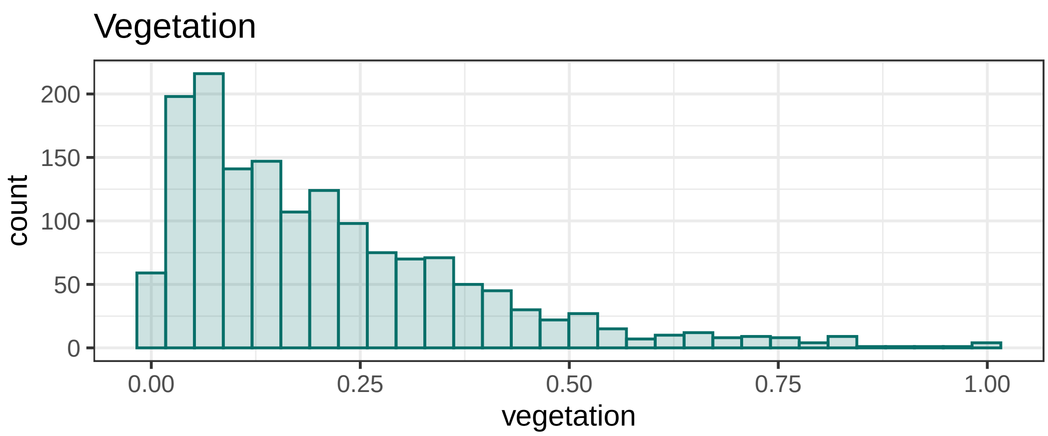
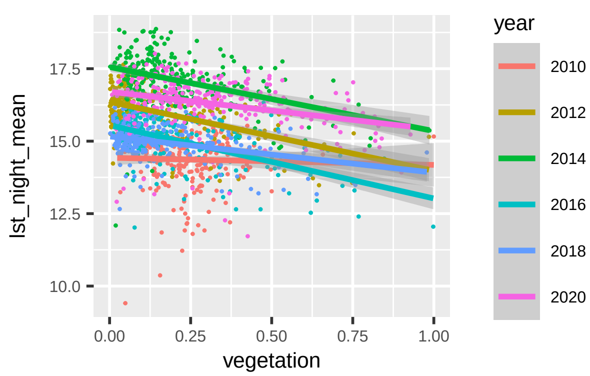
LawnProject Panel Data Analysis / Tensor Flow Data

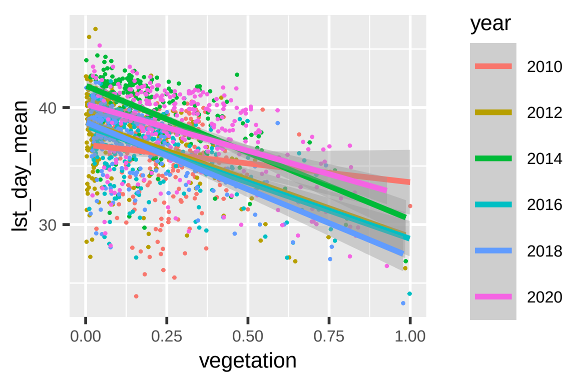
LawnProject Team

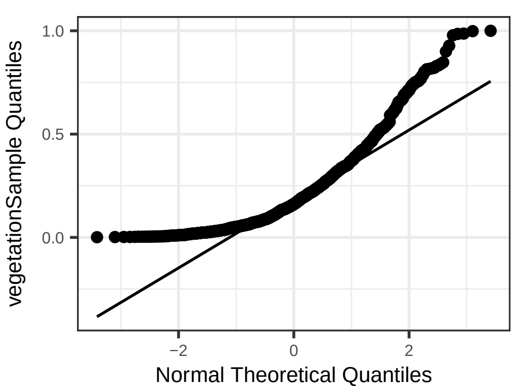
Intro

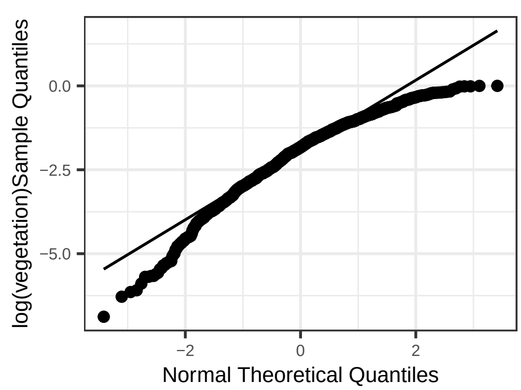
Aggregate Vegetation Area (Trees + Lawn)

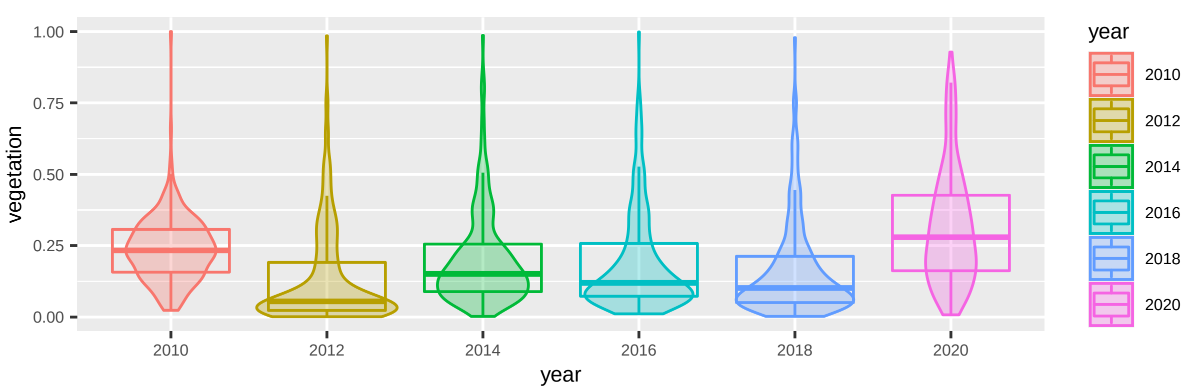




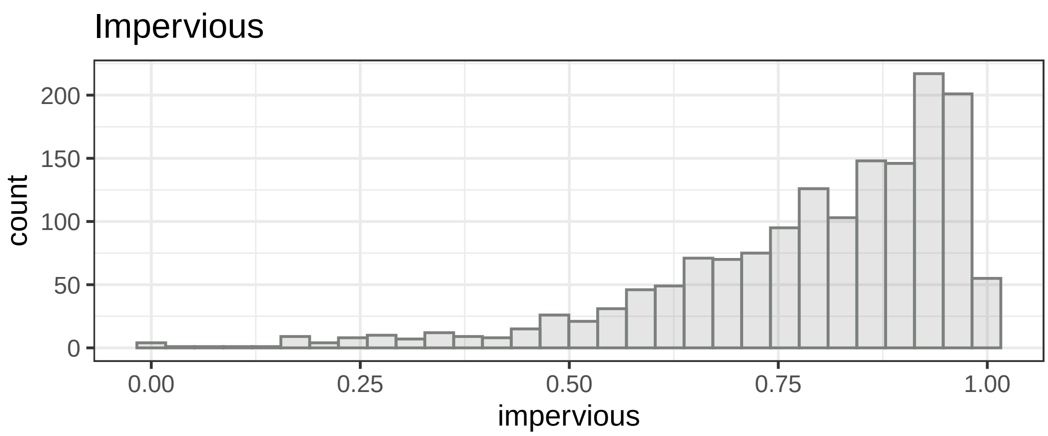


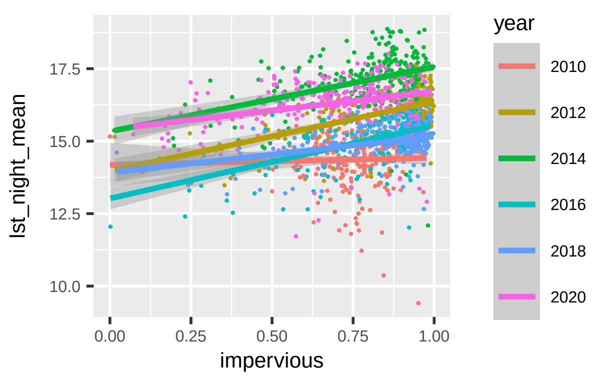


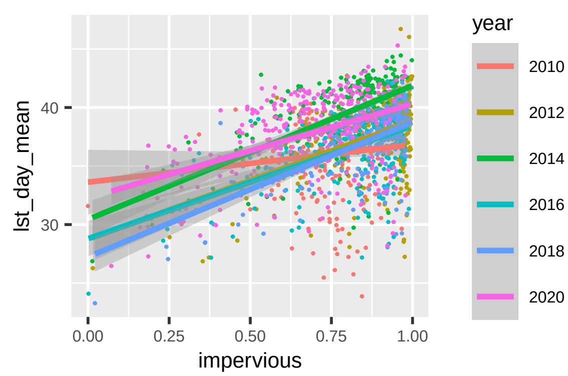


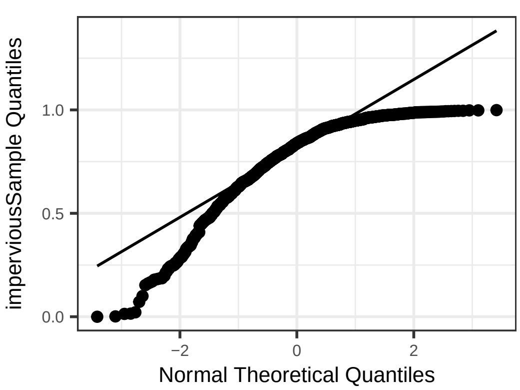


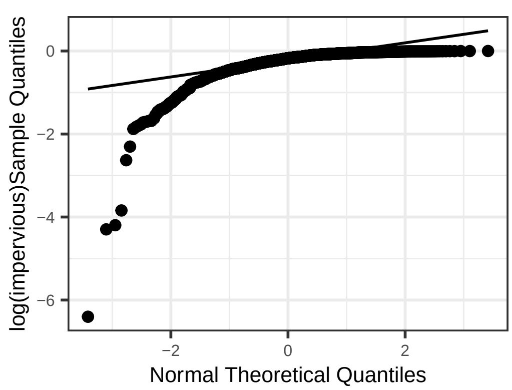
Aggregate Impervious Area (Soil + Turf + Impervious)

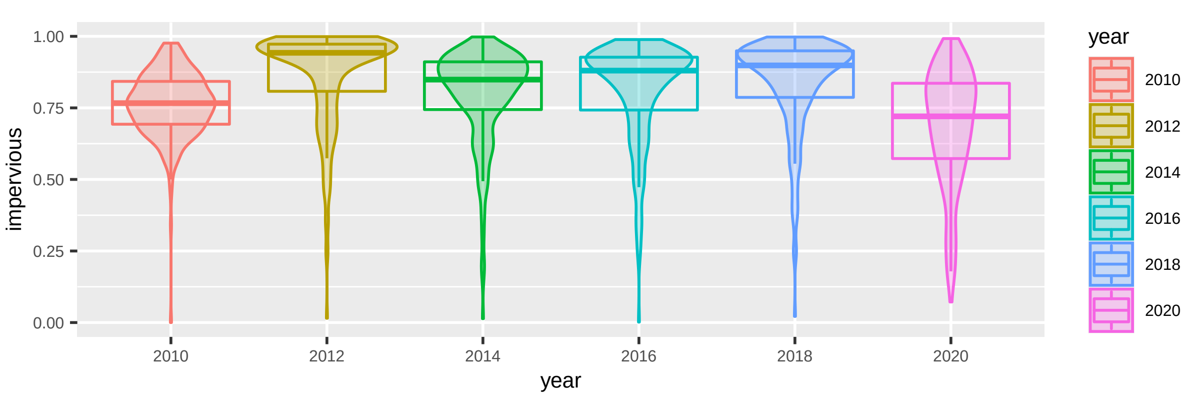




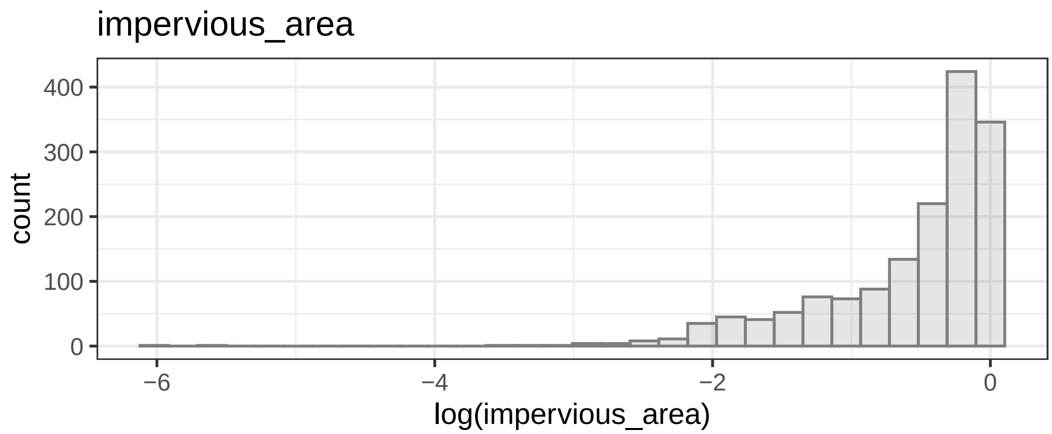


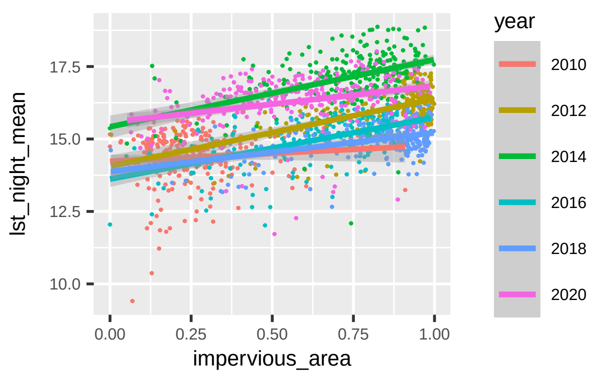


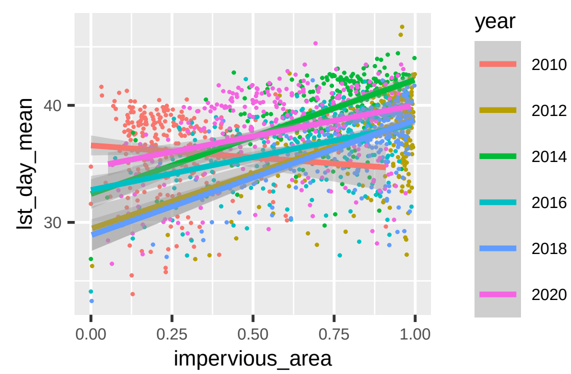


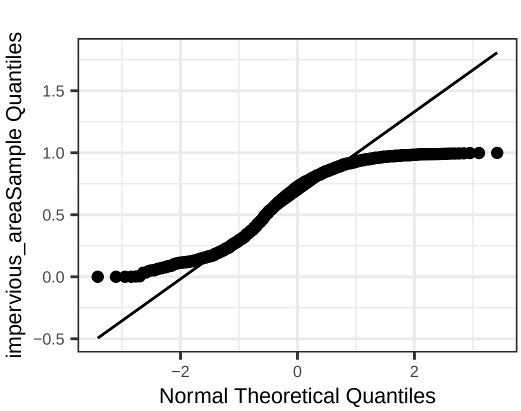


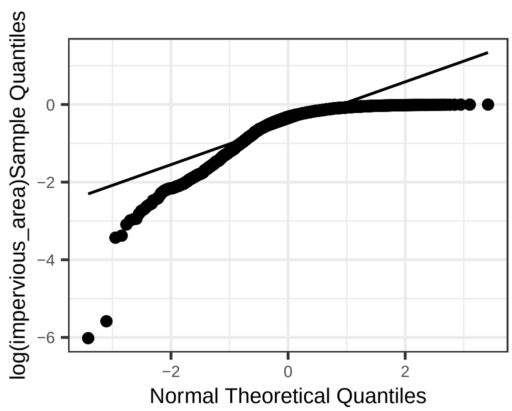
Impervious Area

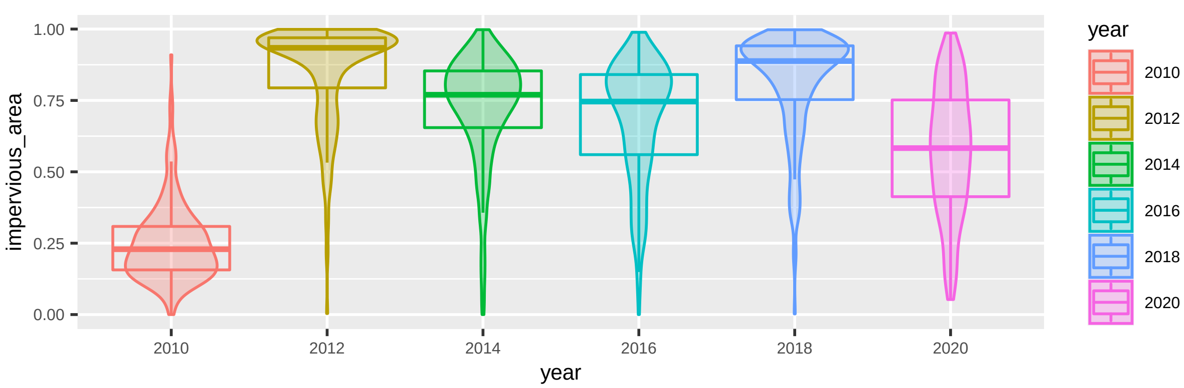




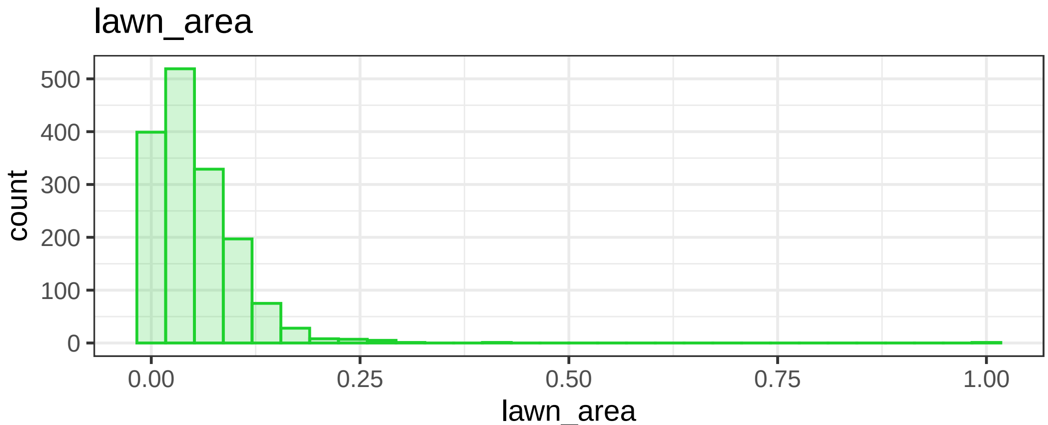


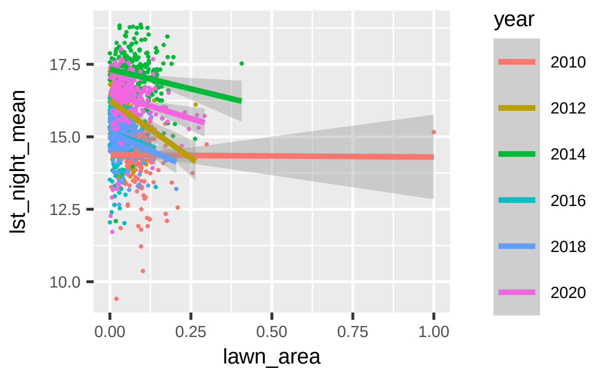


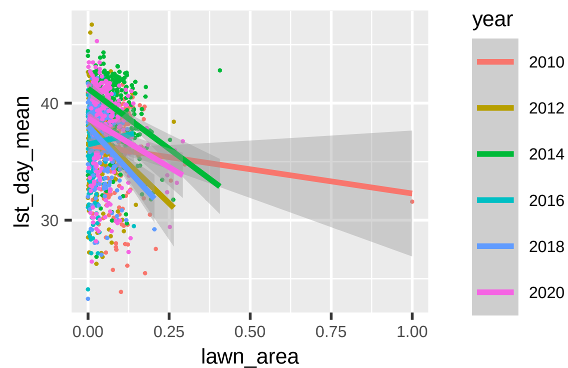


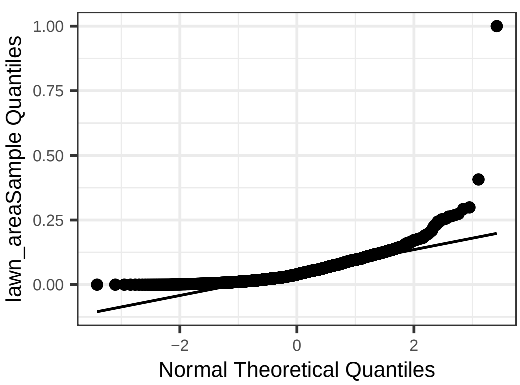


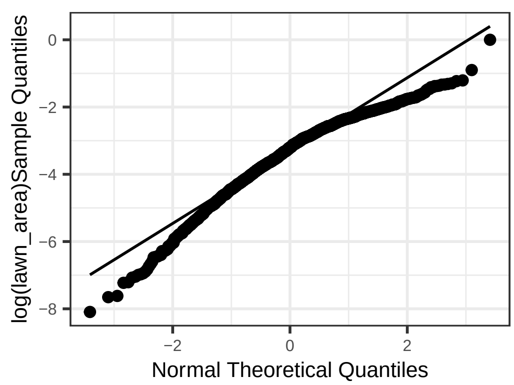
Lawn Area

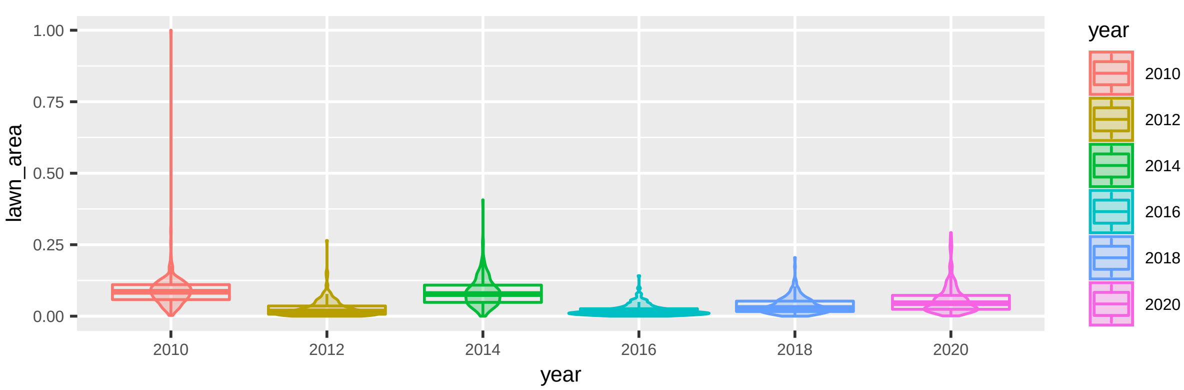

lawn_area
count
500
400
300
200
100
0
0.00
0.25
0.50
0.75
1.00
lawn_area


lst_night_mean
17.5
15.0
12.5
10.0
0.00
0.25
0.50
0.75
1.00
lawn_area
year
2010
2012
2014
2016
2018
2020

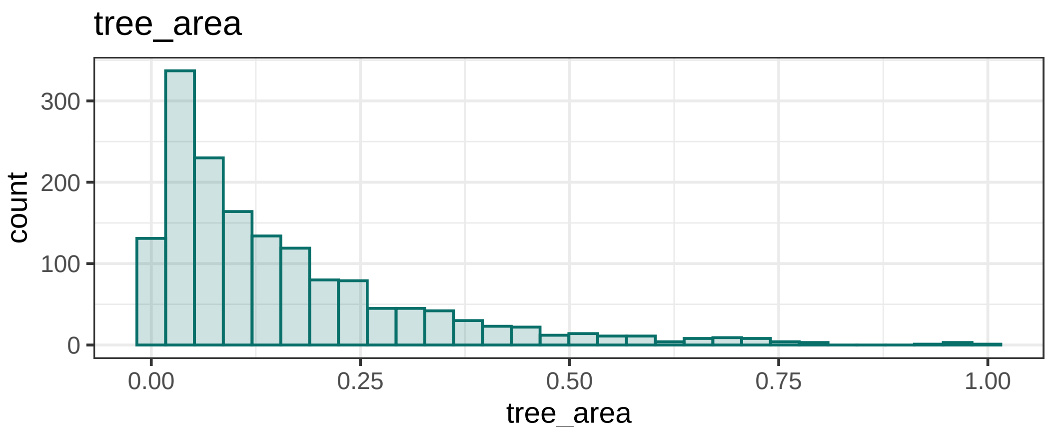

lst_day_mean
40
30
0.00
0.25
0.50
0.75
1.00
lawn_area
year
2010
2012
2014
2016
2018
2020

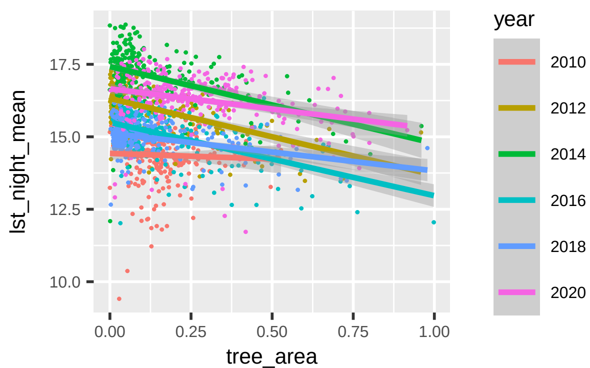

lawn_areaSample Quantiles
1.00
0.75
0.50
0.25
0.00
−2
0
2
Normal Theoretical Quantiles

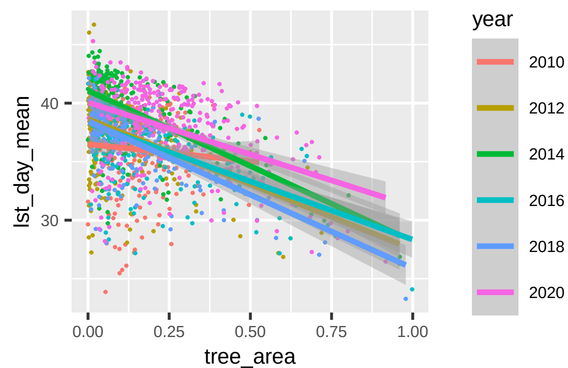

log(lawn_area)Sample Quantiles
0
−2
−4
−6
−8
−2
0
2
Normal Theoretical Quantiles

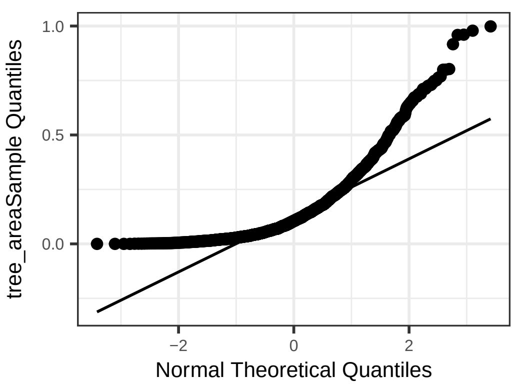

lawn_area
1.00
0.75
0.50
0.25
0.00
2010
2012
2014
2016
2018
2020
year
year
2010
2012
2014
2016
2018
2020

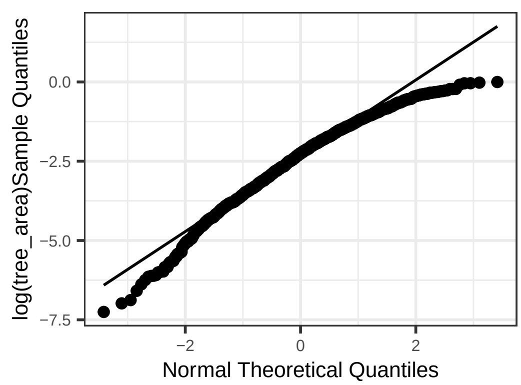
Tree Area

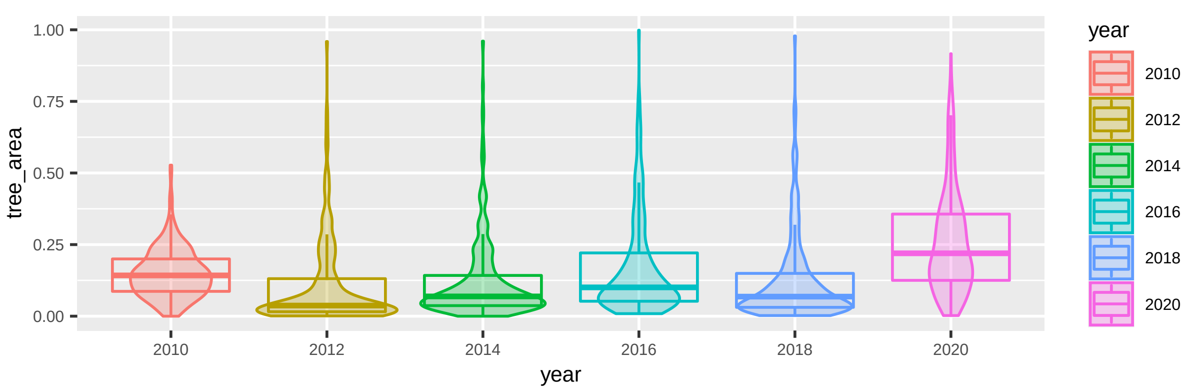

tree_area
count
300
200
100
0
0.00
0.25
0.50
0.75
1.00
tree_area


lst_night_mean
17.5
15.0
12.5
10.0
0.00
0.25
0.50
0.75
1.00
tree_area
year
2010
2012
2014
2016
2018
2020

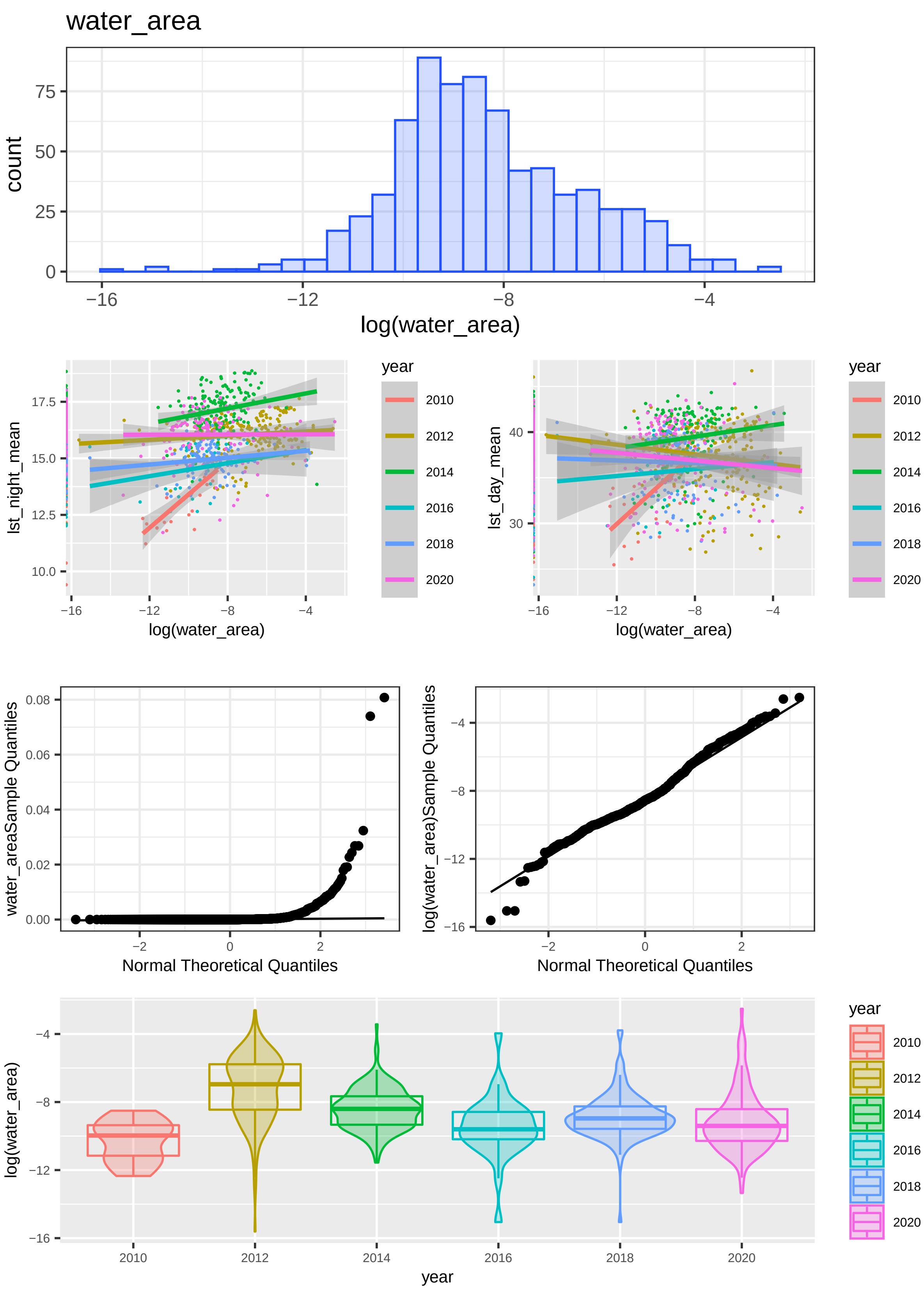

lst_day_mean
40
30
0.00
0.25
0.50
0.75
1.00
tree_area
year
2010
2012
2014
2016
2018
2020


tree_areaSample Quantiles
1.0
0.5
0.0
−2
0
2
Normal Theoretical Quantiles


log(tree_area)Sample Quantiles
0.0
−2.5
−5.0
−7.5
−2
0
2
Normal Theoretical Quantiles

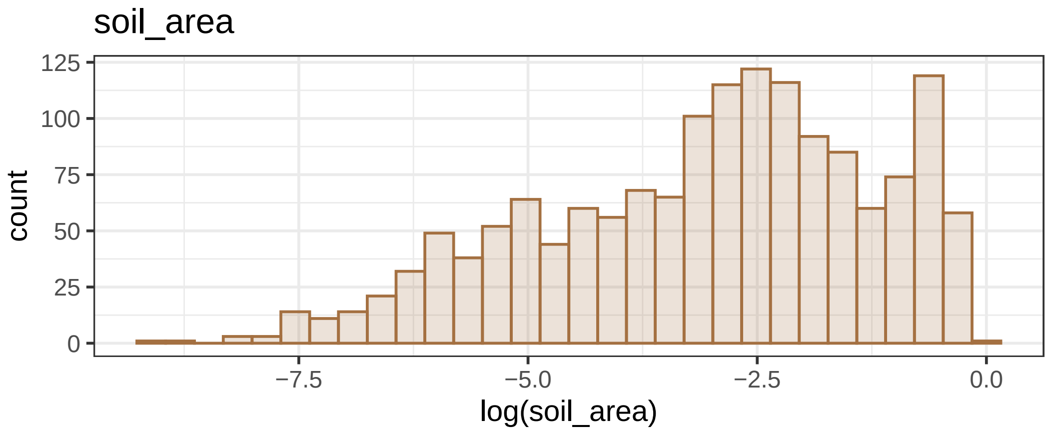

tree_area
1.00
0.75
0.50
0.25
0.00
2010
2012
2014
2016
2018
2020
year
year
2010
2012
2014
2016
2018
2020

Water Area

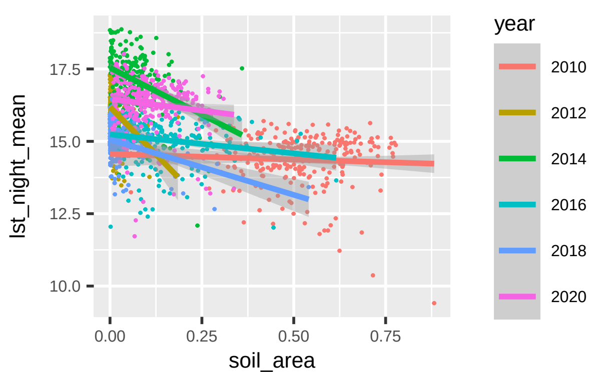


Google Earth: Tensor Flow Model

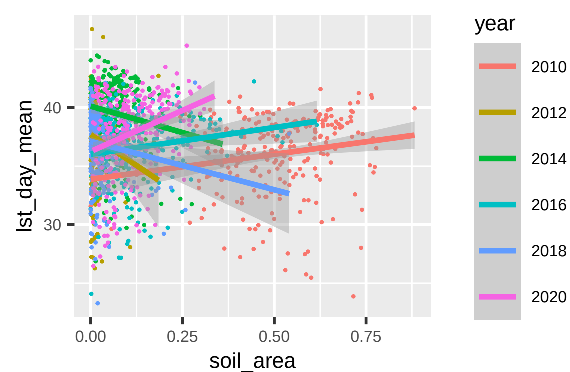
Soil Area



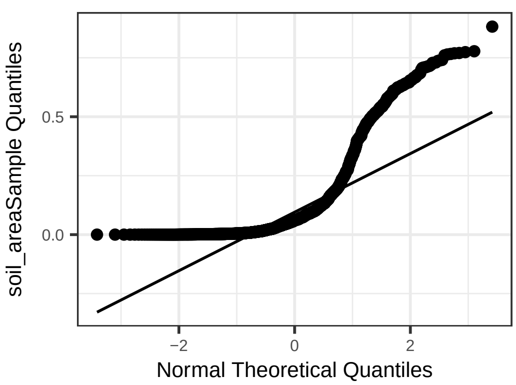
soil_area
count
125
100
75
50
25
0
−7.5
−5.0
−2.5
0.0
log(soil_area)



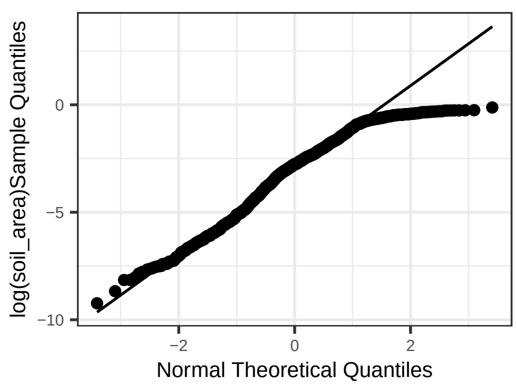
lst_night_mean
17.5
15.0
12.5
10.0
0.00
0.25
0.50
0.75
soil_area
year
2010
2012
2014
2016
2018
2020



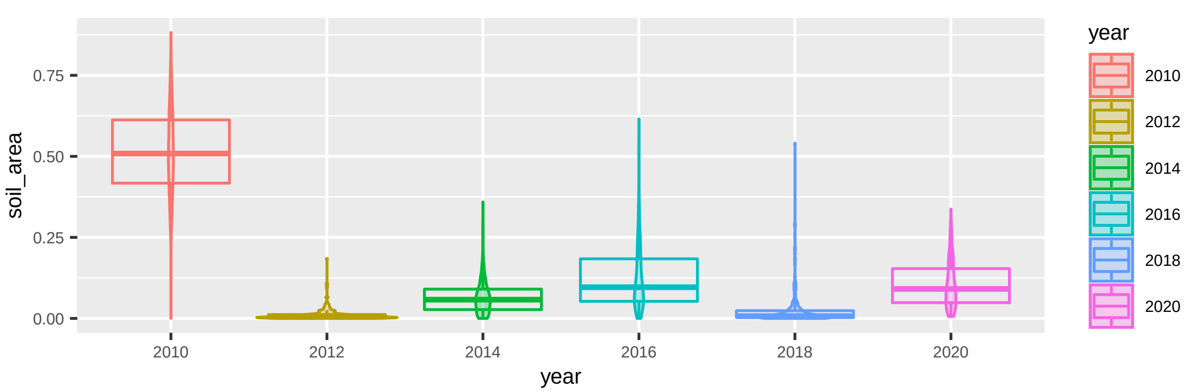
lst_day_mean
40
30
0.00
0.25
0.50
0.75
soil_area
year
2010
2012
2014
2016
2018
2020



soil_areaSample Quantiles
0.5
0.0
−2
0
2
Normal Theoretical Quantiles

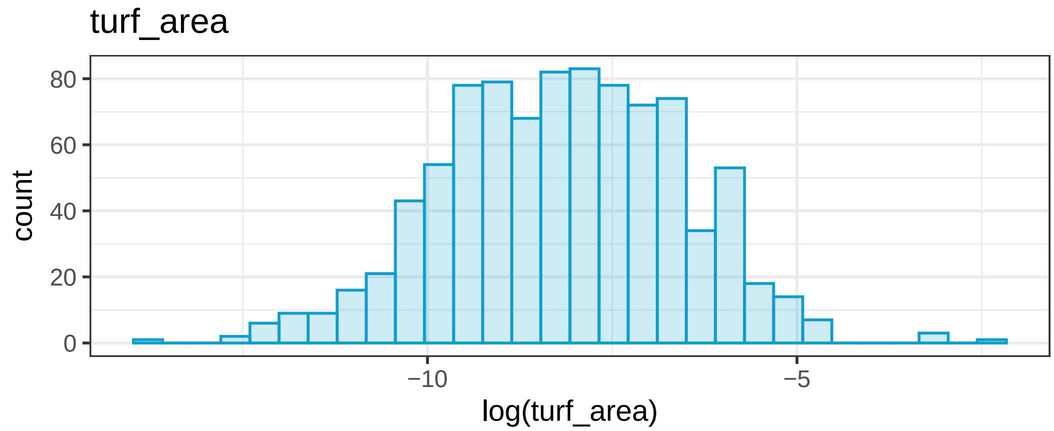


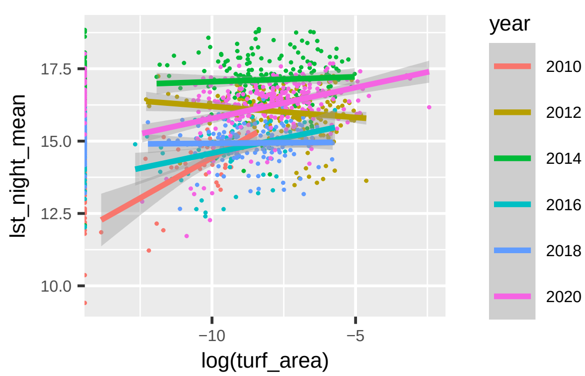
log(soil_area)Sample Quantiles
0
−5
−10
−2
0
2
Normal Theoretical Quantiles

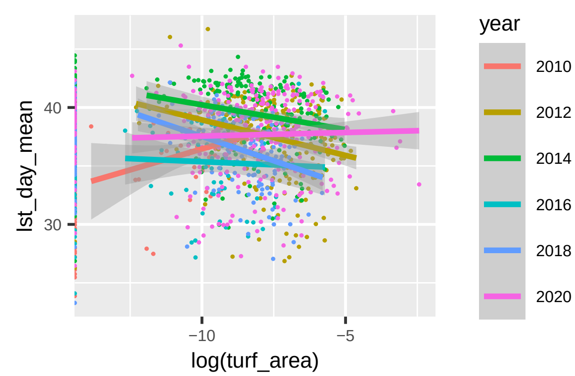


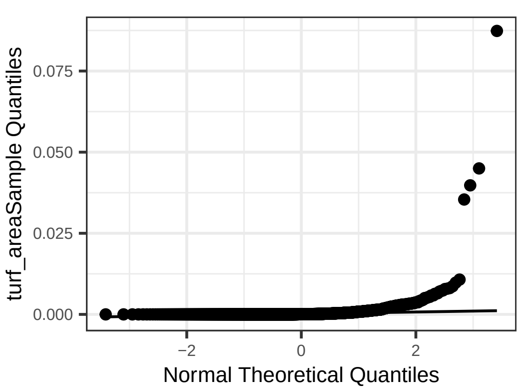
soil_area
0.75
0.50
0.25
0.00
2010
2012
2014
2016
2018
2020
year
year
2010
2012
2014
2016
2018
2020

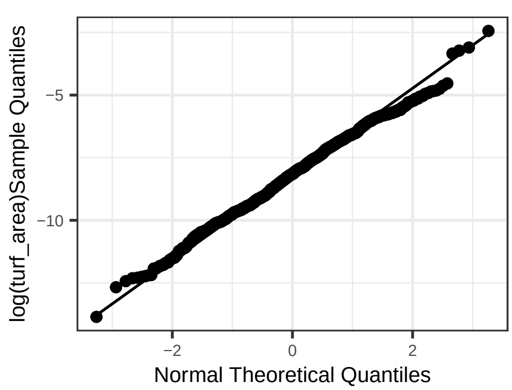
Turf Area

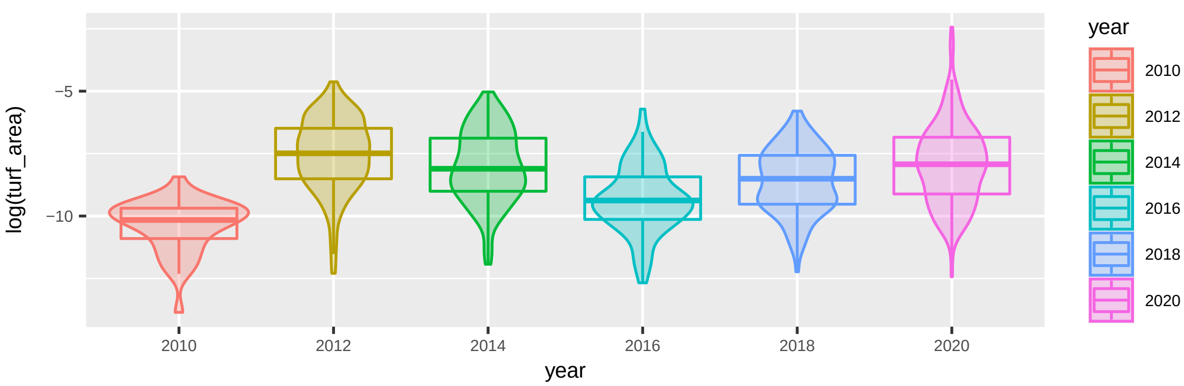

turf_area
count
80
60
40
20
0
−10
−5
log(turf_area)


lst_night_mean
17.5
15.0
12.5
10.0
−10
−5
log(turf_area)
year
2010
2012
2014
2016
2018
2020

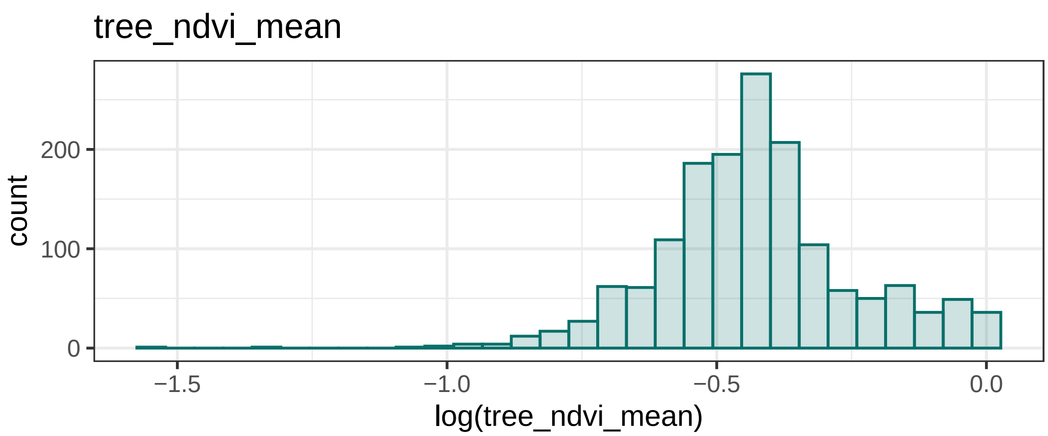

lst_day_mean
40
30
−10
−5
log(turf_area)
year
2010
2012
2014
2016
2018
2020

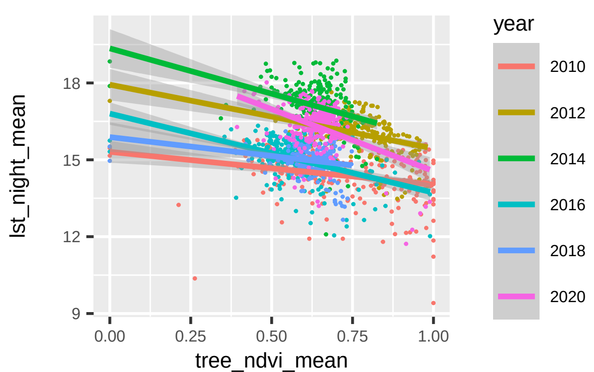

turf_areaSample Quantiles
0.075
0.050
0.025
0.000
−2
0
2
Normal Theoretical Quantiles

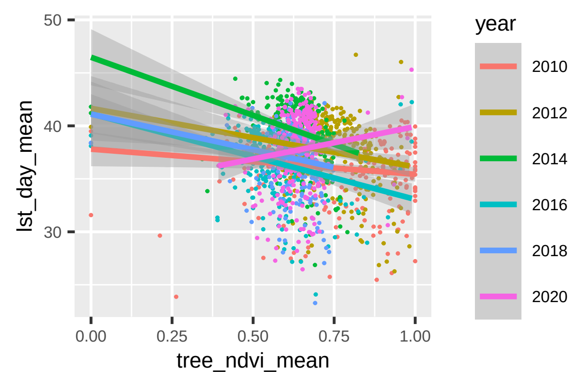

log(turf_area)Sample Quantiles
−5
−10
−2
0
2
Normal Theoretical Quantiles

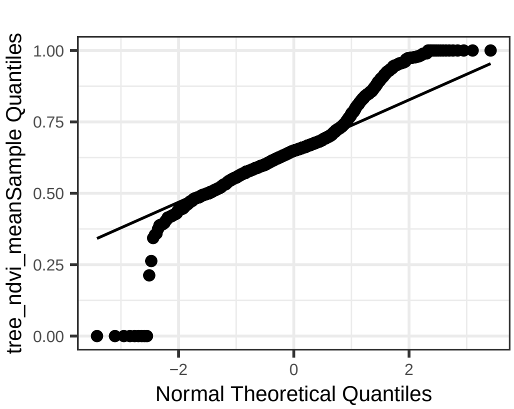

log(turf_area)
−5
−10
2010
2012
2014
2016
2018
2020
year
year
2010
2012
2014
2016
2018
2020

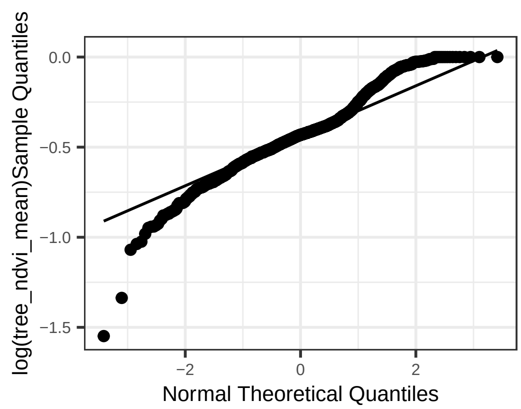
Tree NDVI Mean

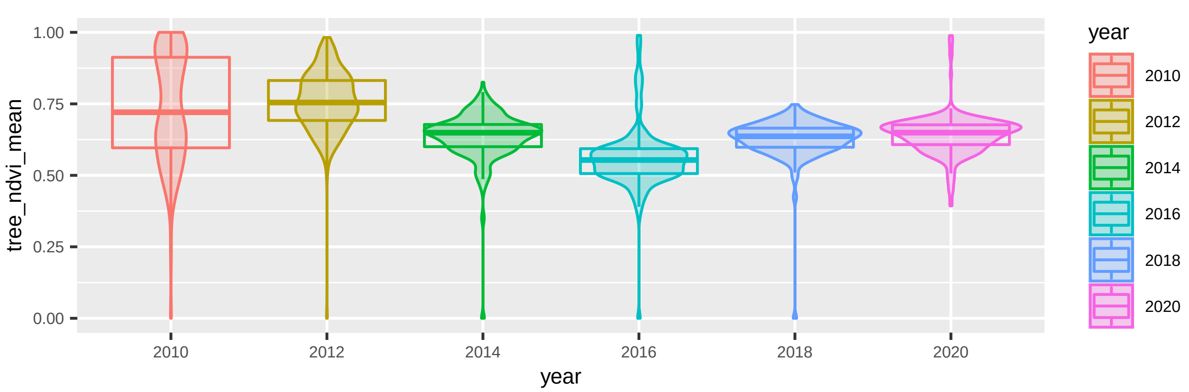






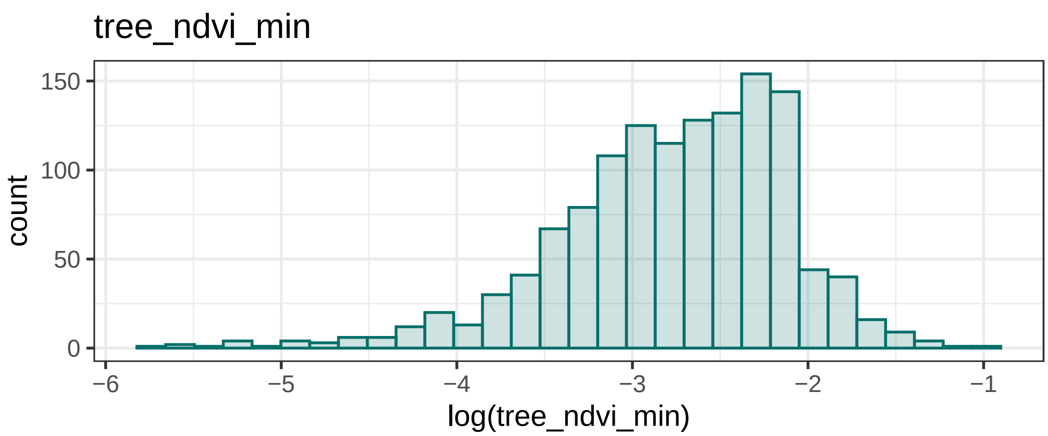


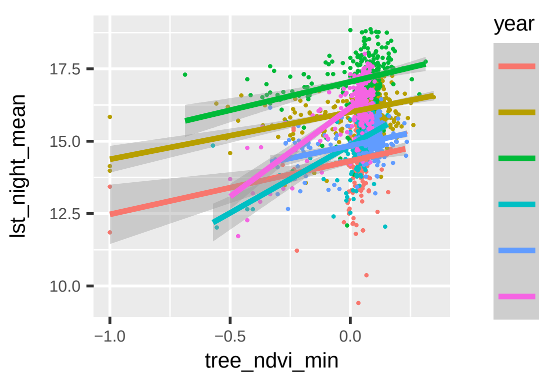


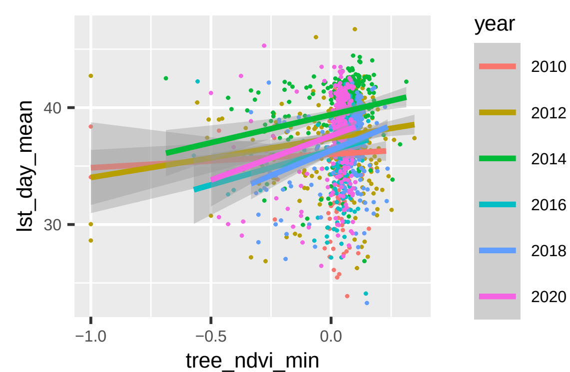


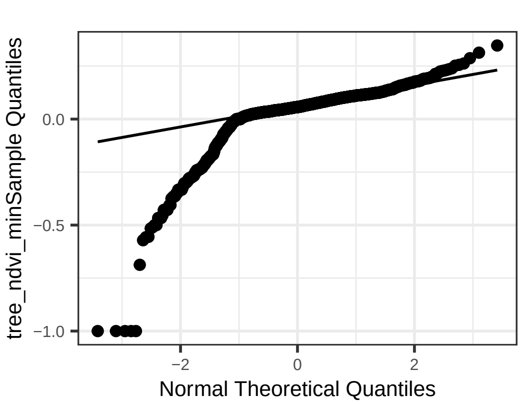
Google Earth: Tensor Flow Model

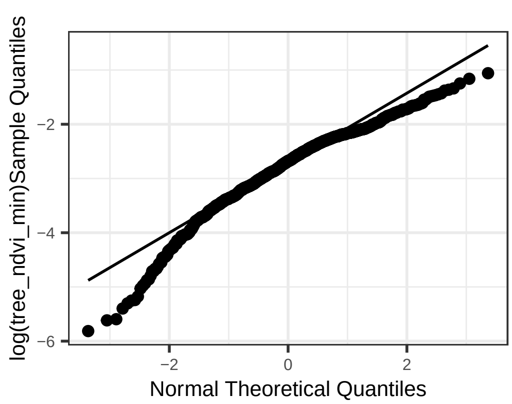
Tree NDVI Min

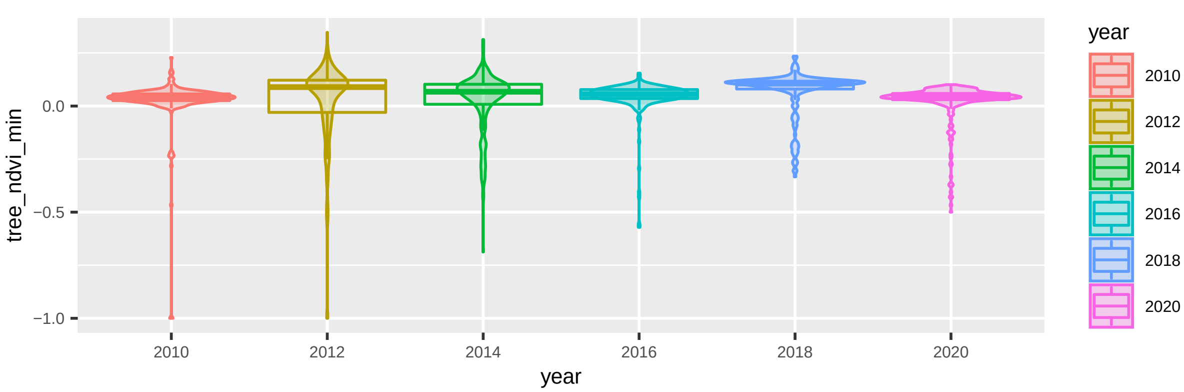




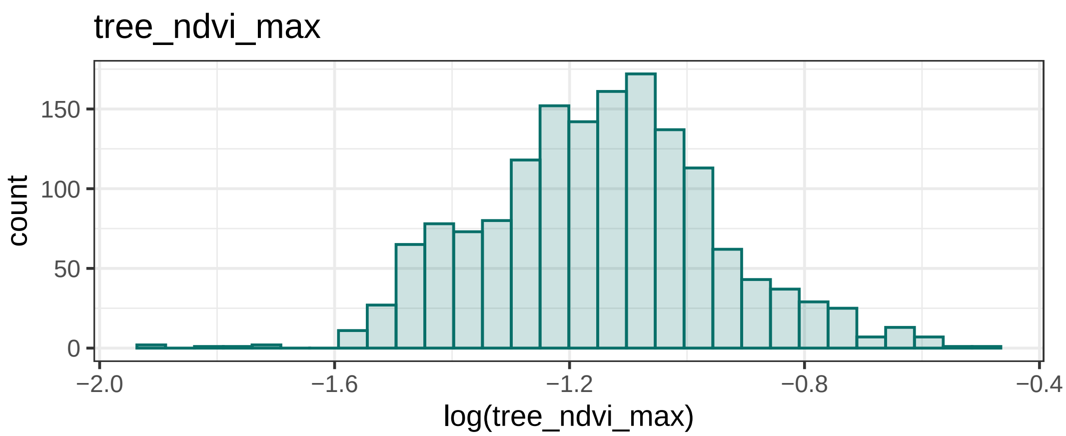


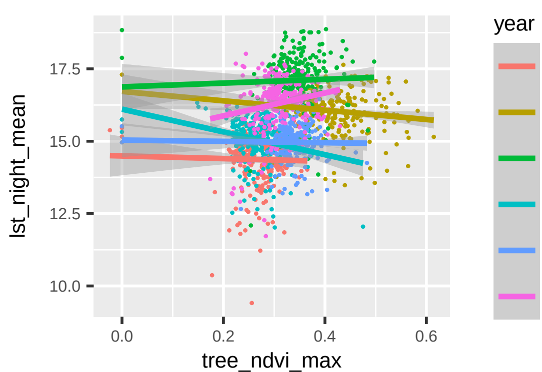


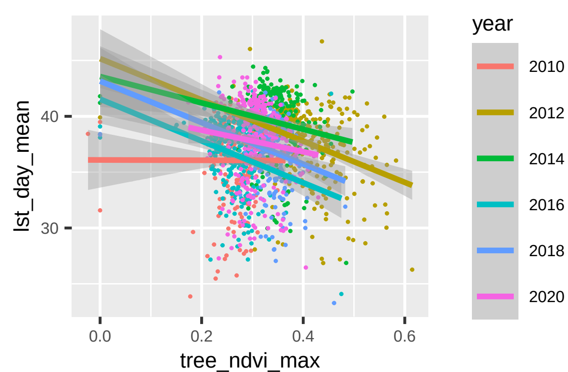


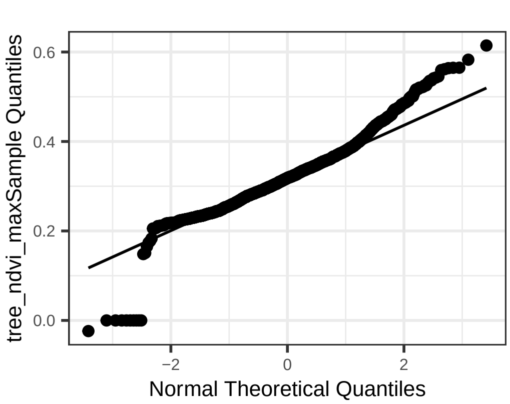


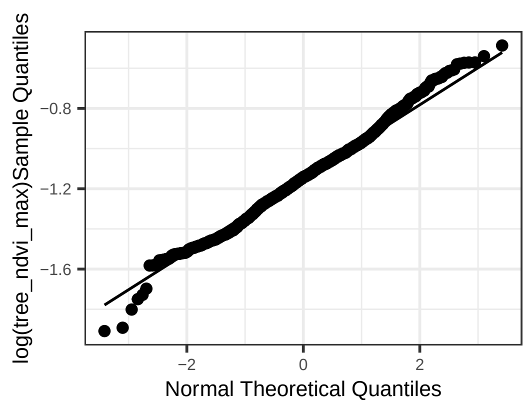
Tree NDVI Max

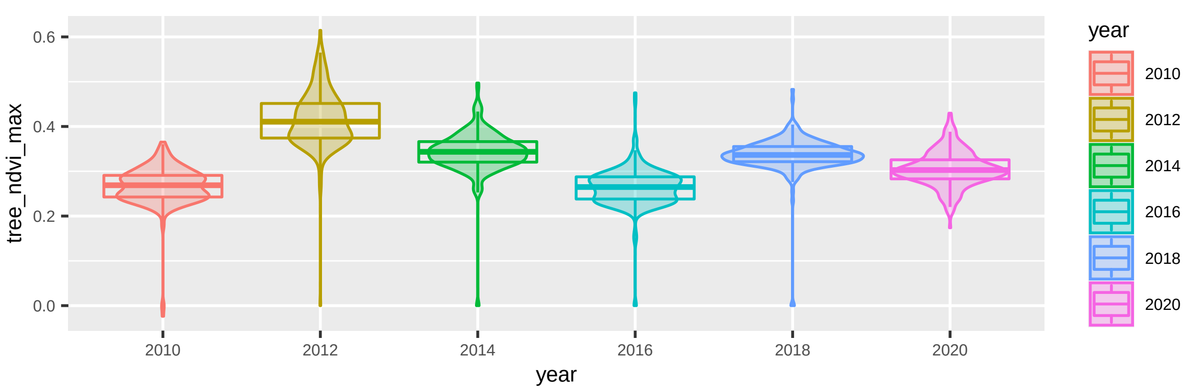




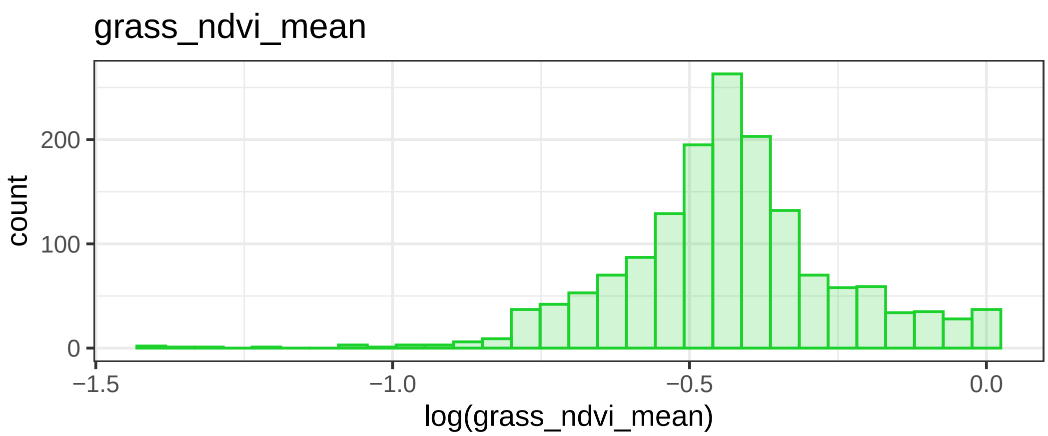


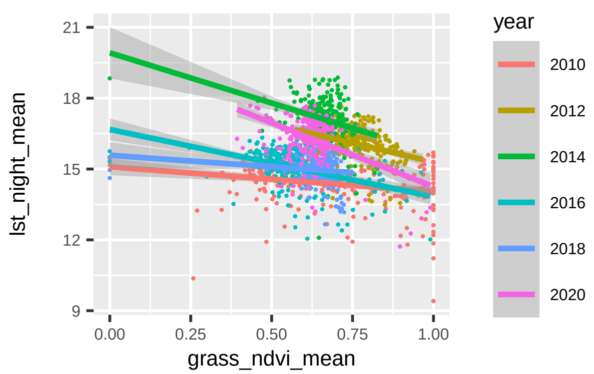


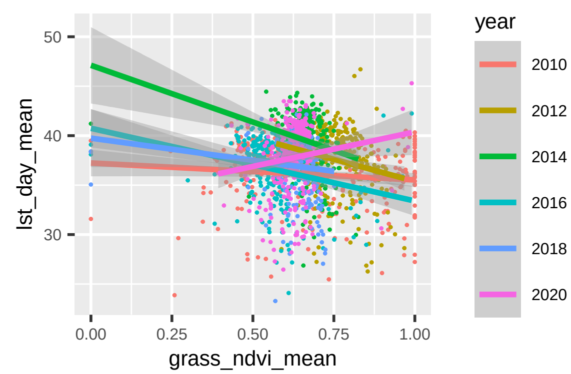


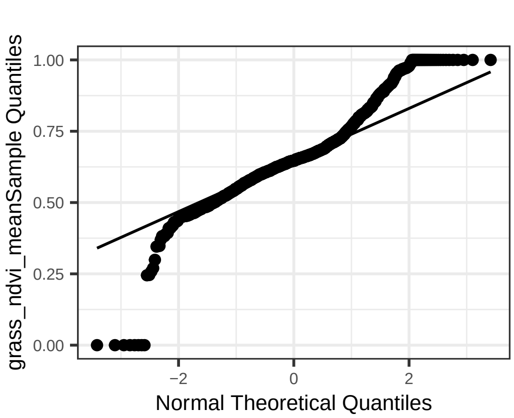


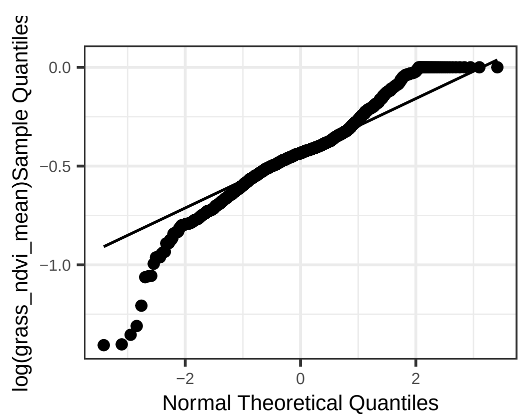
Lawn NDVI Mean

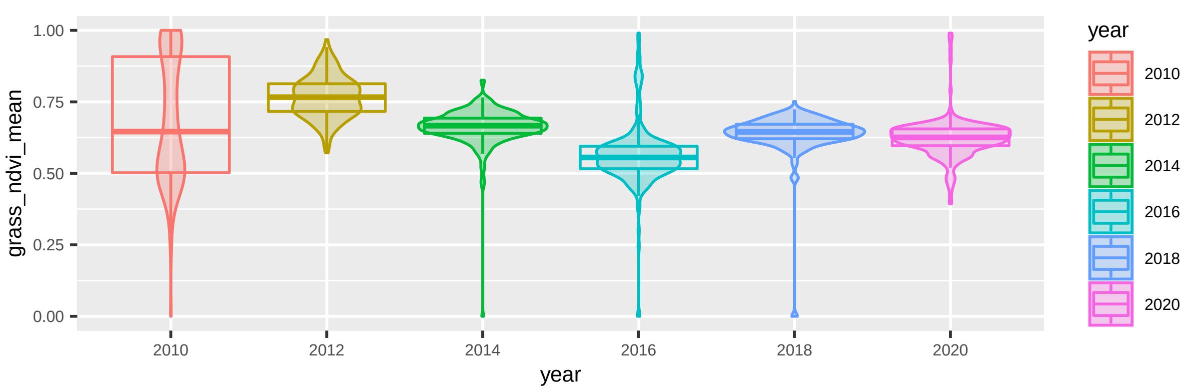




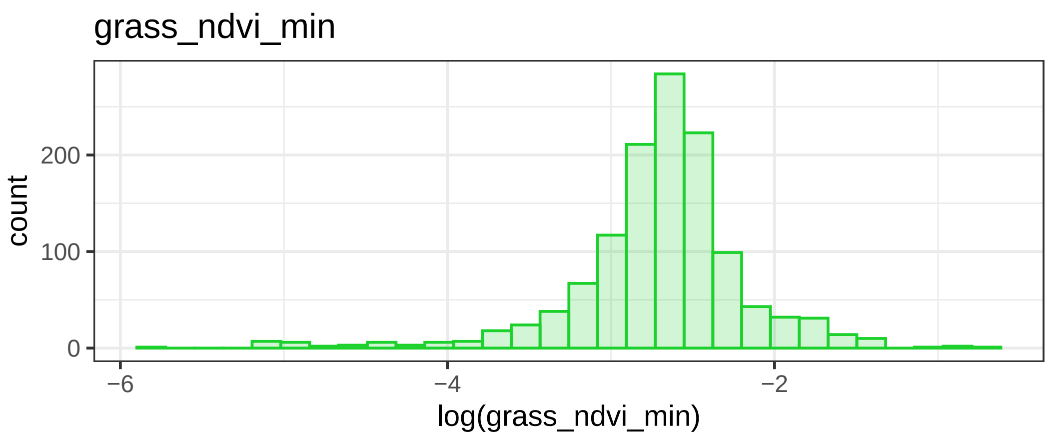


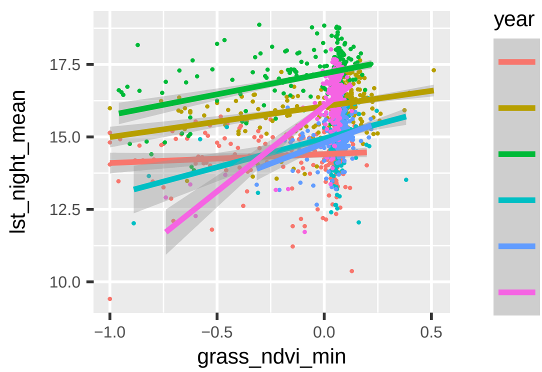


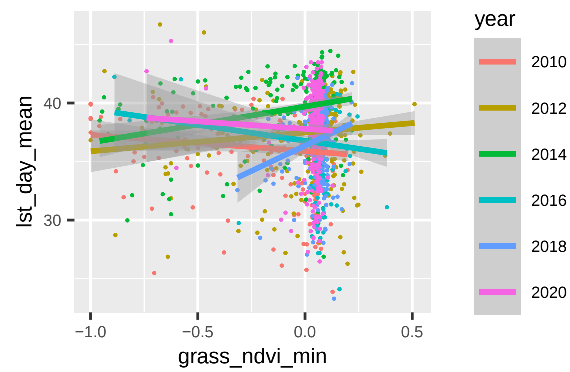


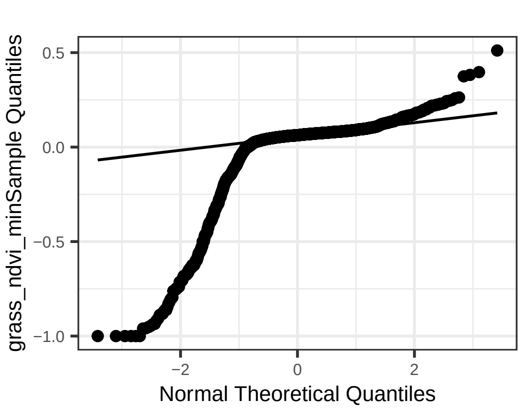


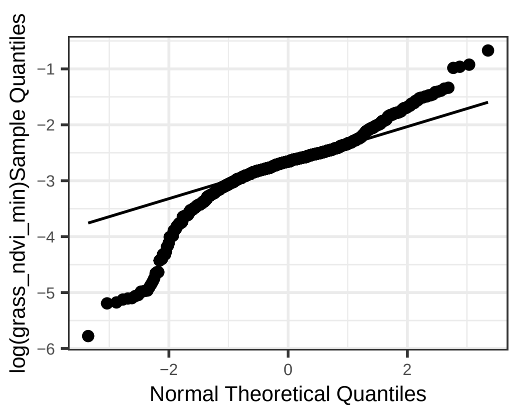
Lawn NDVI Min

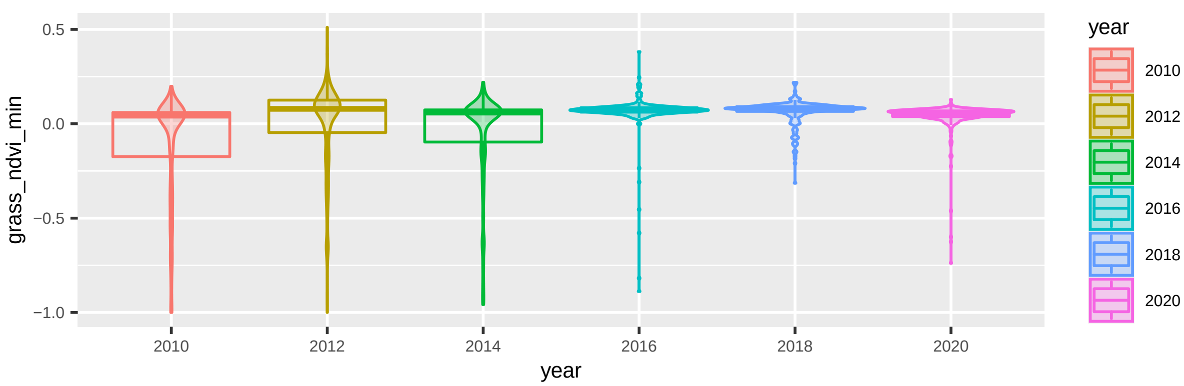




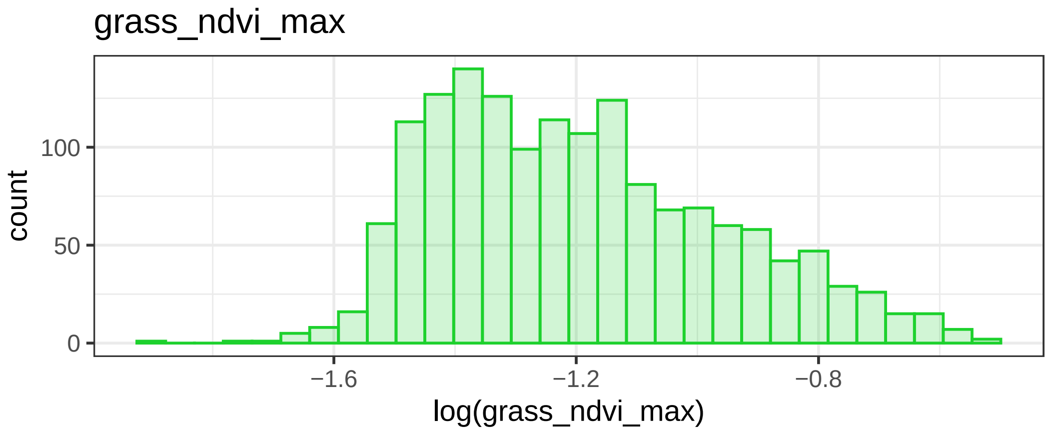


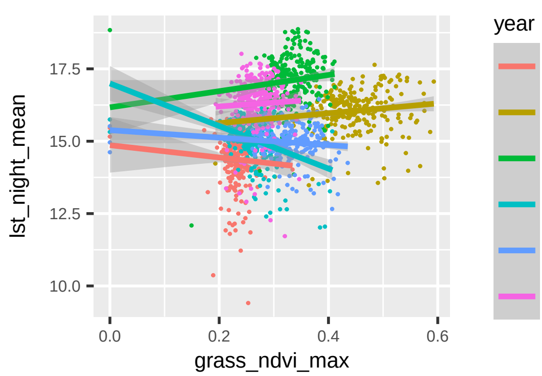


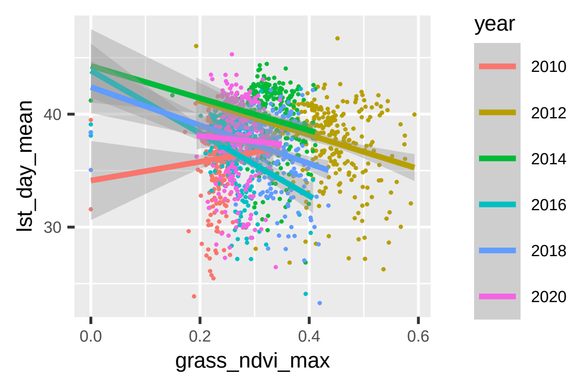


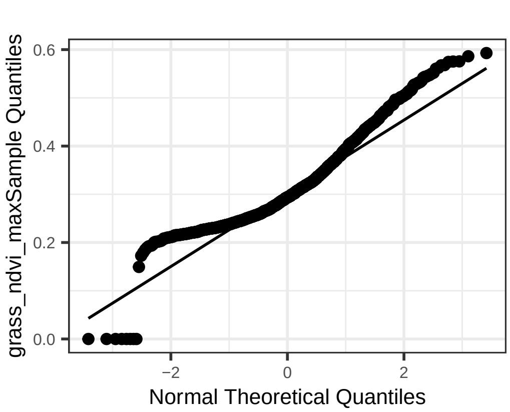


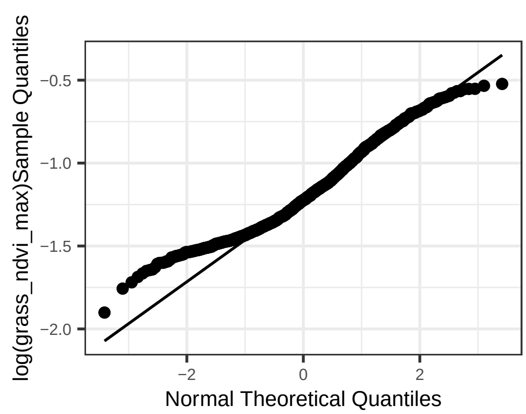
Lawn NDVI Max

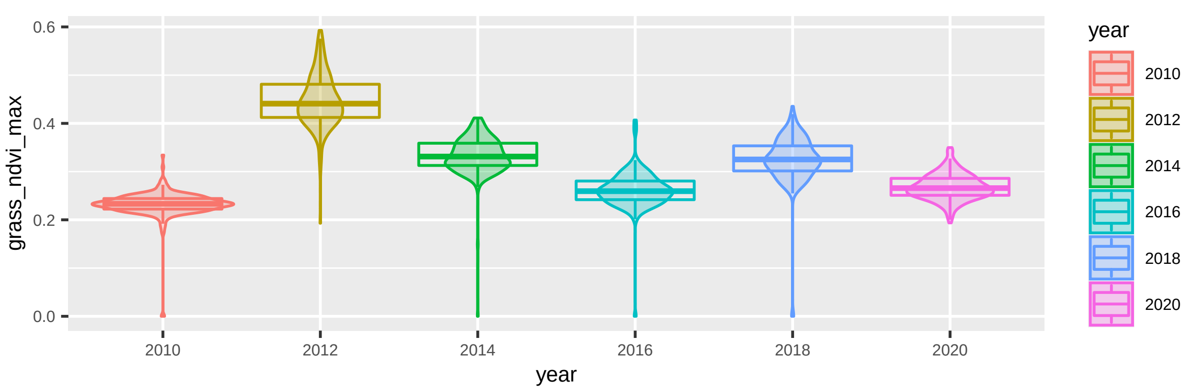




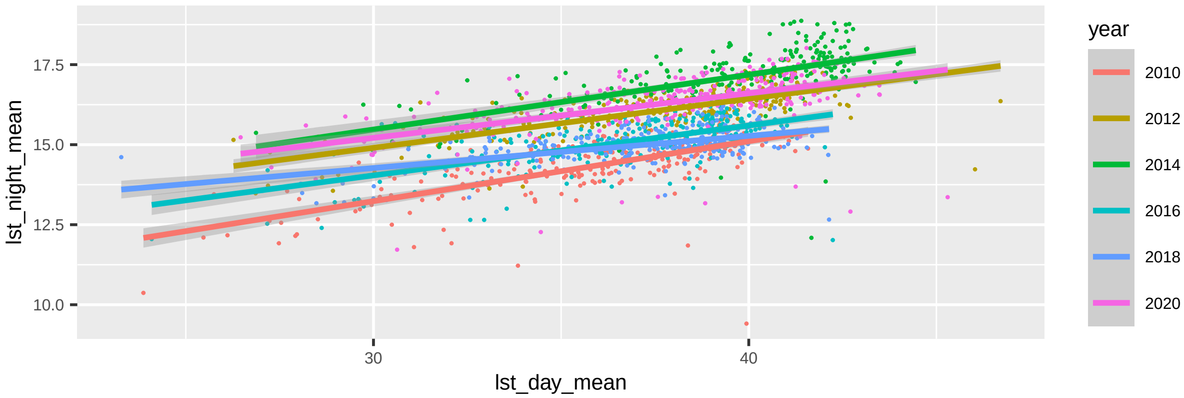


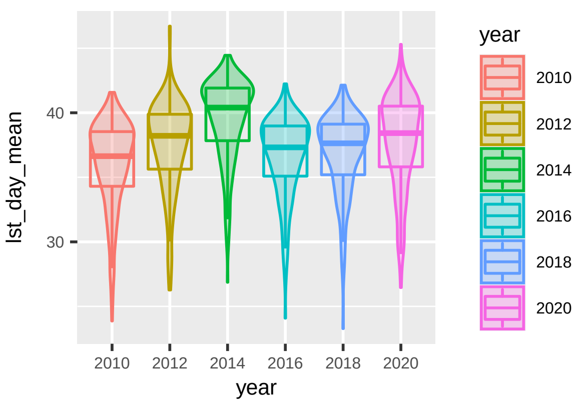


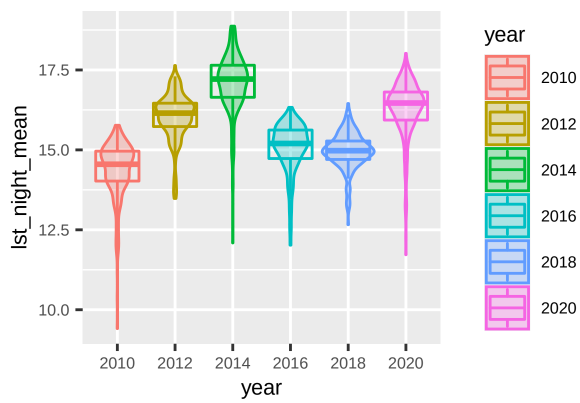


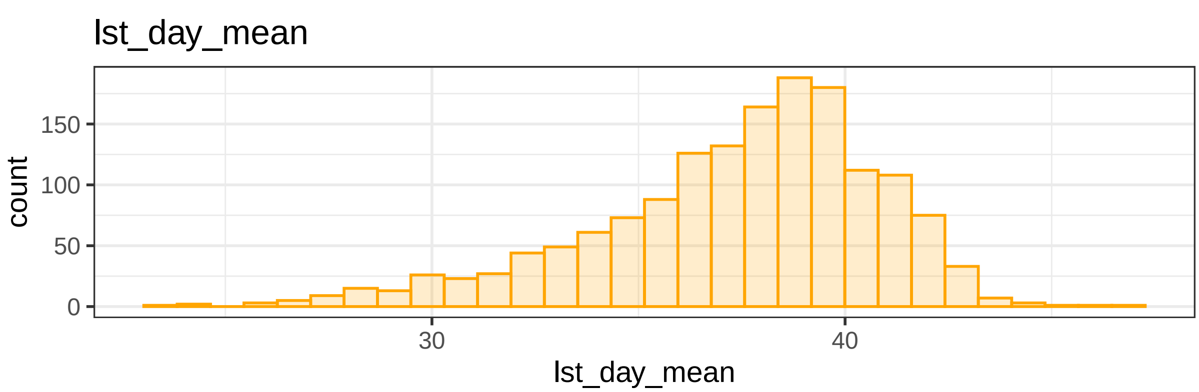


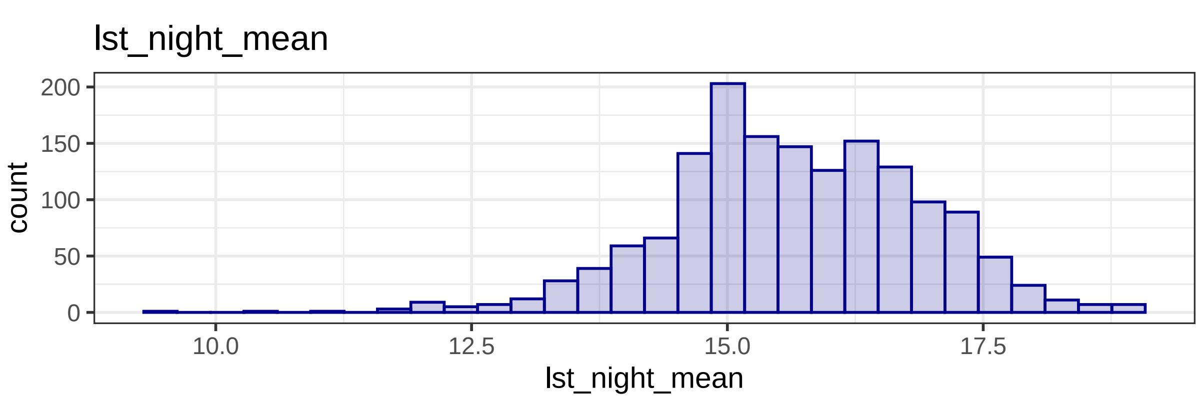
Mean Temperature











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 + so
    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("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 - 7:14:28 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

	Dependent variable:	
	log(lst_day_mean) OLSModel Day Temp	log(lst_night_mean) OLSModel Night Temp
	(1)	(2)
factor(year)2014	3.737^{***} (0.020)	2.922^{***} (0.011)
factor(year)2016	3.644^{***} (0.017)	2.795^{***} (0.009)
factor(year)2018	3.665^{***} (0.019)	2.784^{***} (0.010)
factor(year)2020	3.721^{***} (0.018)	2.901^{***} (0.010)
polygon_area	-0.000^{***} (0.000)	-0.000^{***} (0.000)
tree_area	-0.287^{***} (0.016)	-0.086^{***} (0.009)
lawn_area	-0.225^{***} (0.061)	-0.016 (0.034)
water_area	-2.345^{***} (0.757)	-0.845^{**} (0.419)
soil_area	0.188^{***} (0.032)	-0.070^{***} (0.018)
turf_area	-1.011 (0.650)	0.624^{*} (0.360)
grass_ndvi_mean	-0.041 (0.076)	-0.026 (0.042)
tree_ndvi_mean	0.012 (0.073)	-0.071^{*} (0.040)
Observations	1,057	1,057
R^2	1.000	1.000
Adjusted R^2	1.000	1.000
Residual Std. Error (df = 1045)	0.074	0.041
F Statistic (df = 12; 1045)	$212{,}012.000^{***}$	$401{,}183.300^{***}$
Note:		*p<0.1; **p<0.05; ***p<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 - 7:14:28 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	0.040*** (0.002)	0.037*** (0.003)
factor(year)2016	−0.042*** (0.002)	−0.086*** (0.002)
factor(year)2018	−0.030*** (0.002)	−0.092*** (0.003)
tree_area	−0.020** (0.008)	−0.021** (0.011)
lawn_area	−0.065*** (0.018)	0.056** (0.024)
water_area	−0.138 (0.195)	−0.281 (0.266)
soil_area	0.044*** (0.009)	0.020 (0.013)
turf_area	0.153 (0.237)	0.644** (0.323)
grass_ndvi_mean	0.010 (0.020)	0.011 (0.027)
tree_ndvi_mean	0.020 (0.018)	−0.018 (0.024)
Observations	1,057	1,057
R^2	0.852	0.914
Adjusted R^2	0.794	0.880
F Statistic (df = 10; 759)	437.403***	806.481***
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 = 87.811, df1 = 286, df2 = 759, 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 - 7:14:29 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 = 309.4, 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	0.136** (0.058)	0.372*** (0.074)
lawn_area	0.130 (0.122)	0.249 (0.158)
water_area	1.059 (1.706)	2.456 (2.226)
soil_area	0.300*** (0.068)	0.662*** (0.087)
turf_area	−1.738 (1.993)	−2.204 (2.550)
grass_ndvi_mean	1.136*** (0.152)	2.226*** (0.192)
tree_ndvi_mean	−0.097 (0.146)	−0.110 (0.190)
Observations	1,057	1,057
R^2	0.169	0.174
Adjusted R^2	0.164	0.170
F Statistic	310.425***	1,273.971***
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 - 7:14:30 AM

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

```
##
##  F test for individual effects
##
## data:  log(lst_day_mean) ~ 0 + factor(year) + tree_area + lawn_area +  ...
## F = 757.92, df1 = 3, df2 = 759, p-value < 2.2e-16
## alternative hypothesis: significant effects
```

Table 4:

	Dependent variable:
	log(lst_day_mean)
tree_area	0.011 (0.013)
lawn_area	0.449^{***} (0.028)
water_area	0.151 (0.388)
soil_area	0.022 (0.016)
turf_area	0.534 (0.455)
grass_ndvi_mean	0.261^{***} (0.035)
tree_ndvi_mean	-0.132^{***} (0.033)
Observations	1,057
R^2	0.409
Adjusted R^2	0.181
F Statistic	75.386^{***} (df = 7; 762)
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 = 2.003, df = 1, p-value = 0.157
## 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 = 77.258, df = 1, p-value < 2.2e-16
## alternative hypothesis: serial correlation in idiosyncratic errors
```

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

```
##
## t test of coefficients:
##
##                    Estimate Std. Error  t value  Pr(>|t|)
## factor(year)2014  0.0396623  0.0024143  16.4278 < 2.2e-16 ***
```

```
## factor(year)2016 -0.0415958  0.0017898 -23.2410 < 2.2e-16 ***
## factor(year)2018 -0.0303815  0.0021367 -14.2191 < 2.2e-16 ***
## tree_area        -0.0199003  0.0085040  -2.3401 0.0195358 *
## lawn_area        -0.0651138  0.0185238  -3.5151 0.0004656 ***
## water_area       -0.1378498  0.1827426  -0.7543 0.4508799
## soil_area         0.0439236  0.0092062   4.7711 2.198e-06 ***
## turf_area         0.1532339  0.1796727   0.8528 0.3940116
## grass_ndvi_mean   0.0103740  0.0183325   0.5659 0.5716437
## tree_ndvi_mean    0.0201446  0.0166105   1.2128 0.2255979
## ---
## 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  0.0396623  0.0024143  16.4278 < 2.2e-16 ***
## factor(year)2016 -0.0415958  0.0017898 -23.2410 < 2.2e-16 ***
## factor(year)2018 -0.0303815  0.0021367 -14.2191 < 2.2e-16 ***
## tree_area        -0.0199003  0.0085040  -2.3401 0.0195358 *
## lawn_area        -0.0651138  0.0185238  -3.5151 0.0004656 ***
## water_area       -0.1378498  0.1827426  -0.7543 0.4508799
## soil_area         0.0439236  0.0092062   4.7711 2.198e-06 ***
## turf_area         0.1532339  0.1796727   0.8528 0.3940116
## grass_ndvi_mean   0.0103740  0.0183325   0.5659 0.5716437
## tree_ndvi_mean    0.0201446  0.0166105   1.2128 0.2255979
## ---
## 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 = "are
```

```
##     factor(year)2014 factor(year)2016 factor(year)2018   tree_area  lawn_area
## HC0      0.002414344      0.001789762      0.002136662 0.008503966 0.01852378
## HC1      0.002425846      0.001798288      0.002146841 0.008544480 0.01861203
## HC2      0.002437746      0.001804833      0.002154909 0.008623311 0.01873911
## HC3      0.002462205      0.001820519      0.002173754 0.008747263 0.01895938
## HC4      0.002487928      0.001833629      0.002189318 0.008930426 0.01919561
##     water_area   soil_area turf_area grass_ndvi_mean tree_ndvi_mean
## HC0  0.1827426 0.009206210 0.1796727      0.01833250     0.01661052
## HC1  0.1836132 0.009250070 0.1805287      0.01841984     0.01668965
## HC2  0.2505347 0.009303731 0.1963739      0.01849230     0.01671859
## HC3  0.3611753 0.009409391 0.2159644      0.01865703     0.01682802
## HC4  0.8256154 0.009550660 0.2653934      0.01880047     0.01685394
```

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

```r
print(totalRobust)
```

```
##
## t test of coefficients:
##
##                     Estimate Std. Error  t value  Pr(>|t|)
## factor(year)2014  0.0396623  0.0024143  16.4278 < 2.2e-16 ***
## factor(year)2016 -0.0415958  0.0017898 -23.2410 < 2.2e-16 ***
## factor(year)2018 -0.0303815  0.0021367 -14.2191 < 2.2e-16 ***
## tree_area        -0.0199003  0.0085040  -2.3401 0.0195358 *
## lawn_area        -0.0651138  0.0185238  -3.5151 0.0004656 ***
## water_area       -0.1378498  0.1827426  -0.7543 0.4508799
## soil_area         0.0439236  0.0092062   4.7711 2.198e-06 ***
## turf_area         0.1532339  0.1796727   0.8528 0.3940116
## grass_ndvi_mean   0.0103740  0.0183325   0.5659 0.5716437
## tree_ndvi_mean    0.0201446  0.0166105   1.2128 0.2255979
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
print(cInterval)
```

```
##                        2.5 %       97.5 %
## factor(year)2014  0.03492275  0.044401916
## factor(year)2016 -0.04510929 -0.038082352
## factor(year)2018 -0.03457594 -0.026186997
## tree_area        -0.03659438 -0.003206206
## lawn_area        -0.10147775 -0.028749877
## water_area       -0.49659084  0.220891216
## soil_area         0.02585094  0.061996255
## turf_area        -0.19948064  0.505948344
## grass_ndvi_mean  -0.02561446  0.046362385
## tree_ndvi_mean   -0.01246340  0.052752636
```

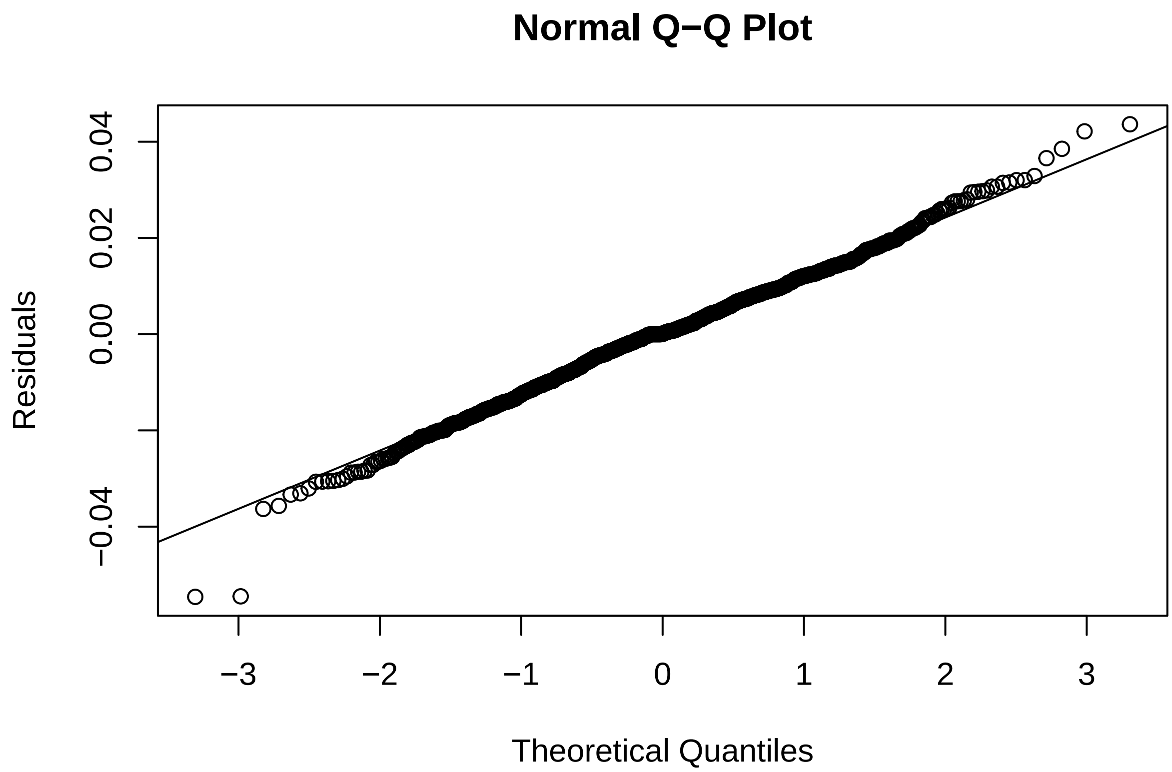
```r
stargazer::stargazer(OLSM1,feM1,reM1,feM2FixedTime,
                     font.size = 'tiny',
                     title = 'Regression Models Sumamry',
                     column.labels = c("OLS", "FixedEffects",
                                       "RandomEffects",
                                       "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 - 7:14:32 AM

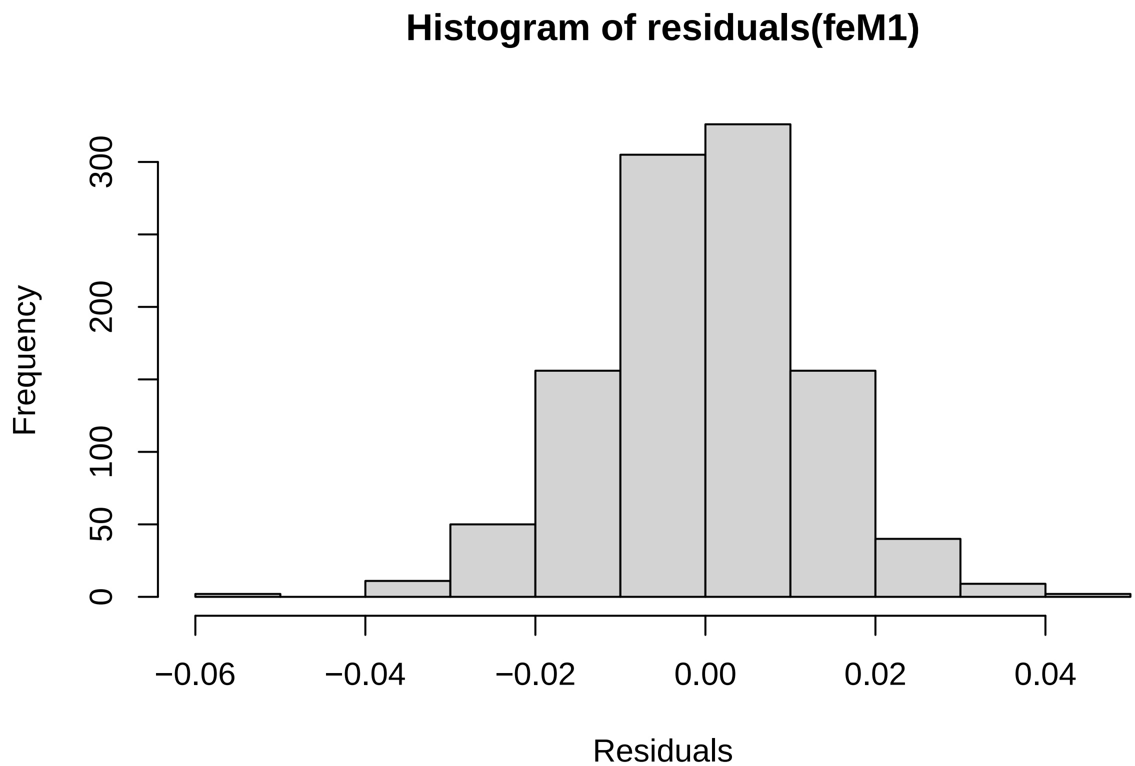
Table 5: Regression Models Sumamry

	Dependent variable:			
	log(lst_day_mean)			
	OLS		*panel linear*	
	OLS	FixedEffects	RandomEffects	FixedEffectsFixedTime
	(1)	(2)	(3)	(4)
factor(year)2014	3.737***	0.040***		
	(0.020)	(0.002)		
factor(year)2016	3.644***	−0.042***		
	(0.017)	(0.002)		
factor(year)2018	3.665***	−0.030***		
	(0.019)	(0.002)		
factor(year)2020	3.721***			
	(0.018)			
polygon_area	−0.000***			
	(0.000)			
tree_area	−0.287***	−0.020**	0.136**	0.011
	(0.016)	(0.008)	(0.058)	(0.013)
lawn_area	−0.225***	−0.065***	0.130	0.449***
	(0.061)	(0.018)	(0.122)	(0.028)
water_area	−2.345***	−0.138	1.059	0.151
	(0.757)	(0.195)	(1.706)	(0.388)
soil_area	0.188***	0.044***	0.300***	0.022
	(0.032)	(0.009)	(0.068)	(0.016)
turf_area	−1.011	0.153	−1.738	0.534
	(0.650)	(0.237)	(1.993)	(0.455)
grass_ndvi_mean	−0.041	0.010	1.136***	0.261***
	(0.076)	(0.020)	(0.152)	(0.035)
tree_ndvi_mean	0.012	0.020	−0.097	−0.132***
	(0.073)	(0.018)	(0.146)	(0.033)
Observations	1,057	1,057	1,057	1,057
R^2	1.000	0.852	0.169	0.409
Adjusted R^2	1.000	0.794	0.164	0.181
Residual Std. Error	0.074 (df = 1045)			
F Statistic	212,012.000*** (df = 12; 1045)	437.403*** (df = 10; 759)	310.425***	75.386*** (df = 7; 762)
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
Residuals