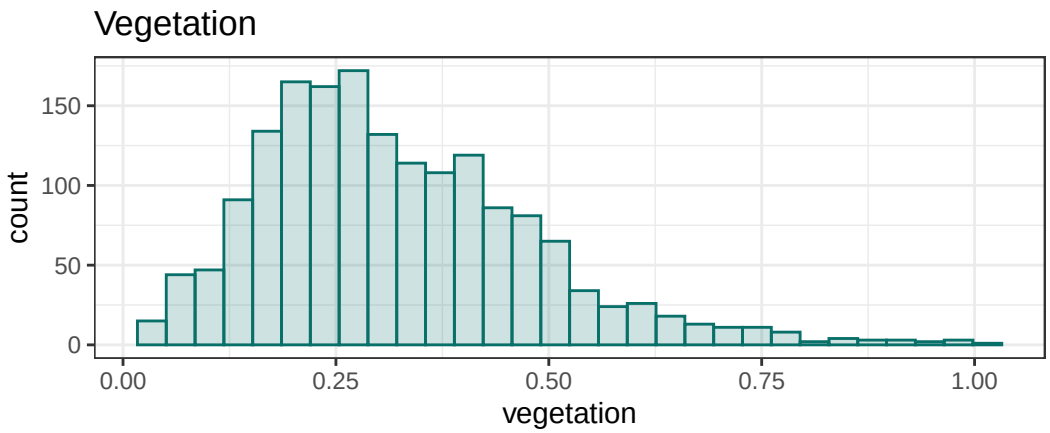
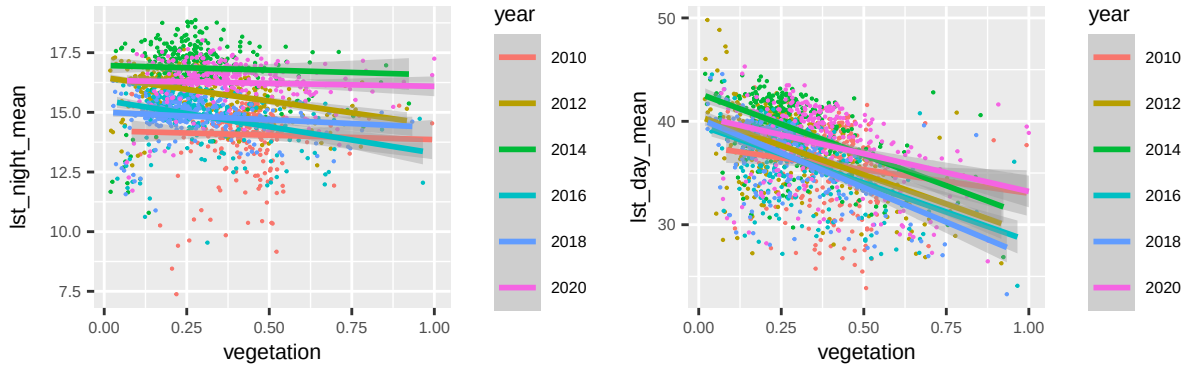
LawnProject Panel Data Analysis / Random Forest Data

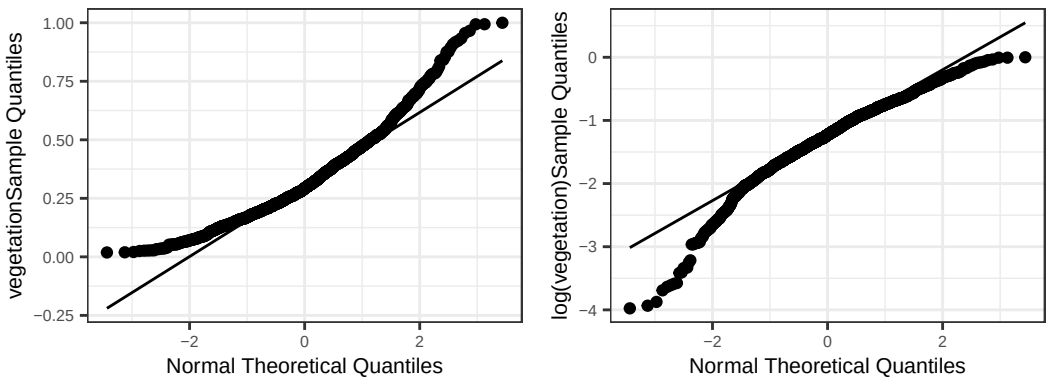
LawnProject Team

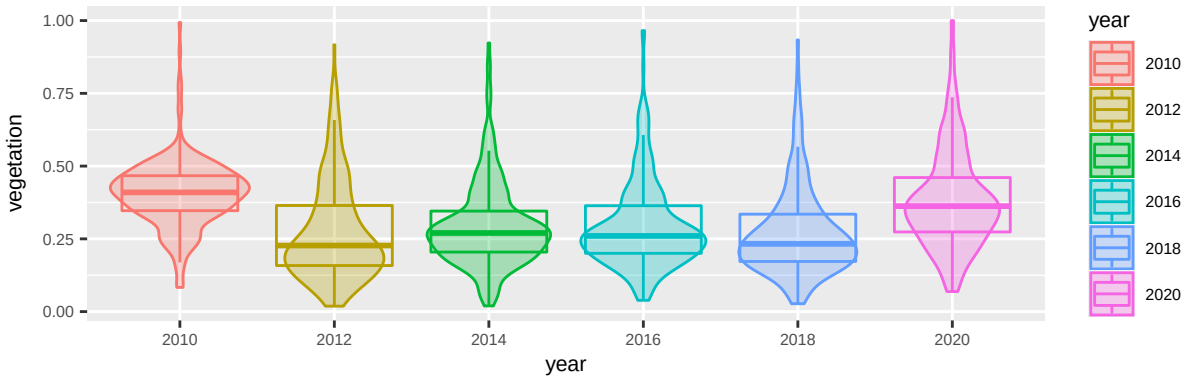
Intro

Aggregate Vegetation Area (Trees + Lawn)



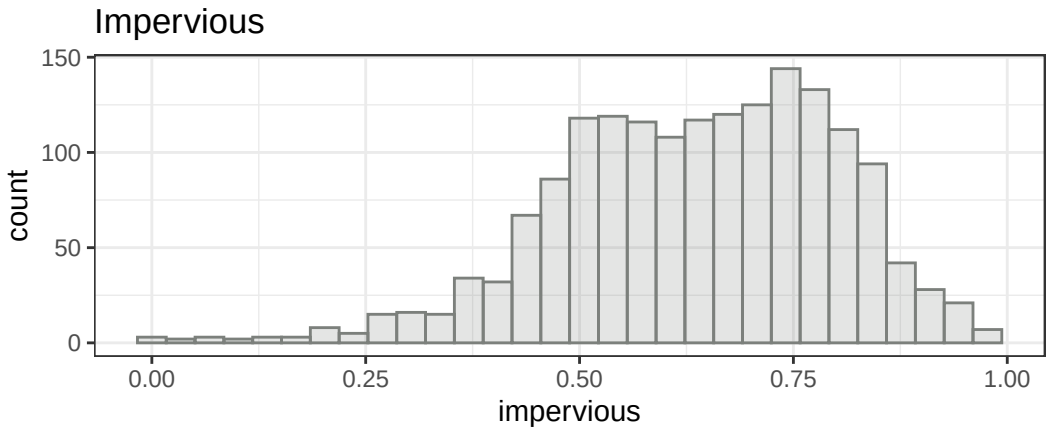


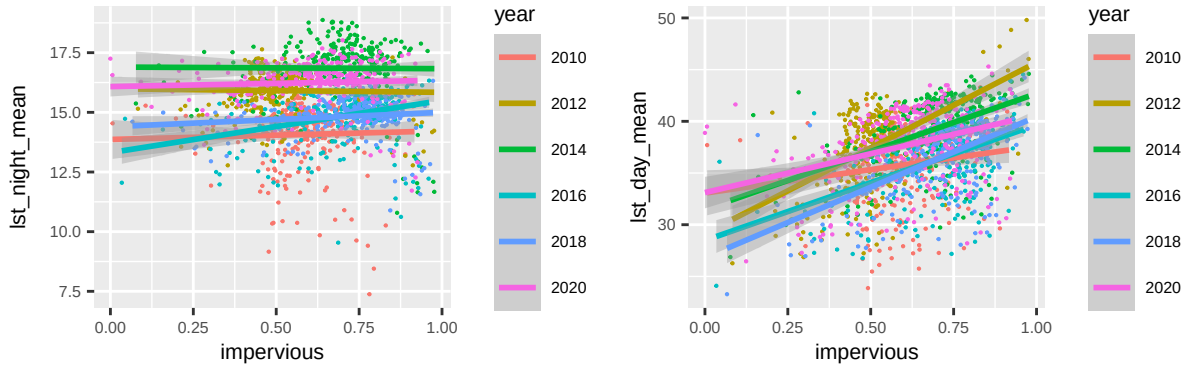


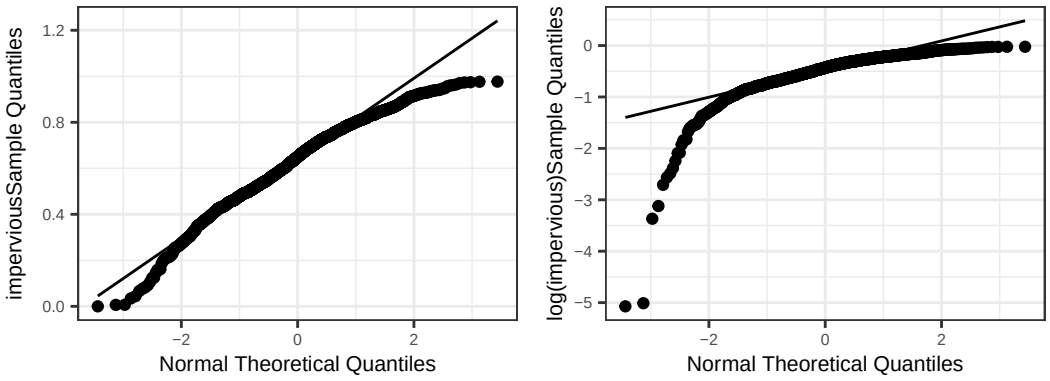


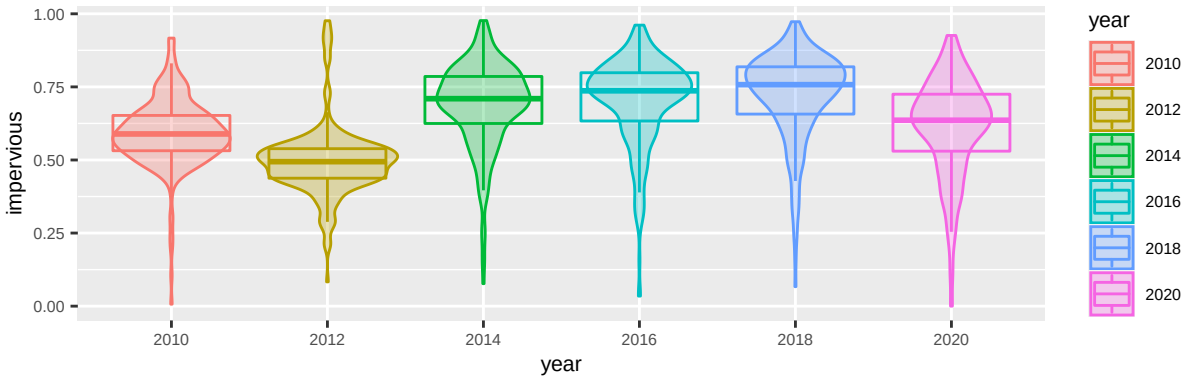
Google Earth: Random Forest Model

Aggregate Impervious Area (Soil + Turf + Impervious)



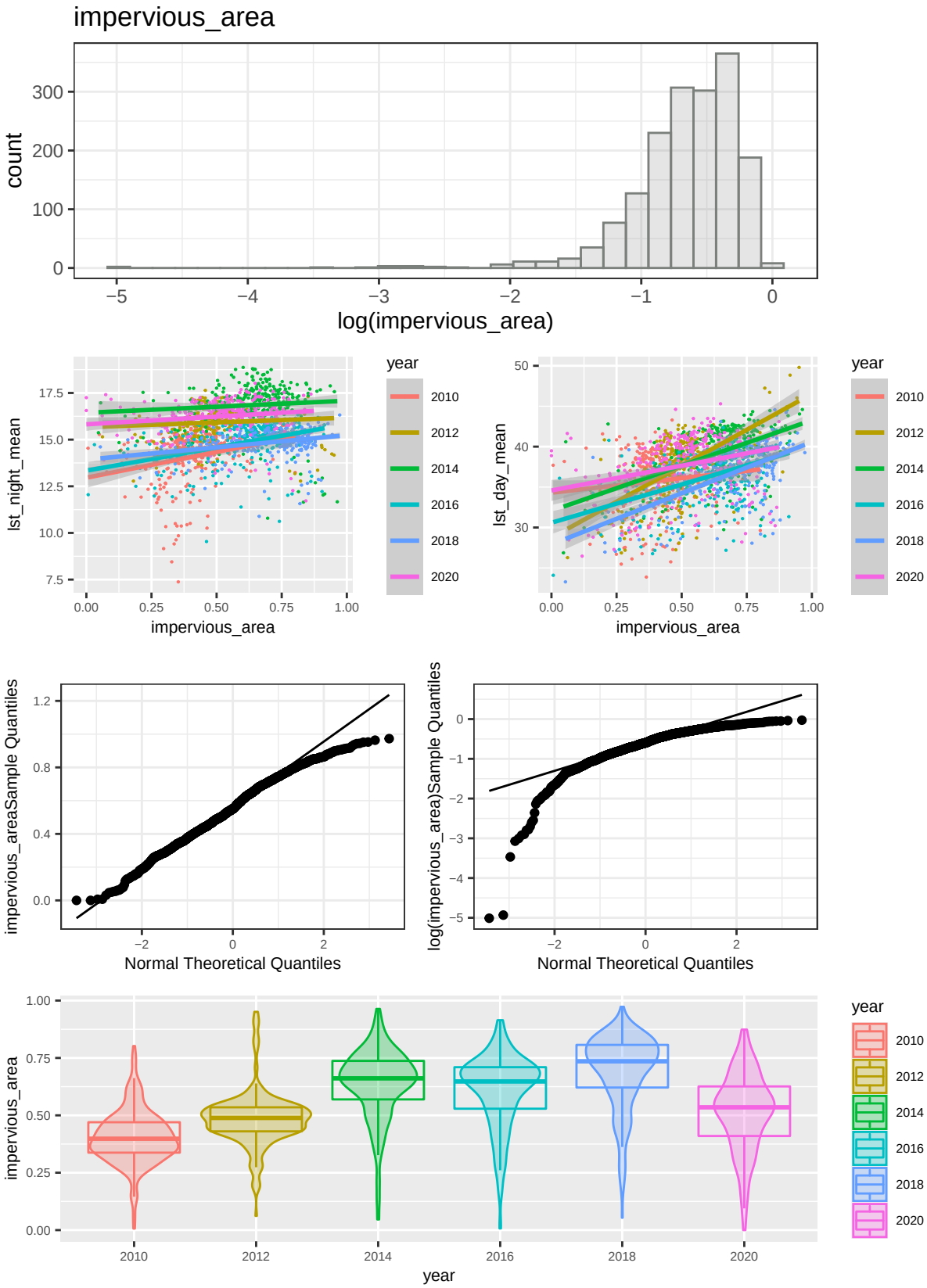






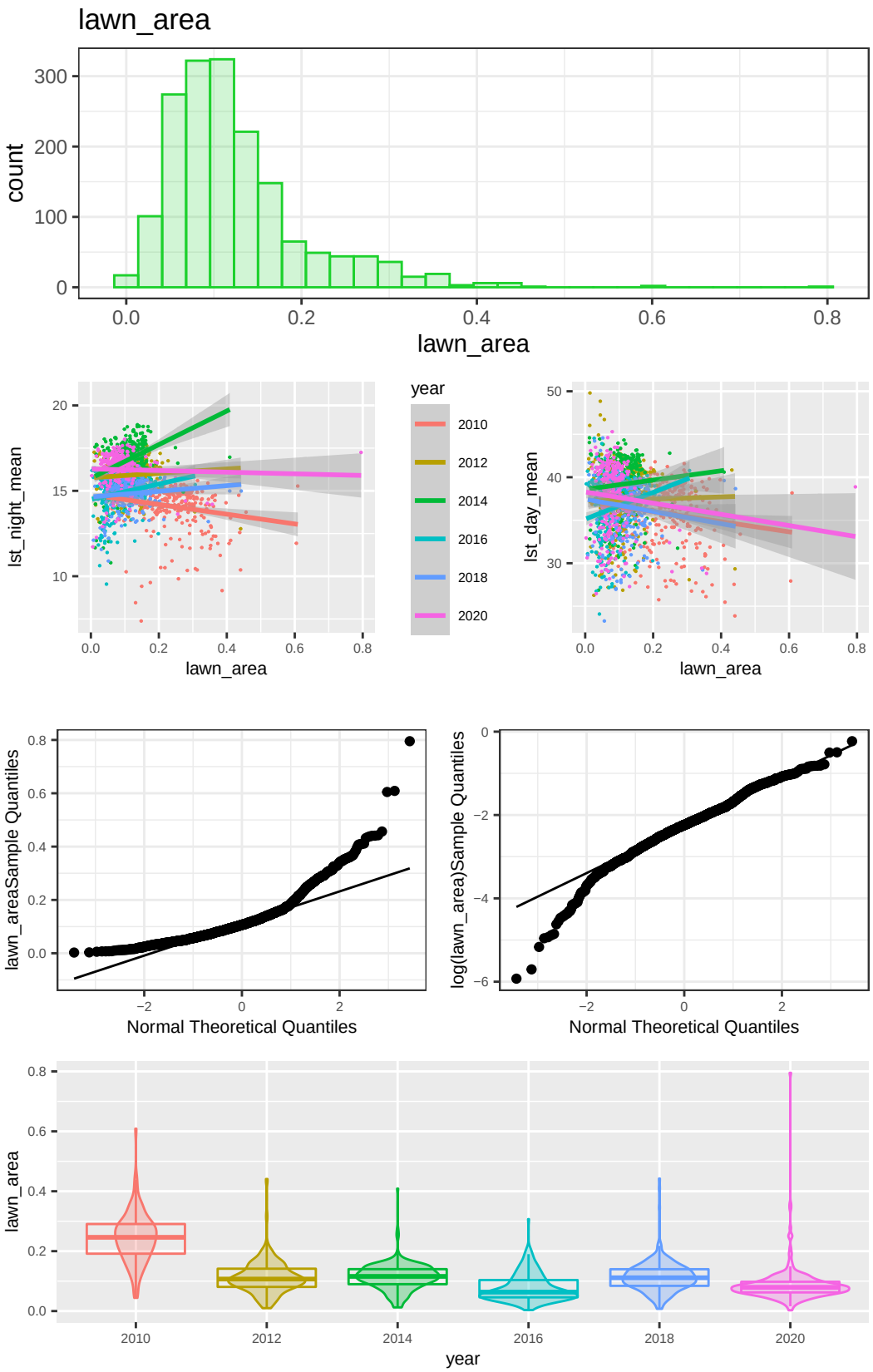
Google Earth: Random Forest Model

Impervious Area



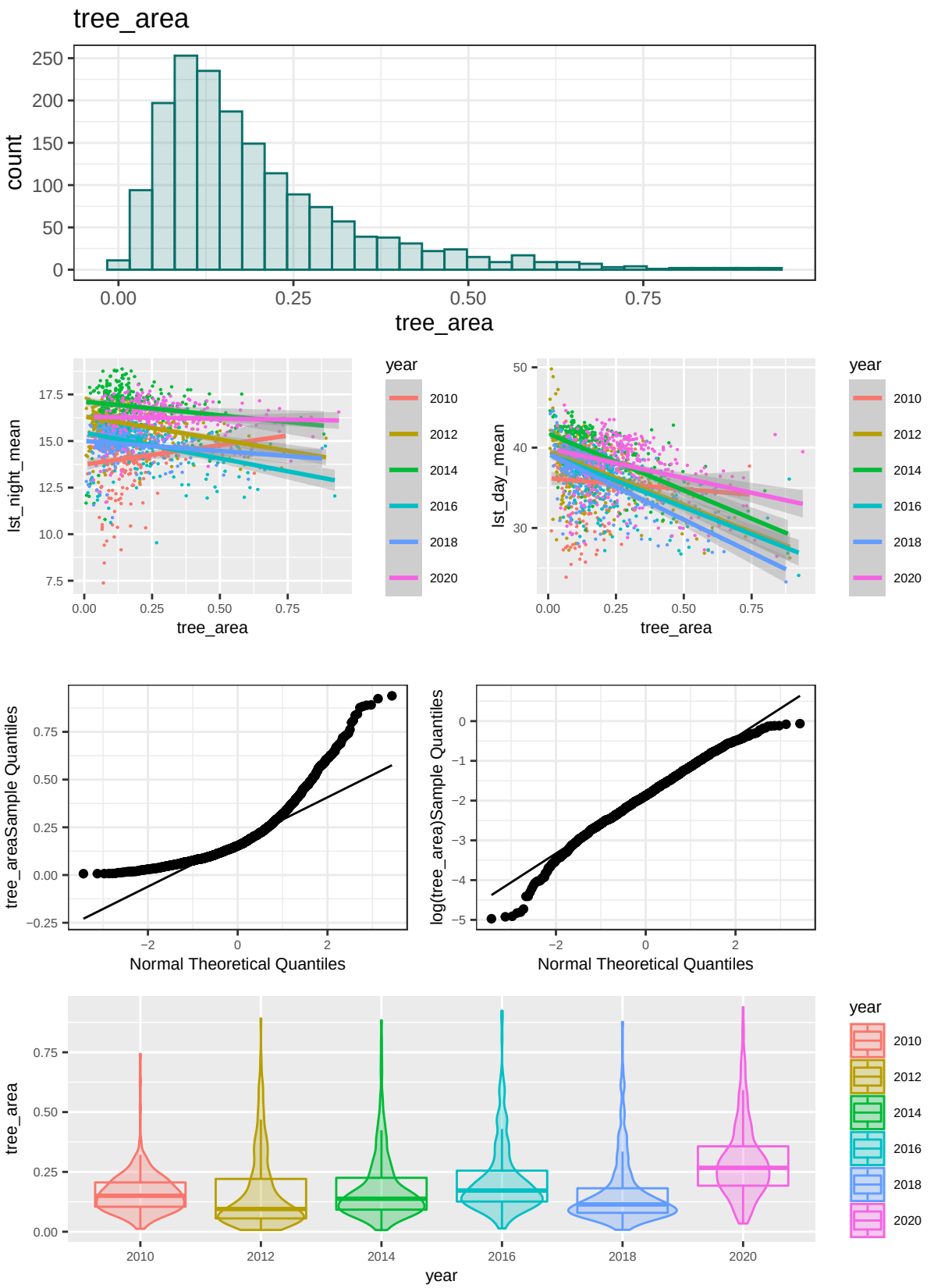
Google Earth: Random Forest Model

Lawn Area



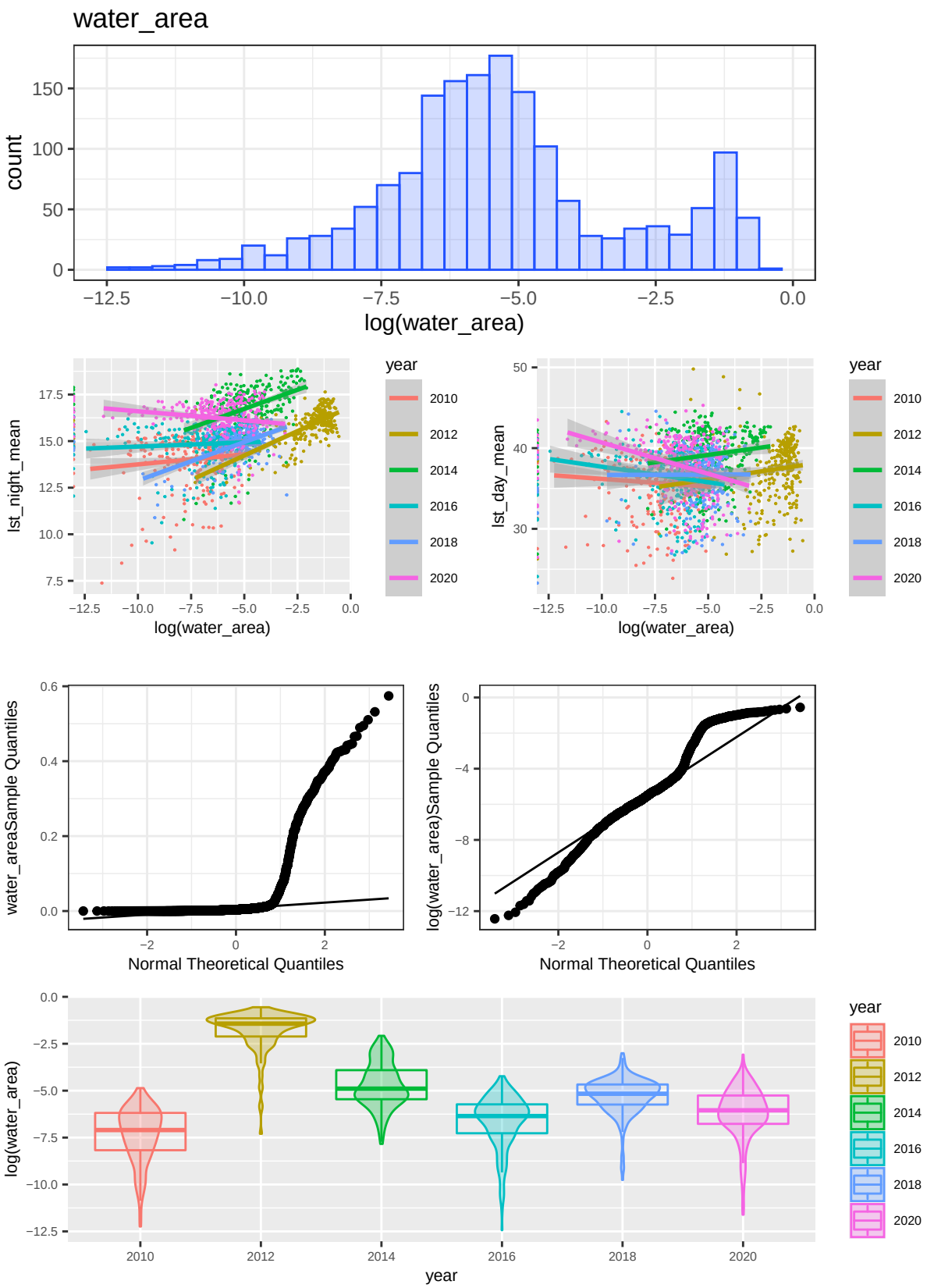
Google Earth: Random Forest Model

Tree Area



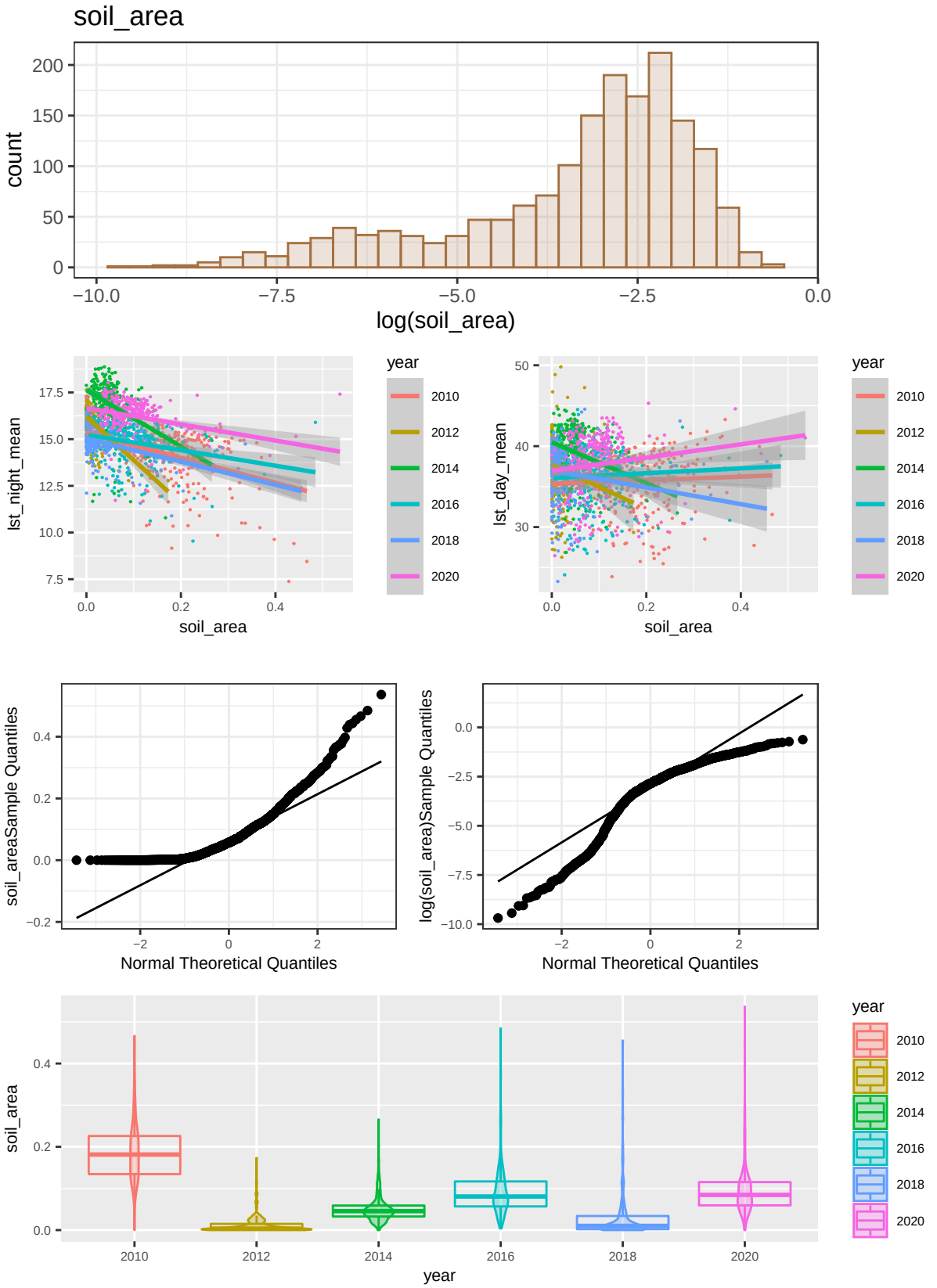
Google Earth: Random Forest Model

Water Area



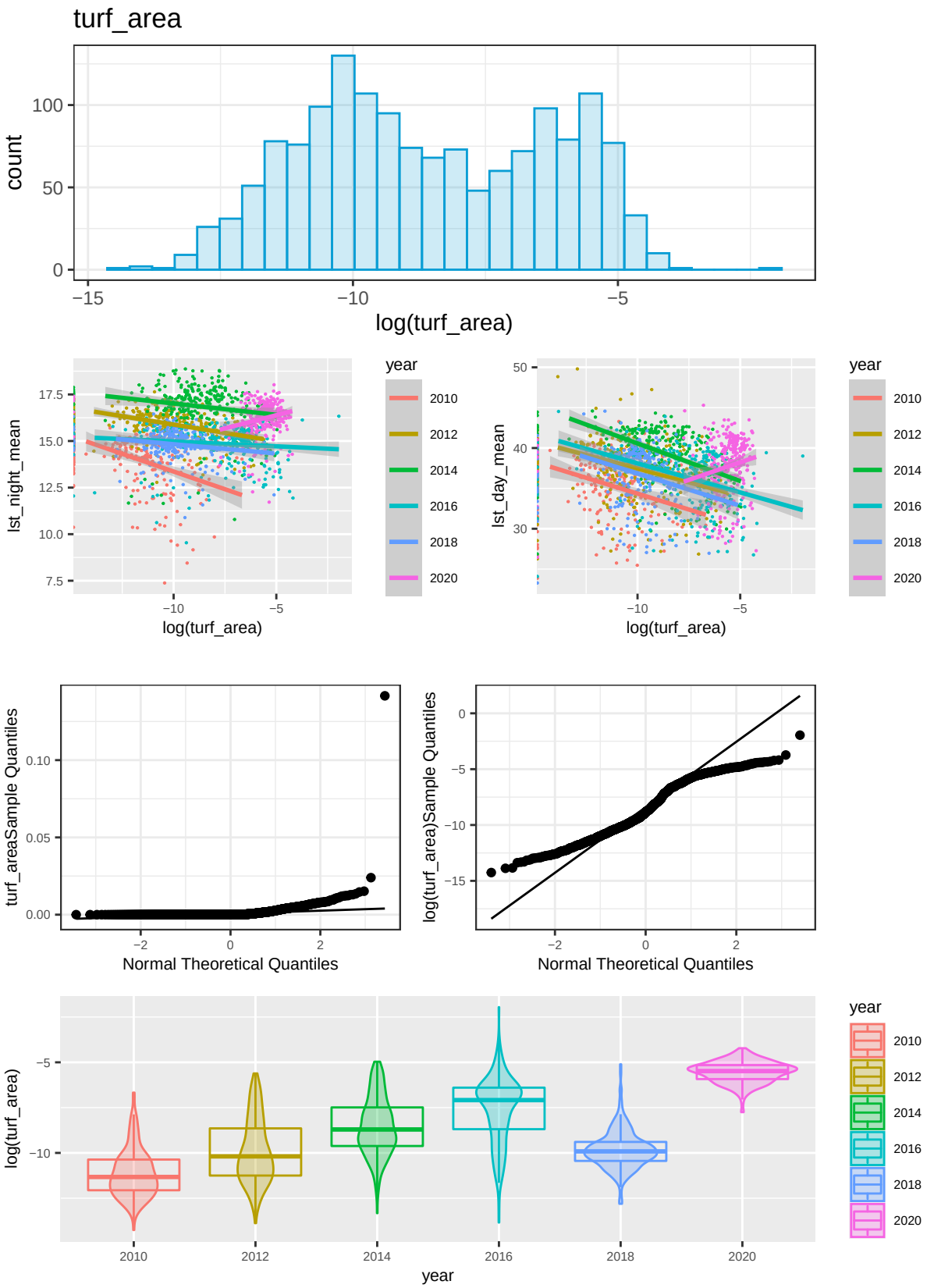
Google Earth: Random Forest Model

Soil Area



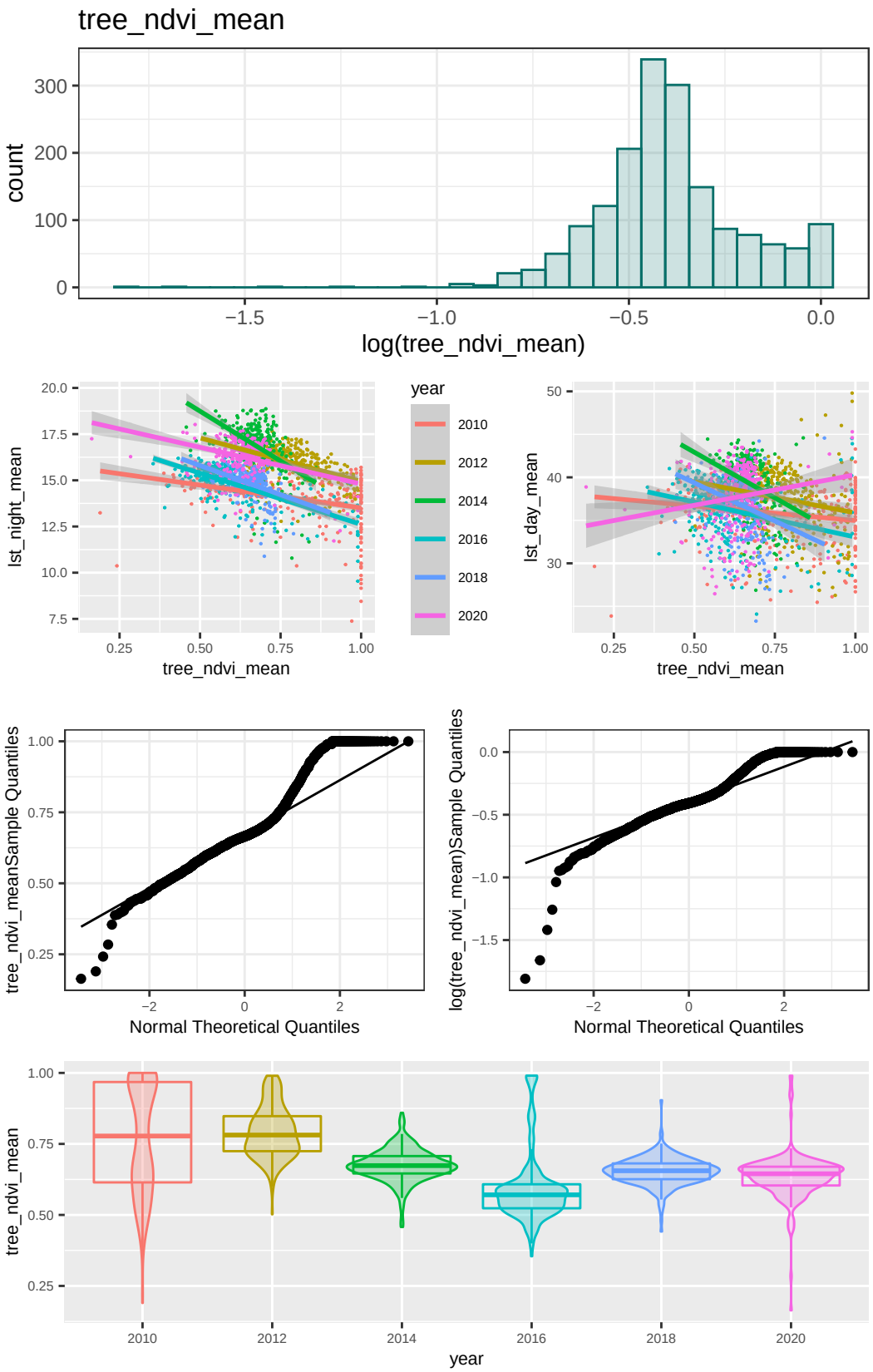
Google Earth: Random Forest Model

Turf Area



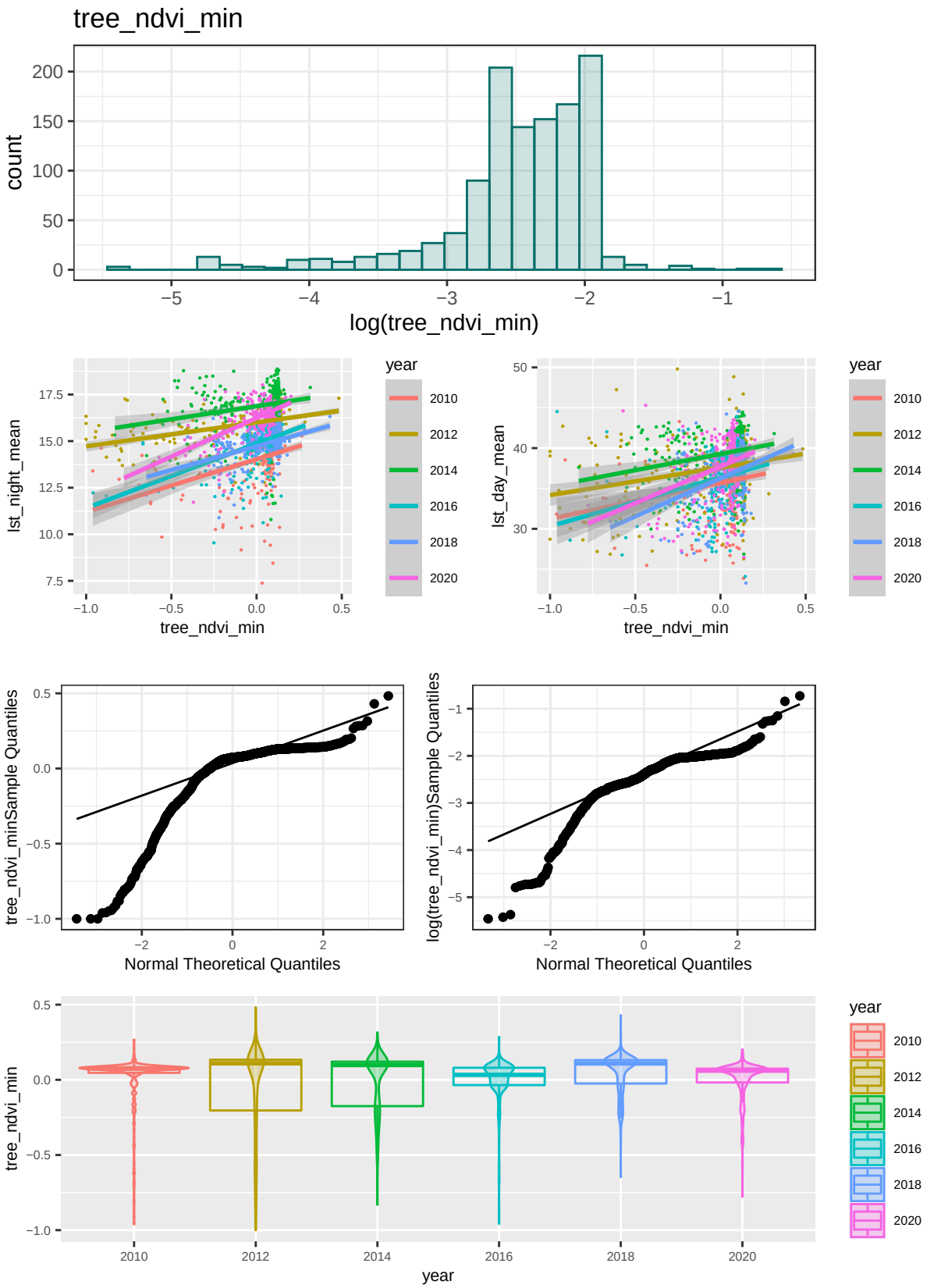
Google Earth: Random Forest Model

Tree NDVI Mean



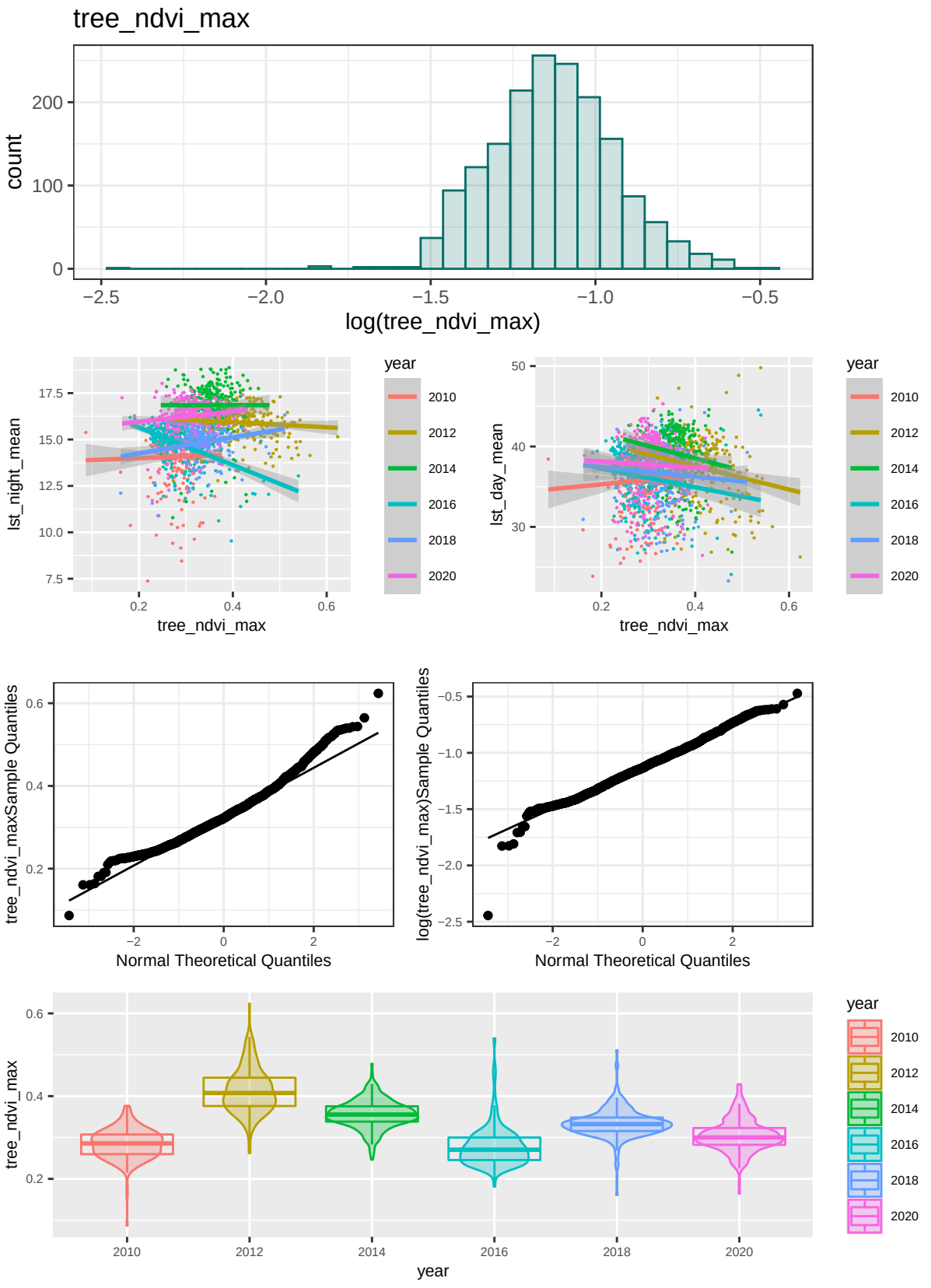
Google Earth: Random Forest Model

Tree NDVI Min



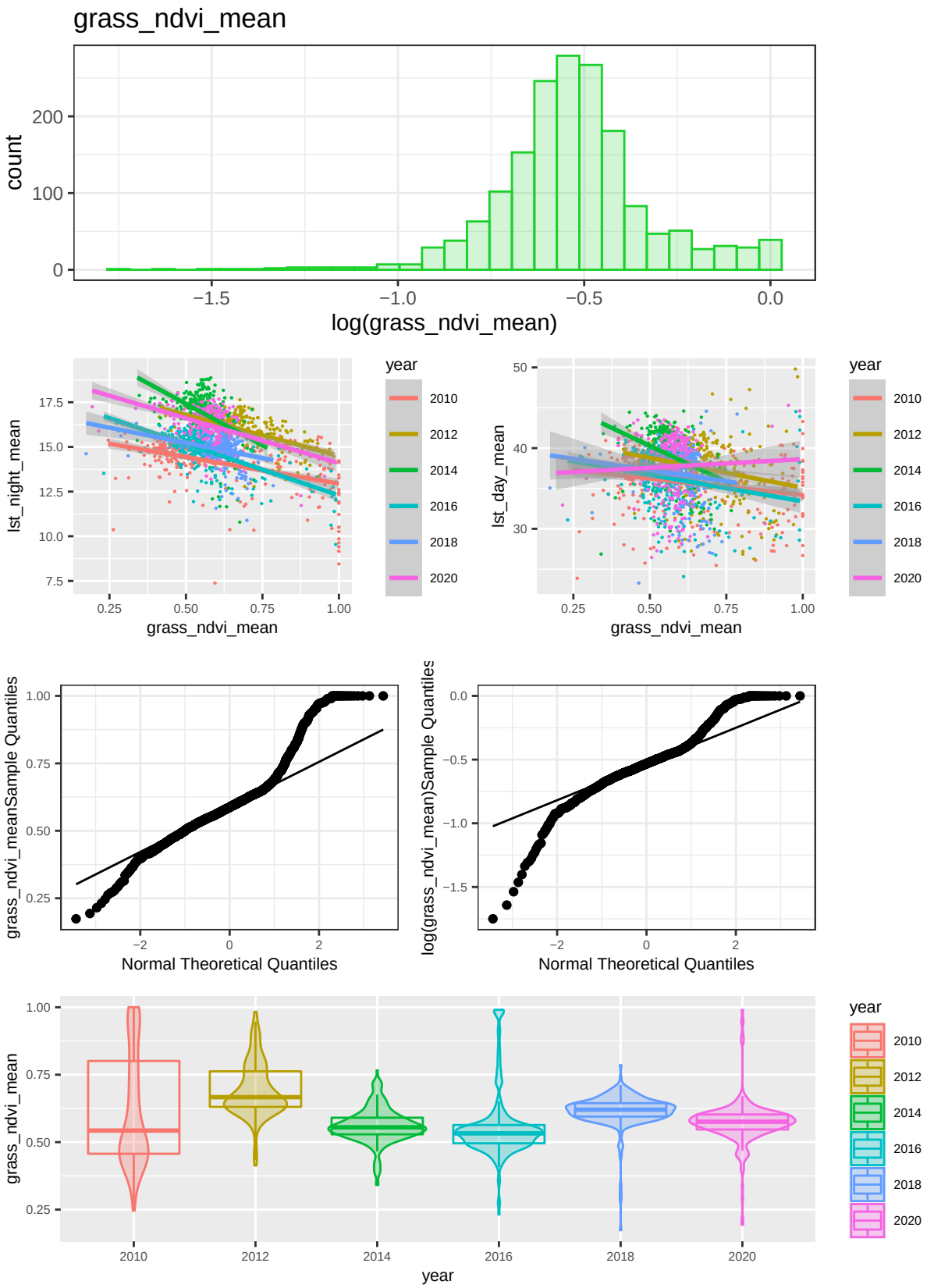
Google Earth: Random Forest Model

Tree NDVI Max



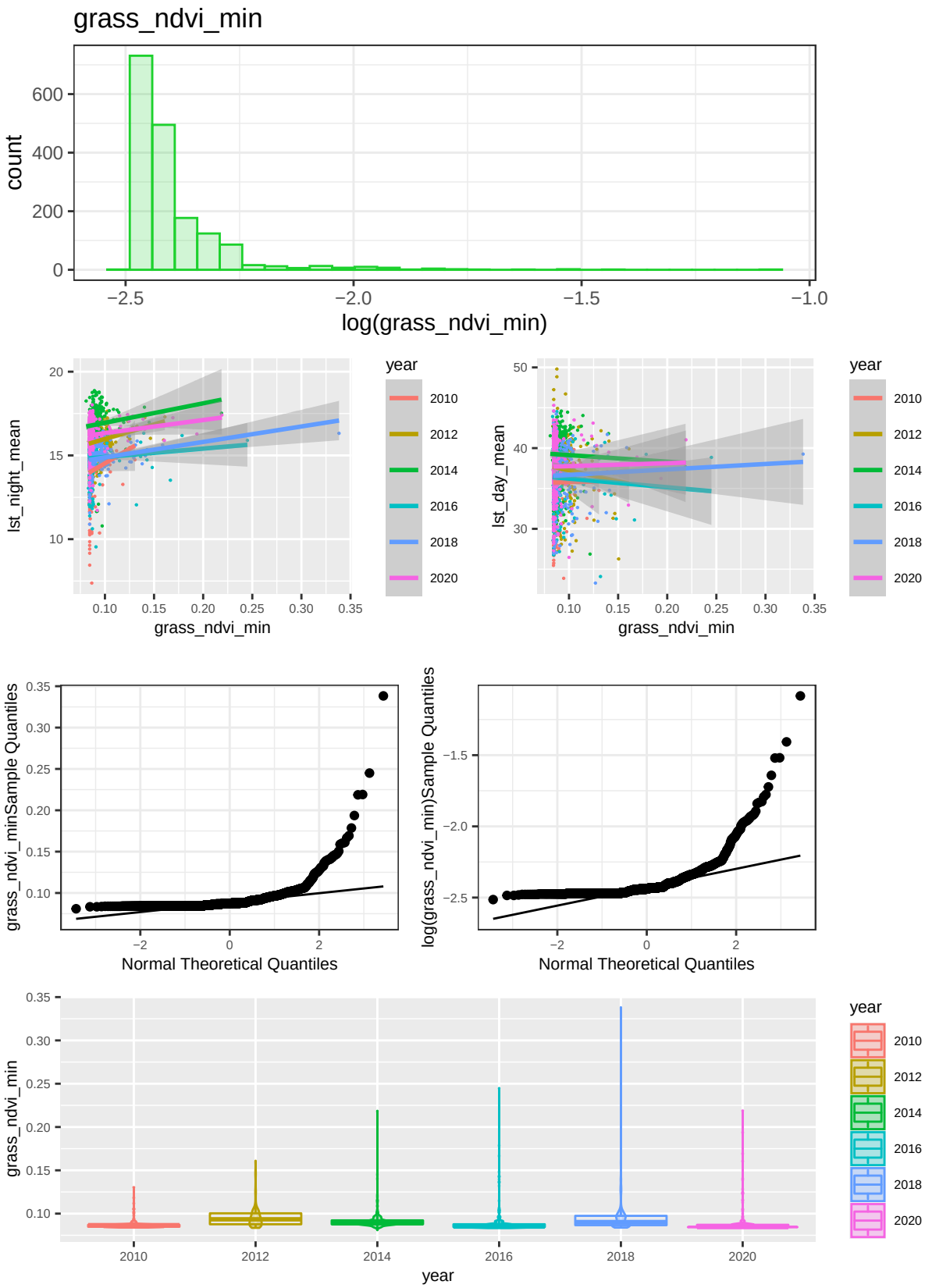
Google Earth: Random Forest Model

Lawn NDVI Mean



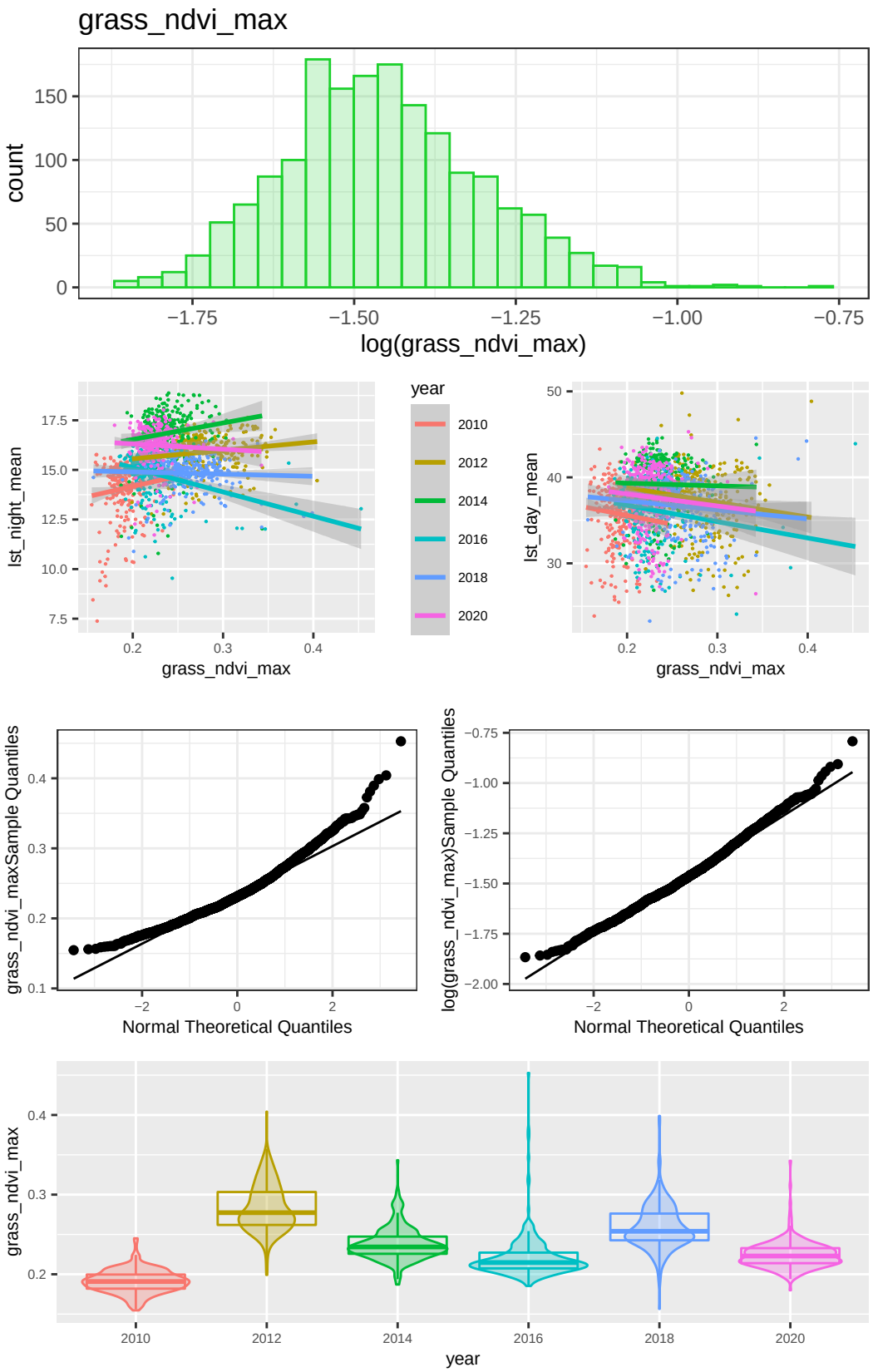
Google Earth: Random Forest Model

Lawn NDVI Min



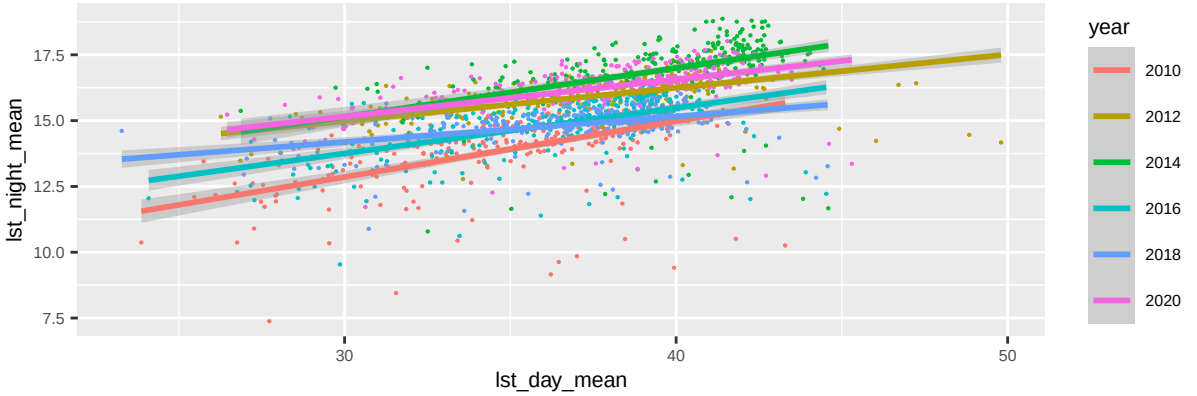
Google Earth: Random Forest Model

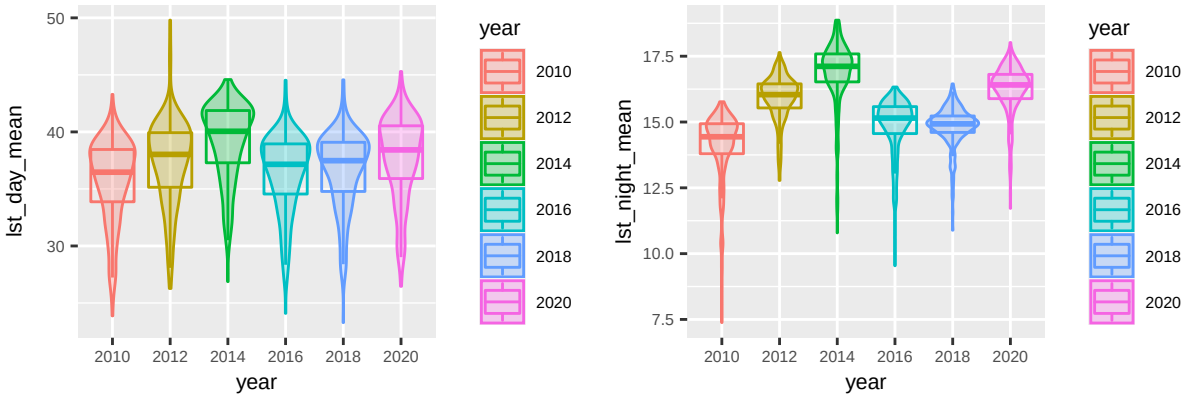
Lawn NDVI Max

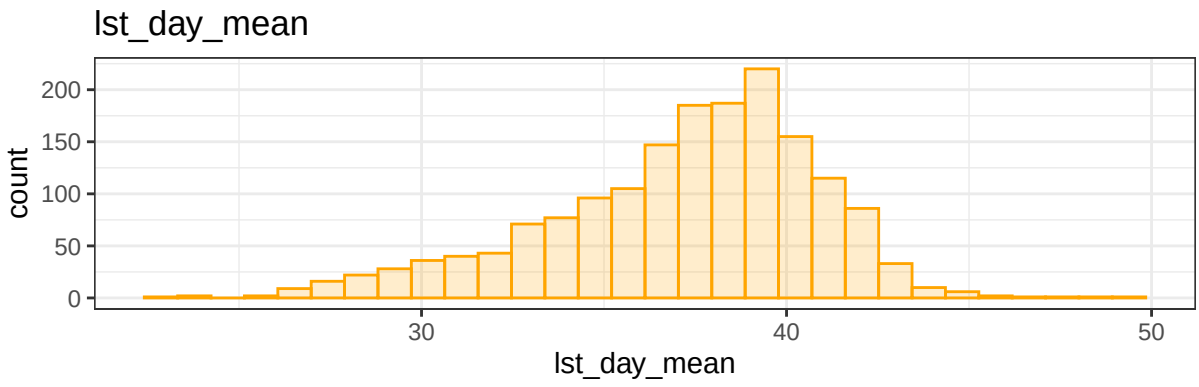


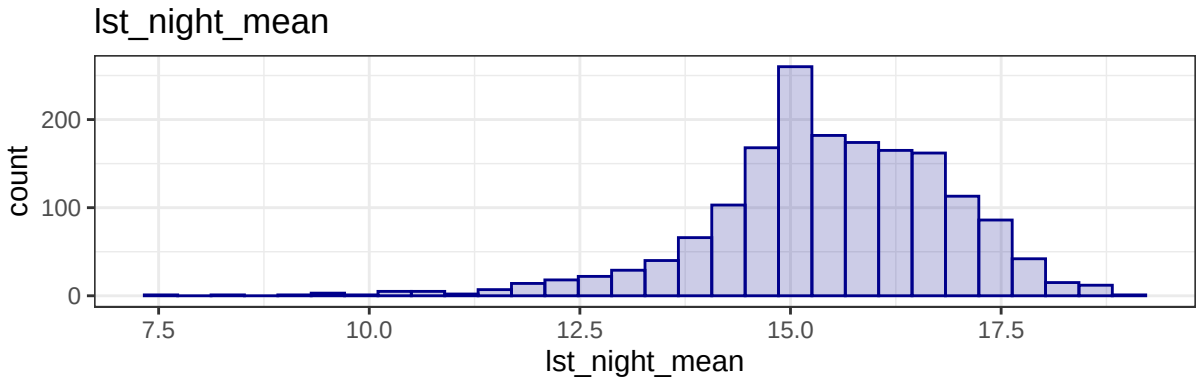
Google Earth: Random Forest Model

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(
  log(lst_day_mean) ~ 0 + factor(year) + tree_area + lawn_area + log(water_area) + log(soil_are
    log(turf_area) + grass_ndvi_mean + tree_ndvi_mean,
  data = microClimatePanel)

OLSM2 <- lm(
  log(lst_night_mean) ~ 0 + factor(year) + tree_area + lawn_area + log(water_area) + log(soil_a
    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: Wed, Jul 27, 2022 - 7:53:58 AM

Second Attempt with a Mixed Effects Model

```
feM1 <- plm(
  log(lst_day_mean) ~ 0 + factor(year) + tree_area + lawn_area + log(water_area) + log(soil_are
    log(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 + log(water_area) + log(soil_a
    log(turf_area) + grass_ndvi_mean + tree_ndvi_mean,
```

Table 1: OLS Model

	Dependent variable:	
	log(lst_day_mean) OLSModel Day Temp	log(lst_night_mean) OLSModel Night Temp
	(1)	(2)
factor(year)2014	3.760^{***} (0.026)	3.100^{***} (0.018)
factor(year)2016	3.696^{***} (0.023)	2.985^{***} (0.016)
factor(year)2018	3.691^{***} (0.027)	2.993^{***} (0.019)
factor(year)2020	3.768^{***} (0.024)	3.086^{***} (0.017)
tree_area	-0.365^{***} (0.019)	-0.020 (0.013)
lawn_area	0.039 (0.052)	0.144^{***} (0.036)
log(water_area)	-0.002 (0.001)	0.006^{***} (0.001)
log(soil_area)	0.002^{*} (0.001)	-0.002^{*} (0.001)
log(turf_area)	-0.002^{**} (0.001)	-0.001 (0.001)
grass_ndvi_mean	-0.117^{*} (0.065)	-0.400^{***} (0.045)
tree_ndvi_mean	0.019 (0.064)	-0.075^{*} (0.045)
Observations	1,132	1,132
R^2	0.999	1.000
Adjusted R^2	0.999	1.000
Residual Std. Error (df = 1121)	0.082	0.058
F Statistic (df = 11; 1121)	$198{,}820.500^{***}$	$235{,}249.600^{***}$
Note:		*p<0.1; **p<0.05; ***p<0.01

```
  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",
                                       "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 - 7:53:59 AM

Table 2: Fixed Effects Model

	Dependent variable:	
	log(lst_day_mean) FixedEffects Day Temp	log(lst_night_mean) FixedEffects Night Temp
	(1)	(2)
factor(year)2014	0.037^{***} (0.002)	0.034^{***} (0.003)
factor(year)2016	-0.043^{***} (0.002)	-0.091^{***} (0.003)
factor(year)2018	-0.038^{***} (0.002)	-0.085^{***} (0.004)
tree_area	-0.065^{***} (0.013)	−0.00003 (0.020)
lawn_area	-0.068^{***} (0.020)	0.071^{**} (0.031)
log(water_area)	-0.002^{***} (0.001)	0.0004 (0.001)
log(soil_area)	−0.001 (0.0003)	0.002^{***} (0.001)
log(turf_area)	−0.0003 (0.0003)	−0.00005 (0.0004)
grass_ndvi_mean	0.042^{**} (0.016)	-0.043^{*} (0.025)
tree_ndvi_mean	0.014 (0.015)	−0.009 (0.023)
Observations	1,132	1,132
R^2	0.824	0.870
Adjusted R^2	0.763	0.824
F Statistic (df = 10; 837)	392.634^{***}	559.040^{***}
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 = 97.752, 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(
  log(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(
  log(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: Wed, Jul 27, 2022 - 7:54:00 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 = 725.76, 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) RandomEffects Day Temp	log(lst_night_mean) RandomEffects Night Temp
	(1)	(2)
tree_area	0.405*** (0.094)	0.730*** (0.098)
lawn_area	0.972*** (0.173)	1.606*** (0.181)
log(water_area)	−0.042*** (0.004)	−0.051*** (0.004)
log(soil_area)	−0.004 (0.003)	−0.001 (0.003)
log(turf_area)	−0.011*** (0.002)	−0.012*** (0.002)
grass_ndvi_mean	0.072 (0.124)	−0.004 (0.132)
tree_ndvi_mean	1.307*** (0.110)	1.836*** (0.115)
Observations	1,132	1,132
R^2	0.020	0.021
Adjusted R^2	0.015	0.016
F Statistic	739.254***	2,395.437***
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 + 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: Wed, Jul 27, 2022 - 7:54:00 AM

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

```
##
##  F test for individual effects
##
## data:  log(lst_day_mean) ~ 0 + factor(year) + tree_area + lawn_area +  ...
## F = 584.46, df1 = 3, df2 = 837, p-value < 2.2e-16
## alternative hypothesis: significant effects
```

Table 4:

	Dependent variable:
	log(lst_day_mean)
tree_area	0.063*** (0.017)
lawn_area	0.229*** (0.032)
log(water_area)	0.008*** (0.001)
log(soil_area)	0.001 (0.001)
log(turf_area)	0.002*** (0.0004)
grass_ndvi_mean	−0.360*** (0.023)
tree_ndvi_mean	0.366*** (0.021)
Observations	1,132
R^2	0.456
Adjusted R^2	0.268
F Statistic	100.664*** (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 = 49.622, df = 1, p-value = 1.864e-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  0.03708130  0.00230223  16.1067 < 2.2e-16 ***
```

```
## factor(year)2016 -0.04283239  0.00203360 -21.0623 < 2.2e-16 ***
## factor(year)2018 -0.03807182  0.00276872 -13.7507 < 2.2e-16 ***
## tree_area        -0.06456127  0.01585514  -4.0719 5.105e-05 ***
## lawn_area        -0.06789676  0.02823292  -2.4049  0.016394 *
## log(water_area)  -0.00233119  0.00057168  -4.0778 4.981e-05 ***
## log(soil_area)   -0.00052546  0.00030555  -1.7197  0.085851 .
## log(turf_area)   -0.00034270  0.00025895  -1.3234  0.186056
## grass_ndvi_mean   0.04239605  0.01599228   2.6510  0.008176 **
## tree_ndvi_mean    0.01430849  0.01551006   0.9225  0.356518
## ---
## 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  0.03708130  0.00230223  16.1067 < 2.2e-16 ***
## factor(year)2016 -0.04283239  0.00203360 -21.0623 < 2.2e-16 ***
## factor(year)2018 -0.03807182  0.00276872 -13.7507 < 2.2e-16 ***
## tree_area        -0.06456127  0.01585514  -4.0719 5.105e-05 ***
## lawn_area        -0.06789676  0.02823292  -2.4049  0.016394 *
## log(water_area)  -0.00233119  0.00057168  -4.0778 4.981e-05 ***
## log(soil_area)   -0.00052546  0.00030555  -1.7197  0.085851 .
## log(turf_area)   -0.00034270  0.00025895  -1.3234  0.186056
## grass_ndvi_mean   0.04239605  0.01599228   2.6510  0.008176 **
## tree_ndvi_mean    0.01430849  0.01551006   0.9225  0.356518
## ---
## 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      0.002302231      0.002033604      0.002768725 0.01585514 0.02823292
## HC1      0.002312467      0.002042647      0.002781036 0.01592563 0.02835846
## HC2      0.002325512      0.002056569      0.002803880 0.01612824 0.02876941
## HC3      0.002350047      0.002080514      0.002840494 0.01641328 0.02932985
## HC4      0.002378747      0.002110571      0.002891514 0.01689290 0.03031314
##     log(water_area) log(soil_area) log(turf_area) grass_ndvi_mean
## HC0    0.0005716837   0.0003055489   0.0002589473      0.01599228
## HC1    0.0005742257   0.0003069075   0.0002600987      0.01606339
## HC2    0.0005897947   0.0003140536   0.0002632202      0.01616236
## HC3    0.0006095076   0.0003230048   0.0002676627      0.01633786
## HC4    0.0006507549   0.0003411647   0.0002748971      0.01653642
##     tree_ndvi_mean
```

```
## HC0     0.01551006
## HC1     0.01557902
## HC2     0.01567376
## HC3     0.01584307
## HC4     0.01603889
```

```
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  0.03708130  0.00230223  16.1067 < 2.2e-16 ***
## factor(year)2016 -0.04283239  0.00203360 -21.0623 < 2.2e-16 ***
## factor(year)2018 -0.03807182  0.00276872 -13.7507 < 2.2e-16 ***
## tree_area        -0.06456127  0.01585514  -4.0719 5.105e-05 ***
## lawn_area        -0.06789676  0.02823292  -2.4049  0.016394 *
## log(water_area)  -0.00233119  0.00057168  -4.0778 4.981e-05 ***
## log(soil_area)   -0.00052546  0.00030555  -1.7197  0.085851 .
## log(turf_area)   -0.00034270  0.00025895  -1.3234  0.186056
## grass_ndvi_mean   0.04239605  0.01599228   2.6510  0.008176 **
## tree_ndvi_mean    0.01430849  0.01551006   0.9225  0.356518
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
print(cInterval)
```

```
##                          2.5 %        97.5 %
## factor(year)2014  0.0325624719  4.160012e-02
## factor(year)2016 -0.0468239575 -3.884083e-02
## factor(year)2018 -0.0435062845 -3.263737e-02
## tree_area        -0.0956817666 -3.344077e-02
## lawn_area        -0.1233124055 -1.248111e-02
## log(water_area)  -0.0034532942 -1.209090e-03
## log(soil_area)   -0.0011251943  7.426968e-05
## log(turf_area)   -0.0008509595  1.655652e-04
## grass_ndvi_mean   0.0110063725  7.378573e-02
## tree_ndvi_mean   -0.0161346879  4.475167e-02
```

```
stargazer::stargazer(OLSM1,feM1,reM1,feM2FixedTime,
                     font.size = 'tiny',
                     title = 'Regression Models Sumamry',
```

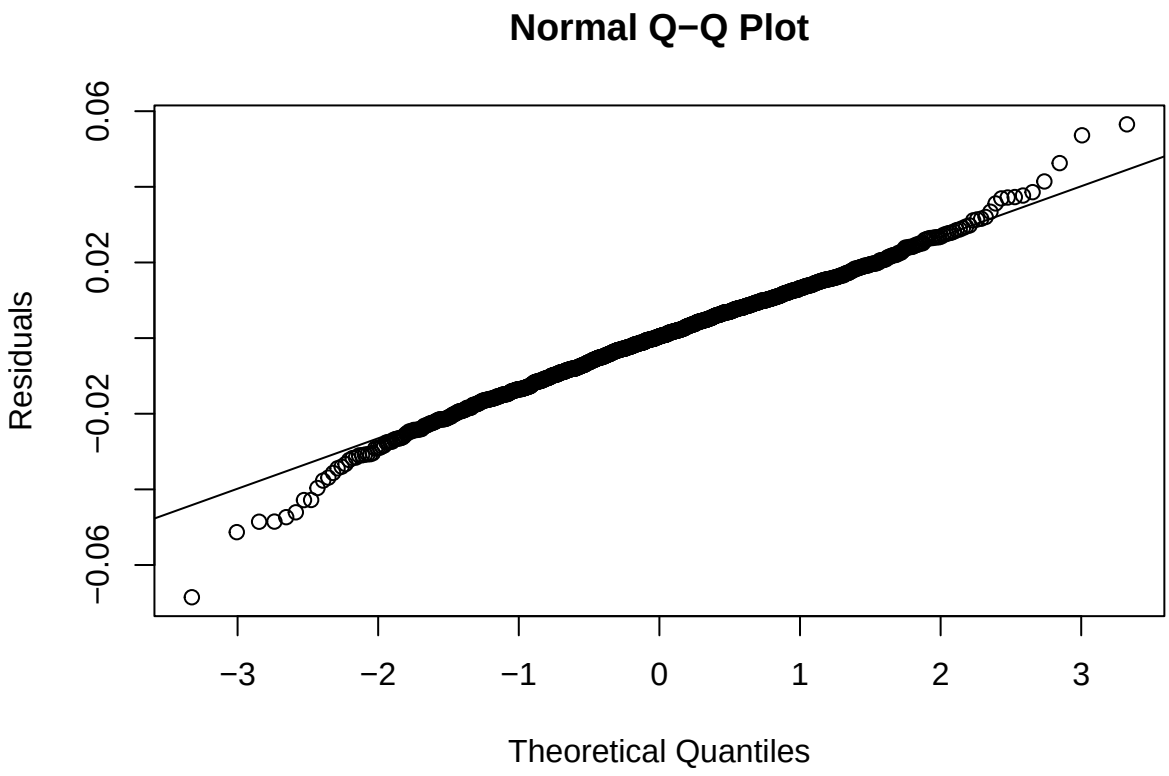
```
                    column.labels = c("OLS", "FixedEffects",
                                      "RandomEffects",
                                      "FixedEffectsFixedTime"))
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Wed, Jul 27, 2022 - 7:54:02 AM

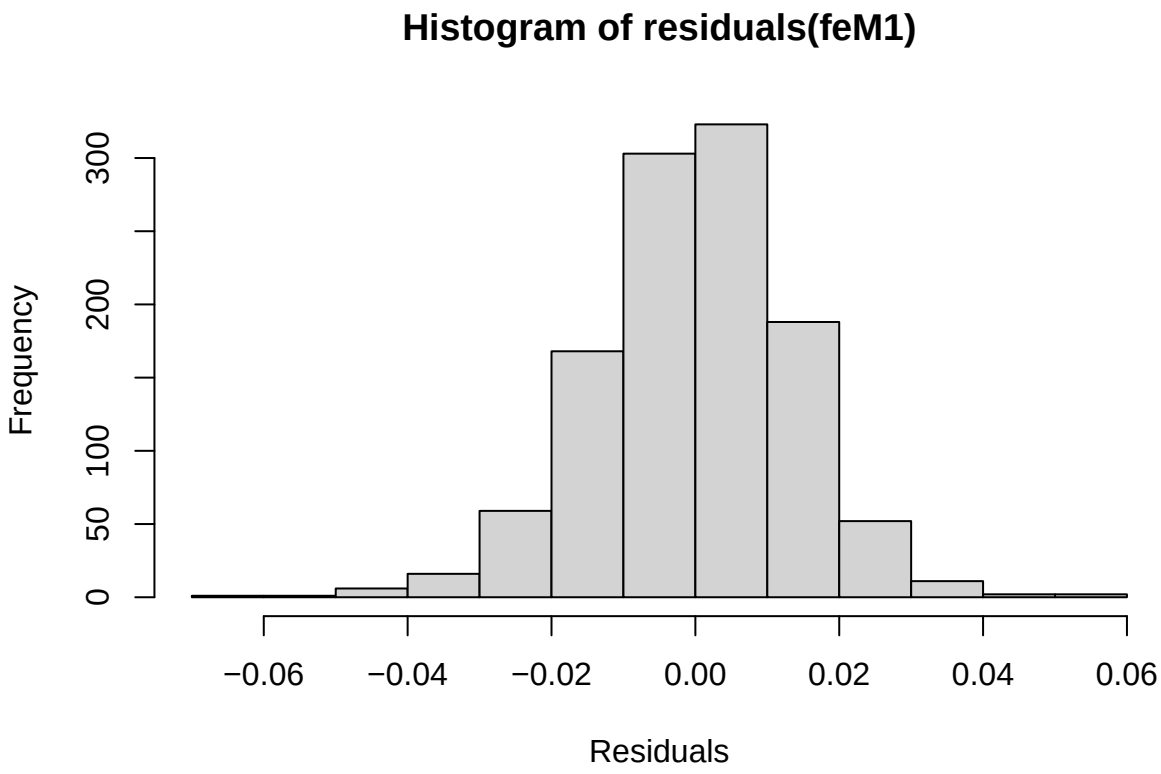
Table 5: Regression Models Sumamry

	Dependent variable:			
	log(lst_day_mean)			
	OLS		*panel linear*	
	OLS	FixedEffects	RandomEffects	FixedEffectsFixedTime
	(1)	(2)	(3)	(4)
factor(year)2014	3.760***	0.037***		
	(0.026)	(0.002)		
factor(year)2016	3.696***	−0.043***		
	(0.023)	(0.002)		
factor(year)2018	3.691***	−0.038***		
	(0.027)	(0.002)		
factor(year)2020	3.768***			
	(0.024)			
tree_area	−0.365***	−0.065***	0.405***	0.063***
	(0.019)	(0.013)	(0.094)	(0.017)
lawn_area	0.039	−0.068***	0.972***	0.229***
	(0.052)	(0.020)	(0.173)	(0.032)
log(water_area)	−0.002	−0.002***	−0.042***	0.008***
	(0.001)	(0.001)	(0.004)	(0.001)
log(soil_area)	0.002*	−0.001	−0.004	0.001
	(0.001)	(0.0003)	(0.003)	(0.001)
log(turf_area)	−0.002**	−0.0003	−0.011***	0.002***
	(0.001)	(0.0003)	(0.002)	(0.0004)
grass_ndvi_mean	−0.117*	0.042**	0.072	−0.360***
	(0.065)	(0.016)	(0.124)	(0.023)
tree_ndvi_mean	0.019	0.014	1.307***	0.366***
	(0.064)	(0.015)	(0.110)	(0.021)
Observations	1,132	1,132	1,132	1,132
R^2	0.999	0.824	0.020	0.456
Adjusted R^2	0.999	0.763	0.015	0.268
Residual Std. Error	0.082 (df = 1121)			
F Statistic	198,820.500*** (df = 11; 1121)	392.634*** (df = 10; 837)	739.254***	100.664*** (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)
Frequency
0
50
100
200
300
−0.06
−0.04
−0.02
0.00
0.02
0.04
0.06
Residuals