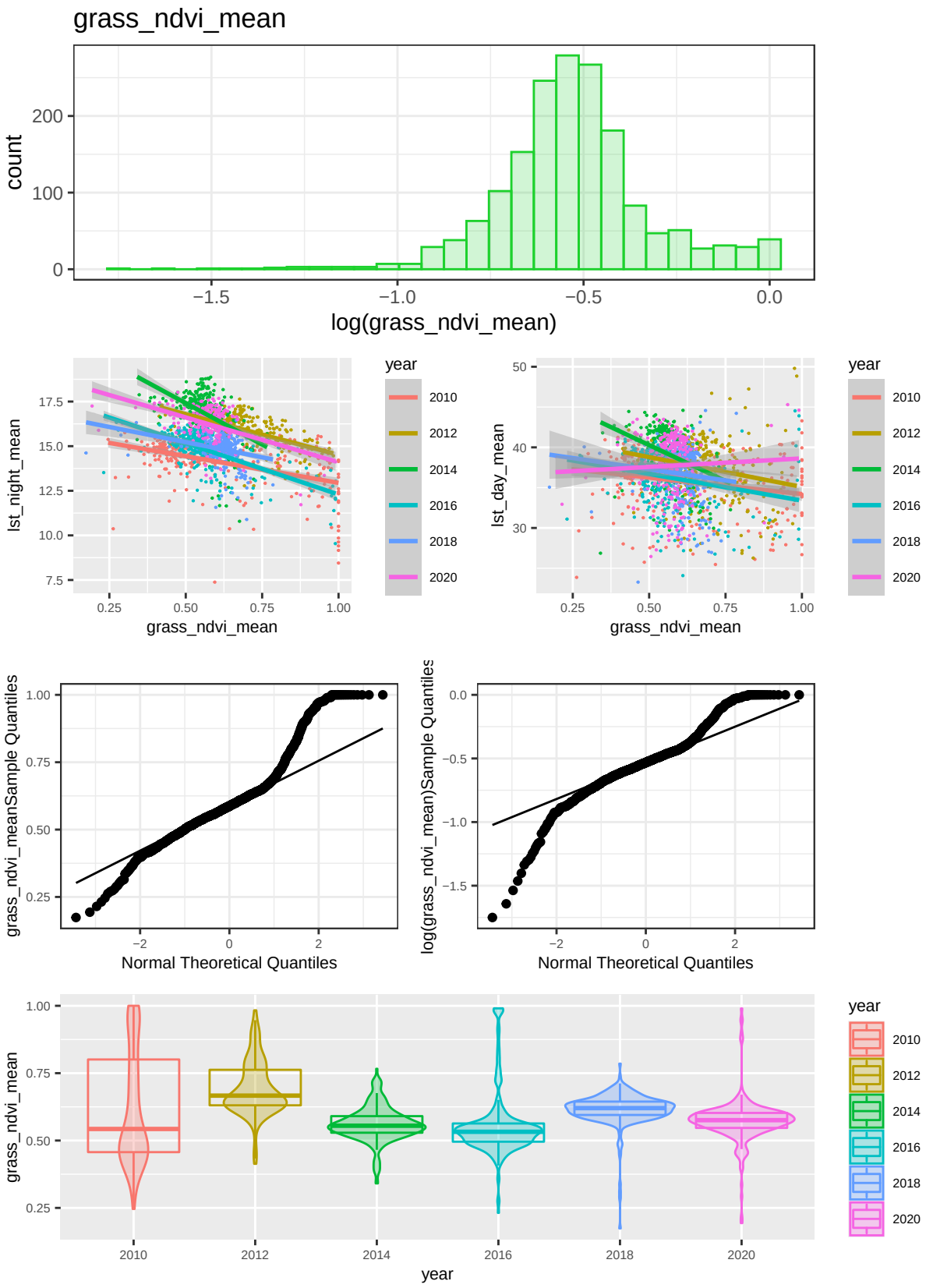
Google Earth: Random Forest Model

Tree NDVI Max



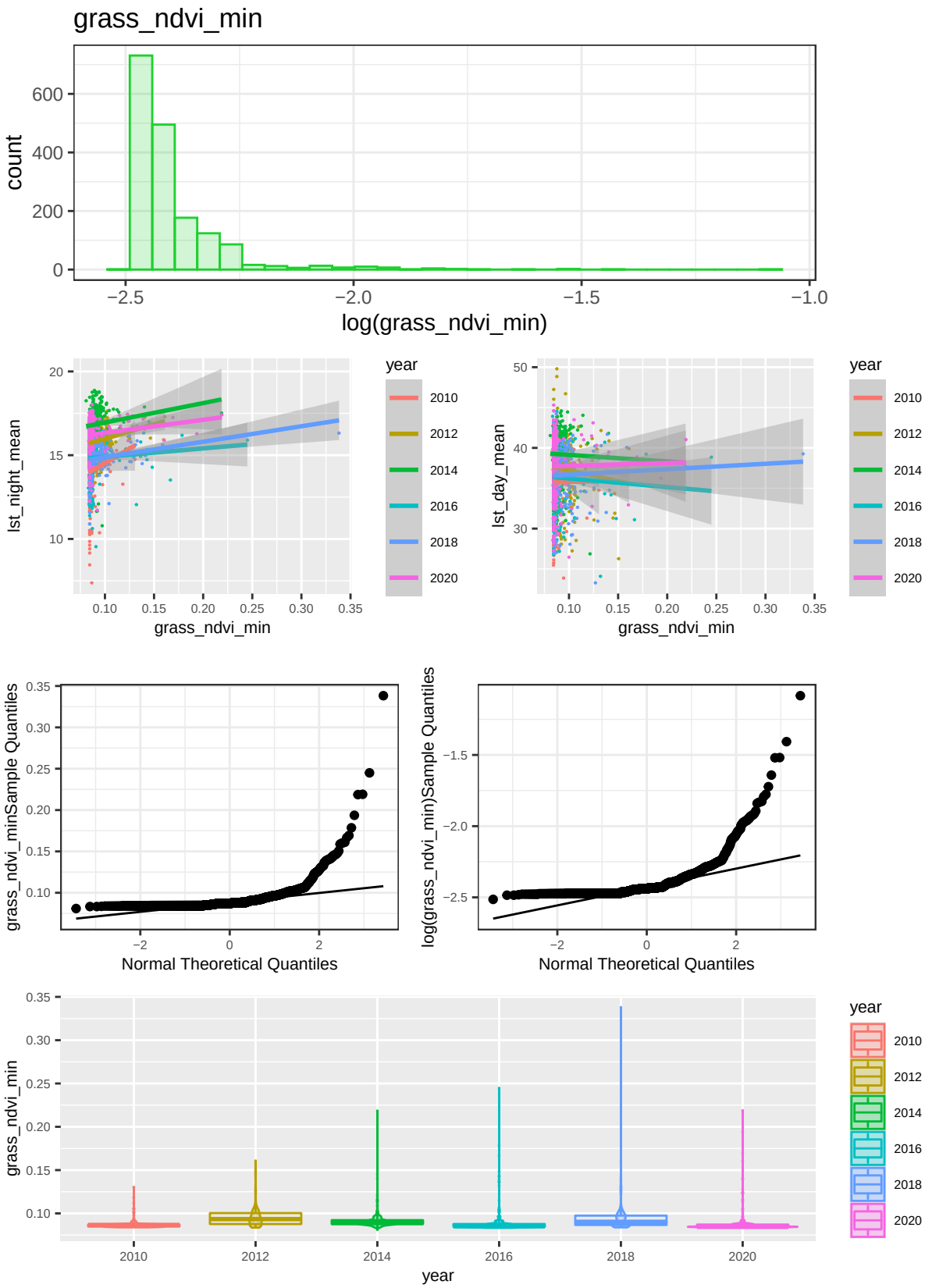
Google Earth: Random Forest Model

Lawn NDVI Mean



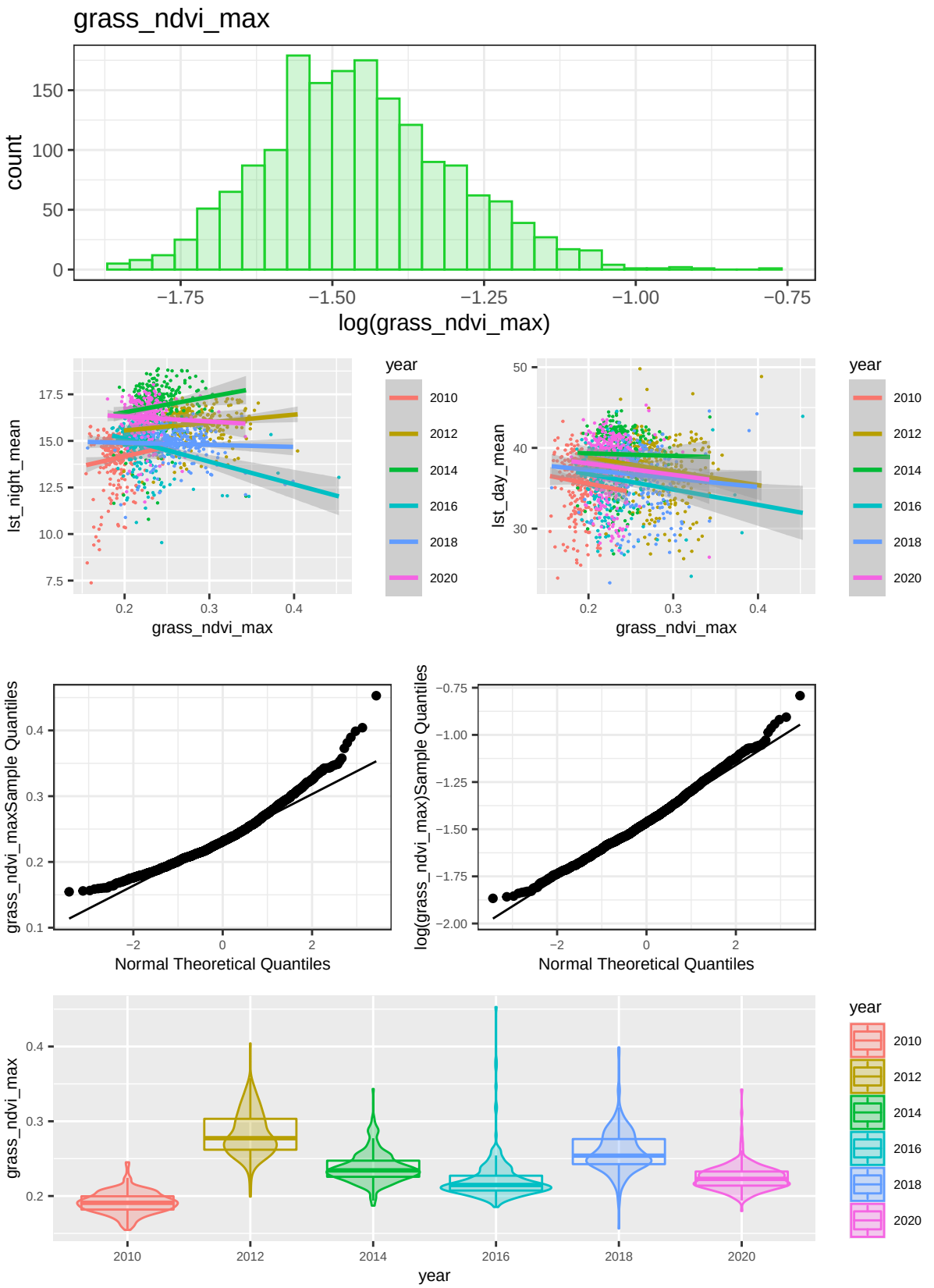
Google Earth: Random Forest Model

Lawn NDVI Min



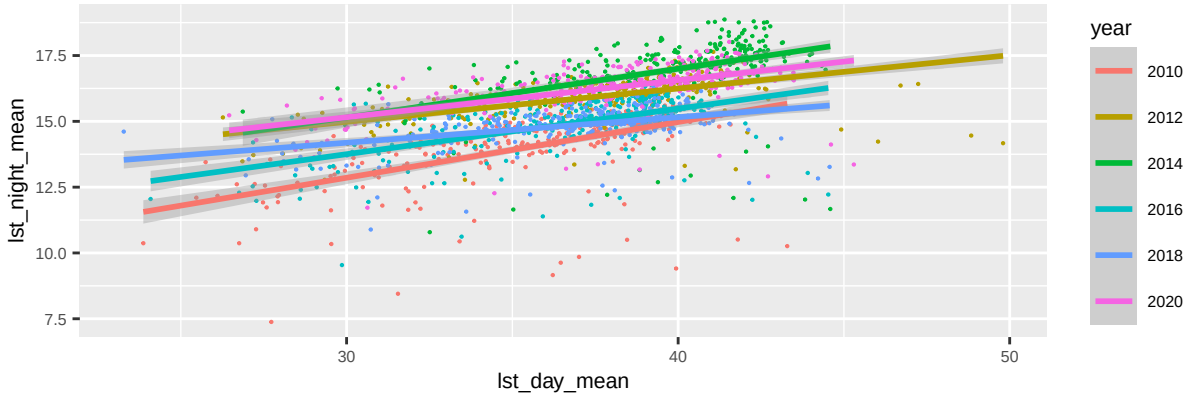
Google Earth: Random Forest Model

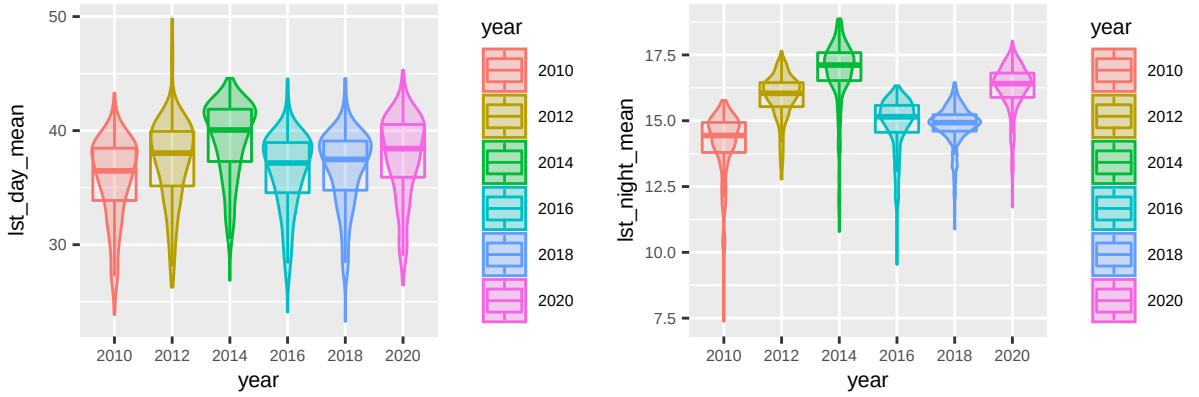
Lawn NDVI Max

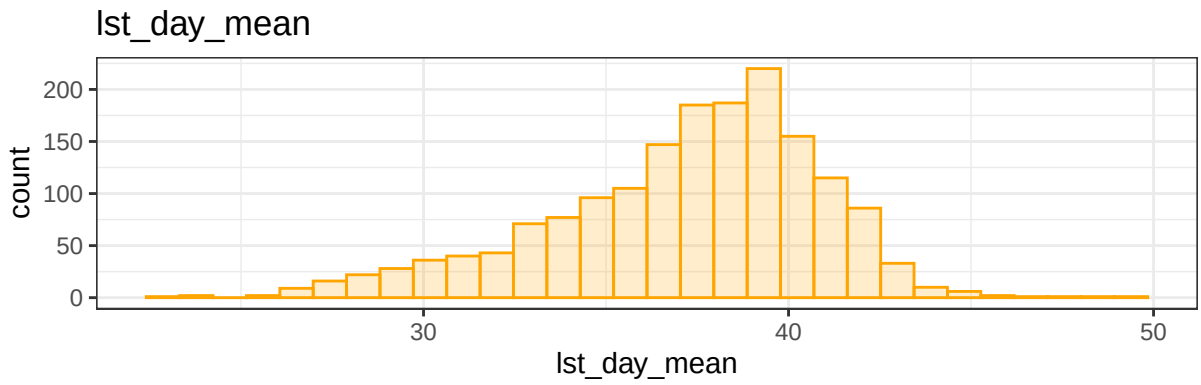


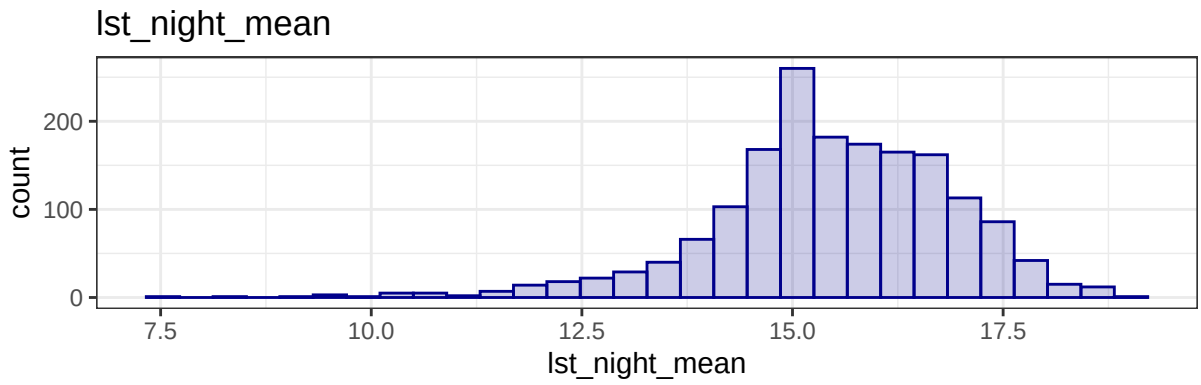
Google Earth: Random Forest Model

Mean Temperature









Google Earth: Random Forest Model

Create MicroClimate Panel Data Set

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

data$water_area <- data$water_area + 0.0000000001
data$soil_area <- data$soil_area + 0.0000000001
data$turf_area <- data$turf_area + 0.0000000001

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

First Attempt with Traditional OLS Model

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

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

stargazer::stargazer(OLSM1,OLSM2, single.row = TRUE,
                     title = 'OLS Model',
                     column.labels = c("OLSModel Day Temp",
                                       "OLSModel 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: Thu, Jul 28, 2022 - 8:35:03 PM

Second Attempt with a Mixed Effects Model

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

feM2 <- plm(
  lst_night_mean ~ 0 + factor(year) + tree_area + lawn_area + log(water_area) + log(soil_area)
    log(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',
                     column.labels = c("FixedEffects Day Temp",
```

Table 1: OLS Model

	Dependent variable:	
	lst_day_mean OLSModel Day Temp	lst_night_mean OLSModel Night Temp
	(1)	(2)
factor(year)2014	42.079^{***} (0.914)	20.770^{***} (0.267)
factor(year)2016	39.704^{***} (0.832)	18.945^{***} (0.244)
factor(year)2018	39.475^{***} (0.948)	19.024^{***} (0.277)
factor(year)2020	42.334^{***} (0.856)	20.501^{***} (0.251)
tree_area	-12.958^{***} (0.674)	-0.425^{**} (0.197)
lawn_area	0.958 (1.838)	2.048^{***} (0.538)
log(water_area)	-0.061^{*} (0.036)	0.079^{***} (0.011)
log(soil_area)	0.082^{*} (0.048)	-0.025^{*} (0.014)
log(turf_area)	-0.083^{**} (0.034)	-0.014 (0.010)
grass_ndvi_mean	-4.152^{*} (2.313)	-5.811^{***} (0.677)
tree_ndvi_mean	1.280 (2.275)	-0.913 (0.665)
Observations	1,132	1,132
R^2	0.994	0.997
Adjusted R^2	0.994	0.997
Residual Std. Error (df = 1121)	2.929	0.857
F Statistic (df = 11; 1121)	$16{,}906.410^{***}$	$34{,}689.580^{***}$
Note:		*p<0.1; **p<0.05; ***p<0.01

```
                                        "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: Thu, Jul 28, 2022 - 8:35:04 PM

Table 2: Fixed Effects Model

	Dependent variable:	
	lst_day_mean FixedEffects Day Temp	lst_night_mean FixedEffects Night Temp
	(1)	(2)
factor(year)2014	1.343*** (0.080)	0.585*** (0.051)
factor(year)2016	−1.567*** (0.071)	−1.391*** (0.046)
factor(year)2018	−1.418*** (0.089)	−1.330*** (0.057)
tree_area	−2.155*** (0.468)	0.017 (0.302)
lawn_area	−2.256*** (0.737)	1.027** (0.474)
log(water_area)	−0.075*** (0.019)	0.006 (0.012)
log(soil_area)	−0.017 (0.012)	0.036*** (0.008)
log(turf_area)	−0.017* (0.010)	−0.002 (0.006)
grass_ndvi_mean	0.985* (0.596)	−0.662* (0.384)
tree_ndvi_mean	1.198** (0.538)	0.095 (0.346)
Observations	1,132	1,132
R^2	0.833	0.879
Adjusted R^2	0.775	0.836
F Statistic (df = 10; 837)	418.404***	605.811***
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:  lst_day_mean ~ 0 + factor(year) + tree_area + lawn_area + log(water_area) +  ...
## F = 93.993, df1 = 284, 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(
  lst_day_mean ~ 0 + tree_area + lawn_area + log(water_area) + log(soil_area) +
    log(turf_area) + grass_ndvi_mean + tree_ndvi_mean,
  index = c(zipcode, year), data = microClimatePanel, model = 'random')

reM2 <- plm(
  lst_night_mean ~ 0 + tree_area + lawn_area + log(water_area) + log(soil_area) +
    log(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: Thu, Jul 28, 2022 - 8:35:06 PM

Table 3: Random Effects Model

	Dependent variable:	
	lst_day_mean RandomEffects Day Temp	lst_night_mean RandomEffects Night Temp
	(1)	(2)
tree_area	7.058*** (1.441)	5.744*** (0.708)
lawn_area	21.620*** (2.668)	15.092*** (1.348)
log(water_area)	−0.718*** (0.060)	−0.258*** (0.029)
log(soil_area)	−0.040 (0.049)	0.031 (0.026)
log(turf_area)	−0.165*** (0.035)	−0.034* (0.018)
grass_ndvi_mean	−3.796* (1.960)	−4.919*** (1.022)
tree_ndvi_mean	33.604*** (1.705)	19.198*** (0.896)
Observations	1,132	1,132
R^2	0.058	0.094
Adjusted R^2	0.053	0.090
F Statistic	3,365.535***	6,607.022***
Note:		*p<0.1; **p<0.05; ***p<0.01

Compare Performance of Mixed Effects vs Random Effects Models

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

```
##
##  Hausman Test
##
## data:  lst_day_mean ~ 0 + factor(year) + tree_area + lawn_area + log(water_area) +  ...
## chisq = 4238, 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).
```

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(
  lst_day_mean ~ 0 + tree_area + lawn_area + log(water_area) + log(soil_area) +
    log(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: Thu, Jul 28, 2022 - 8:35:07 PM

Table 4:

	Dependent variable:
	lst_day_mean
tree_area	2.565*** (0.631)
lawn_area	8.581*** (1.154)
log(water_area)	0.303*** (0.030)
log(soil_area)	0.036* (0.020)
log(turf_area)	0.065*** (0.015)
grass_ndvi_mean	−13.743*** (0.825)
tree_ndvi_mean	14.032*** (0.750)
Observations	1,132
R^2	0.478
Adjusted R^2	0.297
F Statistic	109.708*** (df = 7; 840)
Note:	*p<0.1; **p<0.05; ***p<0.01

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

```
##
##  F test for individual effects
##
## data:  lst_day_mean ~ 0 + factor(year) + tree_area + lawn_area + log(water_area) +  ...
## F = 595.33, df1 = 3, df2 = 837, p-value < 2.2e-16
## alternative hypothesis: significant effects
```

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

```
##
##  Lagrange Multiplier Test - time effects (Breusch-Pagan) for unbalanced
##  panels
##
## data:  lst_day_mean ~ 0 + factor(year) + tree_area + lawn_area + log(water_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:  lst_day_mean ~ 0 + factor(year) + tree_area + lawn_area + log(water_area) +  ...
## chisq = 49.753, df = 1, p-value = 1.744e-12
## alternative hypothesis: serial correlation in idiosyncratic errors
```

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

```
##
## t test of coefficients:
##
##                   Estimate Std. Error  t value  Pr(>|t|)
## factor(year)2014  1.342987   0.077123  17.4136 < 2.2e-16 ***
## factor(year)2016 -1.566925   0.069203 -22.6423 < 2.2e-16 ***
## factor(year)2018 -1.417975   0.094855 -14.9489 < 2.2e-16 ***
## tree_area        -2.155056   0.512658  -4.2037 2.909e-05 ***
## lawn_area        -2.256392   0.906433  -2.4893   0.01299 *
## log(water_area)  -0.074947   0.017742  -4.2243 2.660e-05 ***
## log(soil_area)   -0.016669   0.010594  -1.5734   0.11601
## log(turf_area)   -0.017084   0.009136  -1.8699   0.06184 .
## grass_ndvi_mean   0.984963   0.580048   1.6981   0.08987 .
## tree_ndvi_mean    1.198089   0.563376   2.1266   0.03374 *
## ---
## 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  1.342987   0.077123  17.4136 < 2.2e-16 ***
## factor(year)2016 -1.566925   0.069203 -22.6423 < 2.2e-16 ***
## factor(year)2018 -1.417975   0.094855 -14.9489 < 2.2e-16 ***
## tree_area        -2.155056   0.512658  -4.2037 2.909e-05 ***
## lawn_area        -2.256392   0.906433  -2.4893   0.01299 *
## log(water_area)  -0.074947   0.017742  -4.2243 2.660e-05 ***
## log(soil_area)   -0.016669   0.010594  -1.5734   0.11601
## log(turf_area)   -0.017084   0.009136  -1.8699   0.06184 .
## grass_ndvi_mean   0.984963   0.580048   1.6981   0.08987 .
## tree_ndvi_mean    1.198089   0.563376   2.1266   0.03374 *
## ---
## 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, method = "ar
```

```
##     factor(year)2014 factor(year)2016 factor(year)2018 tree_area lawn_area
## HC0       0.07712266       0.06920345       0.09485494 0.5126580 0.9064332
## HC1       0.07746558       0.06951115       0.09527671 0.5149375 0.9104636
## HC2       0.07788296       0.06984899       0.09602538 0.5206905 0.9213472
## HC3       0.07867625       0.07051884       0.09724551 0.5290508 0.9369258
## HC4       0.07958025       0.07123518       0.09896360 0.5424832 0.9626473
##     log(water_area) log(soil_area) log(turf_area) grass_ndvi_mean
## HC0      0.01774208     0.01059424    0.009135963       0.5800483
## HC1      0.01782097     0.01064135    0.009176585       0.5826274
## HC2      0.01813795     0.01088292    0.009285425       0.5861647
## HC3      0.01856396     0.01118719    0.009440793       0.5924203
## HC4      0.01938445     0.01180339    0.009694735       0.5992896
##     tree_ndvi_mean
## HC0      0.5633764
## HC1      0.5658814
## HC2      0.5690925
## HC3      0.5749515
## HC4      0.5813872
```

```
totalRobust <- coeftest(feM1, vcov = vcovHC(feM1, type = 'HC0'))
cInterval <- coefci(feM1, vcov. = vcovHC(feM1, type = 'HC0'))
```

```
print(totalRobust)
```

```
##
## t test of coefficients:
##
##                   Estimate Std. Error  t value  Pr(>|t|)
## factor(year)2014  1.342987   0.077123  17.4136 < 2.2e-16 ***
## factor(year)2016 -1.566925   0.069203 -22.6423 < 2.2e-16 ***
## factor(year)2018 -1.417975   0.094855 -14.9489 < 2.2e-16 ***
```

```
## tree_area        -2.155056   0.512658  -4.2037 2.909e-05 ***
## lawn_area        -2.256392   0.906433  -2.4893   0.01299 *
## log(water_area)  -0.074947   0.017742  -4.2243 2.660e-05 ***
## log(soil_area)   -0.016669   0.010594  -1.5734   0.11601
## log(turf_area)   -0.017084   0.009136  -1.8699   0.06184 .
## grass_ndvi_mean   0.984963   0.580048   1.6981   0.08987 .
## tree_ndvi_mean    1.198089   0.563376   2.1266   0.03374 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
print(cInterval)
```

```
##                          2.5 %        97.5 %
## factor(year)2014  1.19161020  1.4943632484
## factor(year)2016 -1.70275736 -1.4310920040
## factor(year)2018 -1.60415610 -1.2317930992
## tree_area        -3.16130259 -1.1488101196
## lawn_area        -4.03554081 -0.4772424380
## log(water_area)  -0.10977146 -0.0401230837
## log(soil_area)   -0.03746331  0.0041254944
## log(turf_area)   -0.03501584  0.0008483342
## grass_ndvi_mean  -0.15355743  2.1234827556
## tree_ndvi_mean    0.09229206  2.3038850927
```

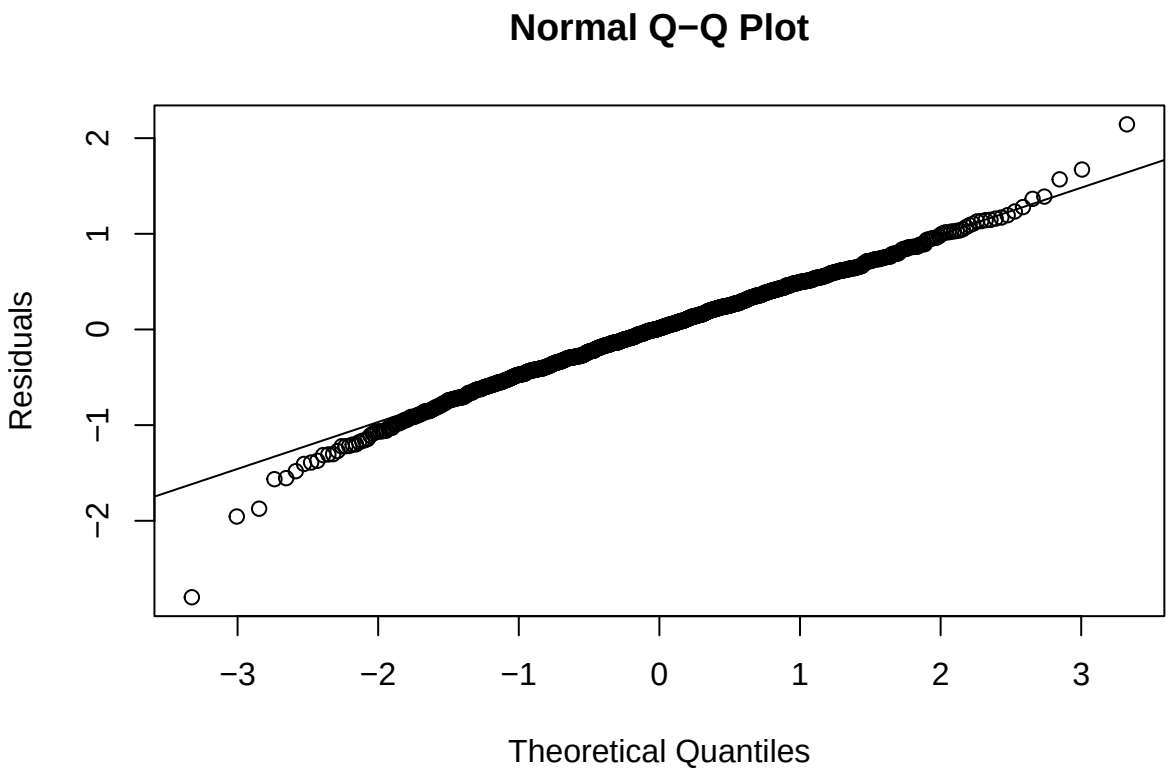
```
stargazer::stargazer(OLSM1,feM1,reM1,feM2FixedTime,
                     font.size = 'tiny',
                     title = 'Regression Models Sumamry',
                     column.labels = c("OLS", "FixedEffects",
                                       "RandomEffects",
                                       "FixedEffectsFixedTime"))
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Thu, Jul 28, 2022 - 8:35:09 PM

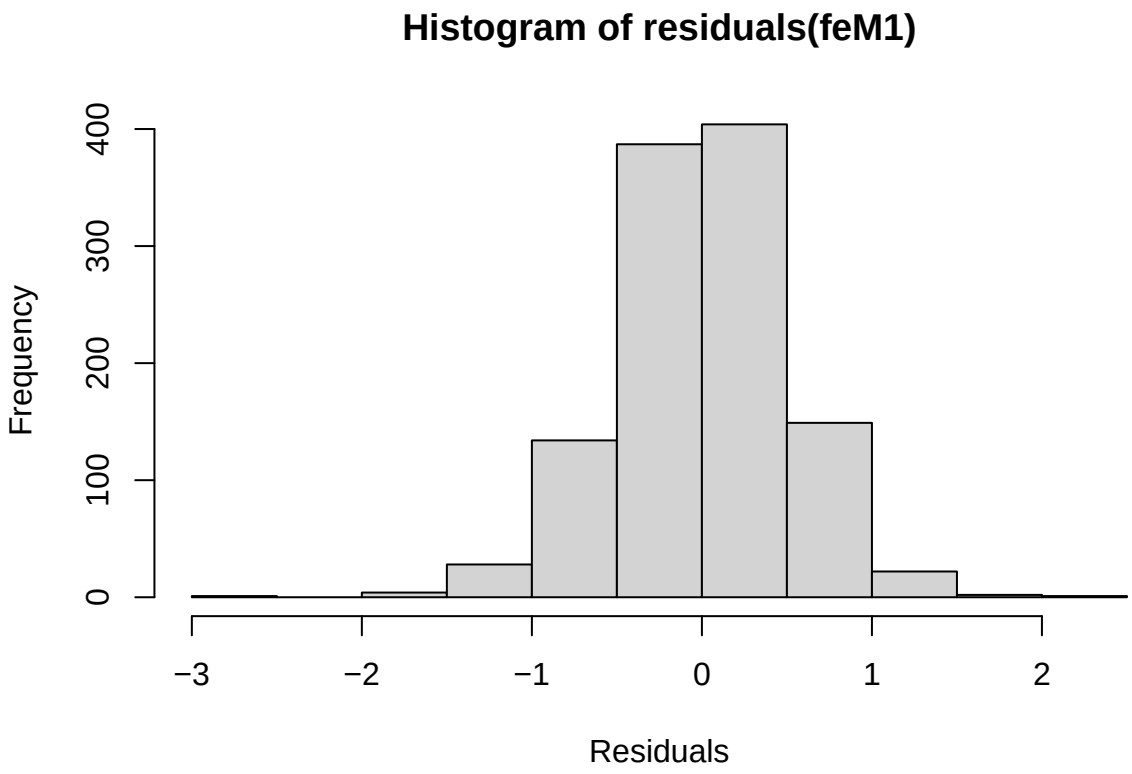
```
qqnorm(residuals(feM1), ylab = 'Residuals')
qqline(residuals(feM1))
```

Table 5: Regression Models Sumamry

	Dependent variable:			
	lst_day_mean			
	OLS		*panel linear*	
	OLS	FixedEffects	RandomEffects	FixedEffectsFixedTime
	(1)	(2)	(3)	(4)
factor(year)2014	42.079***	1.343***		
	(0.914)	(0.080)		
factor(year)2016	39.704***	−1.567***		
	(0.832)	(0.071)		
factor(year)2018	39.475***	−1.418***		
	(0.948)	(0.089)		
factor(year)2020	42.334***			
	(0.856)			
tree_area	−12.958***	−2.155***	7.058***	2.565***
	(0.674)	(0.468)	(1.441)	(0.631)
lawn_area	0.958	−2.256***	21.620***	8.581***
	(1.838)	(0.737)	(2.668)	(1.154)
log(water_area)	−0.061*	−0.075***	−0.718***	0.303***
	(0.036)	(0.019)	(0.060)	(0.030)
log(soil_area)	0.082*	−0.017	−0.040	0.036*
	(0.048)	(0.012)	(0.049)	(0.020)
log(turf_area)	−0.083**	−0.017*	−0.165***	0.065***
	(0.034)	(0.010)	(0.035)	(0.015)
grass_ndvi_mean	−4.152*	0.985*	−3.796*	−13.743***
	(2.313)	(0.596)	(1.960)	(0.825)
tree_ndvi_mean	1.280	1.198**	33.604***	14.032***
	(2.275)	(0.538)	(1.705)	(0.750)
Observations	1,132	1,132	1,132	1,132
R^2	0.994	0.833	0.058	0.478
Adjusted R^2	0.994	0.775	0.053	0.297
Residual Std. Error	2.929 (df = 1121)			
F Statistic	16,906.410*** (df = 11; 1121)	418.404*** (df = 10; 837)	3,365.535***	109.708*** (df = 7; 840)
Note:				*p<0.1; **p<0.05; ***p<0.01




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


Histogram of residuals(feM1)
Frequency
0
100
200
300
400
−3
−2
−1
0
1
2
Residuals