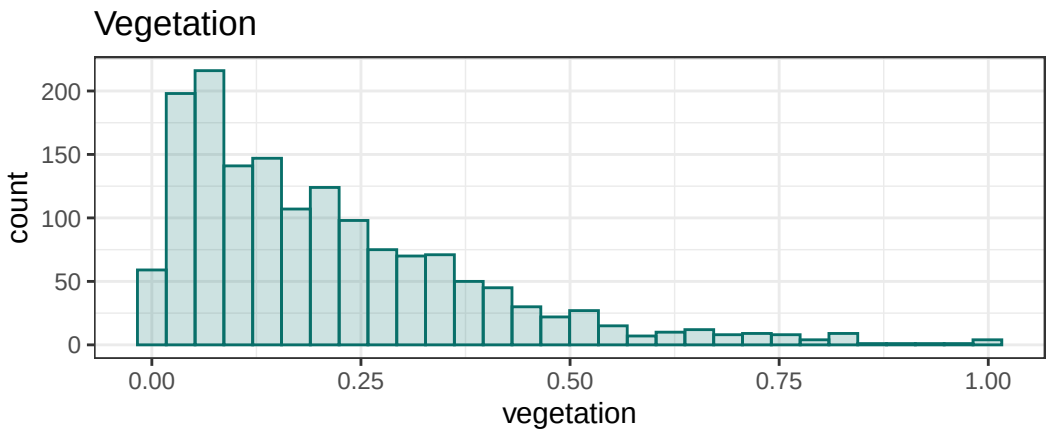
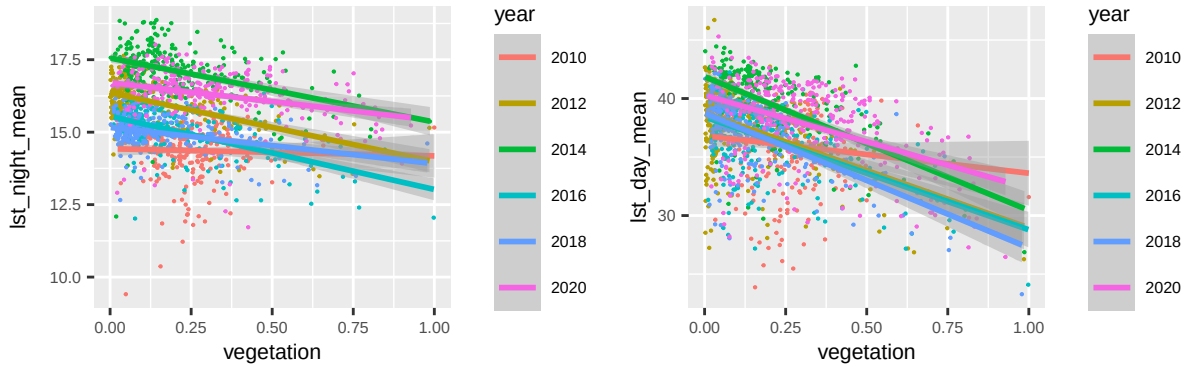
LawnProject Panel Data Analysis / Tensor Flow Data

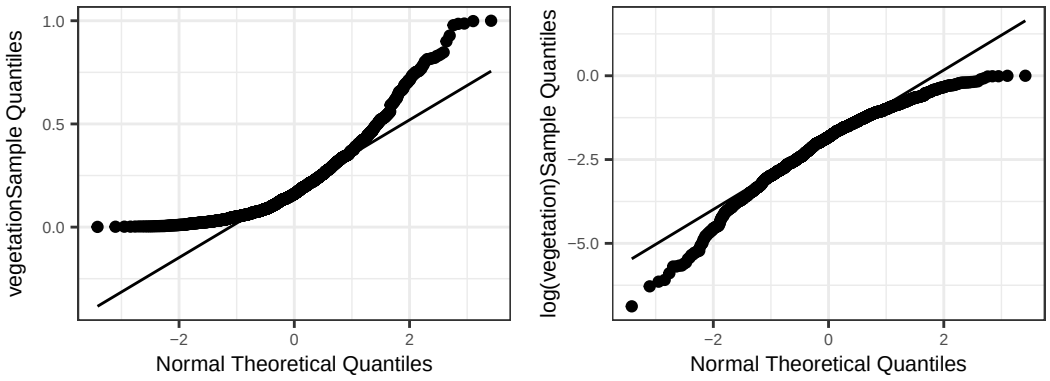
LawnProject Team

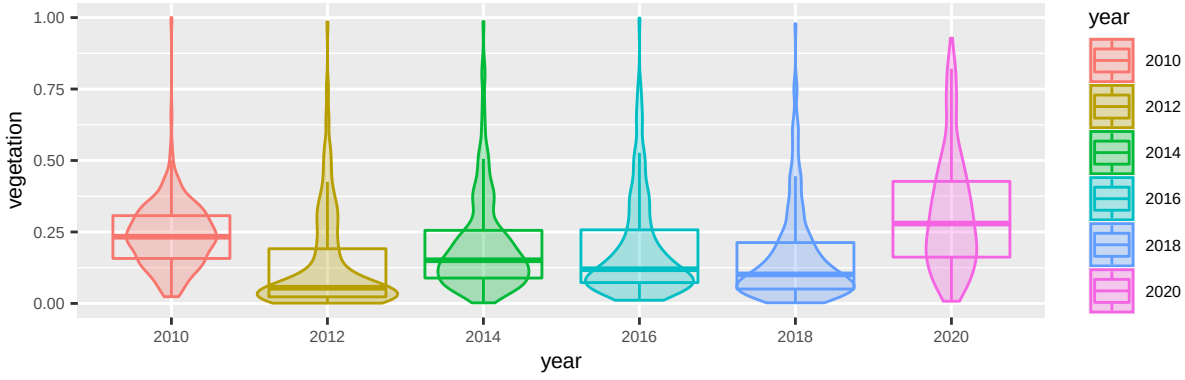
Intro

Aggregate Vegetation Area (Trees + Lawn)



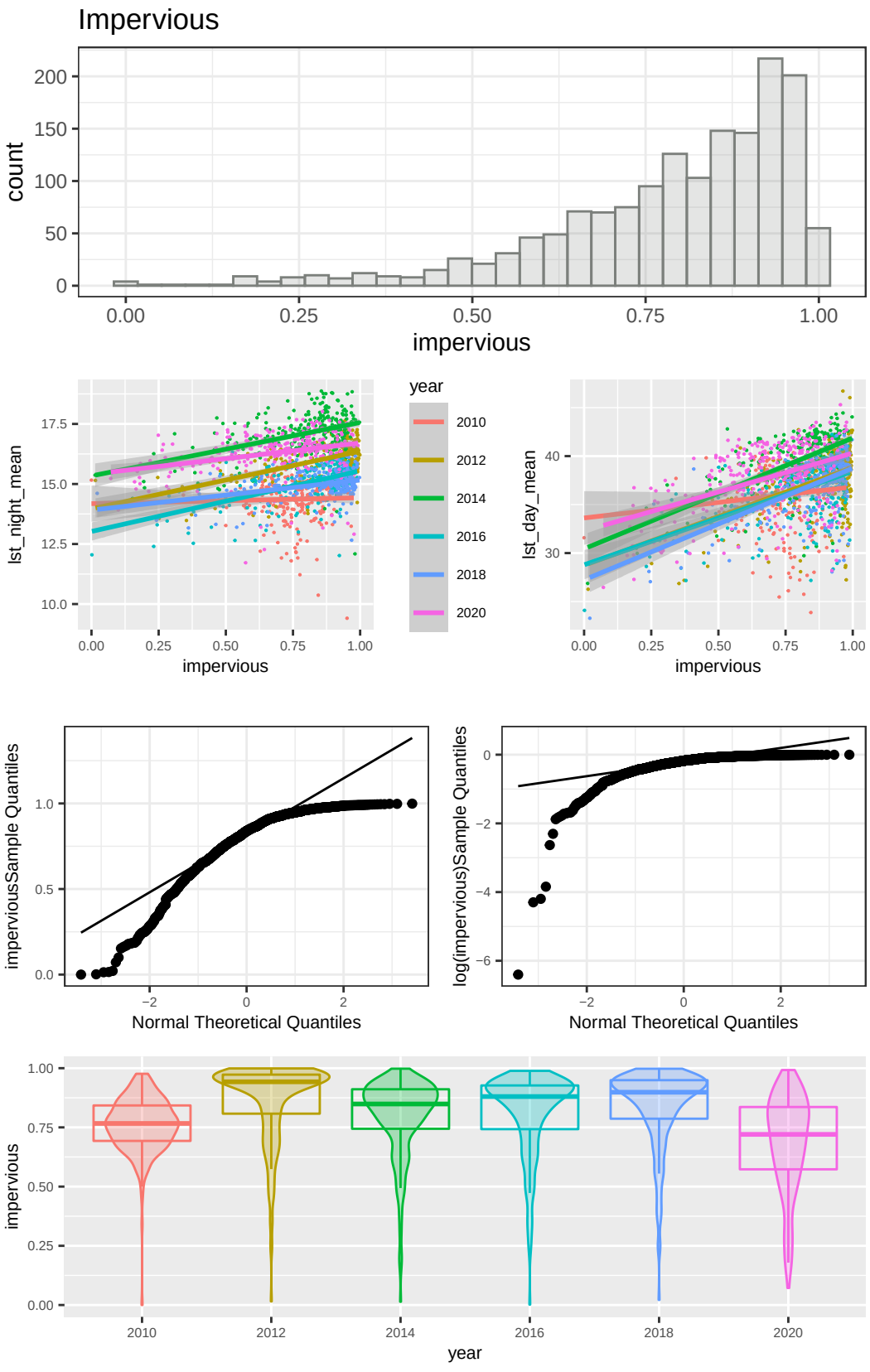






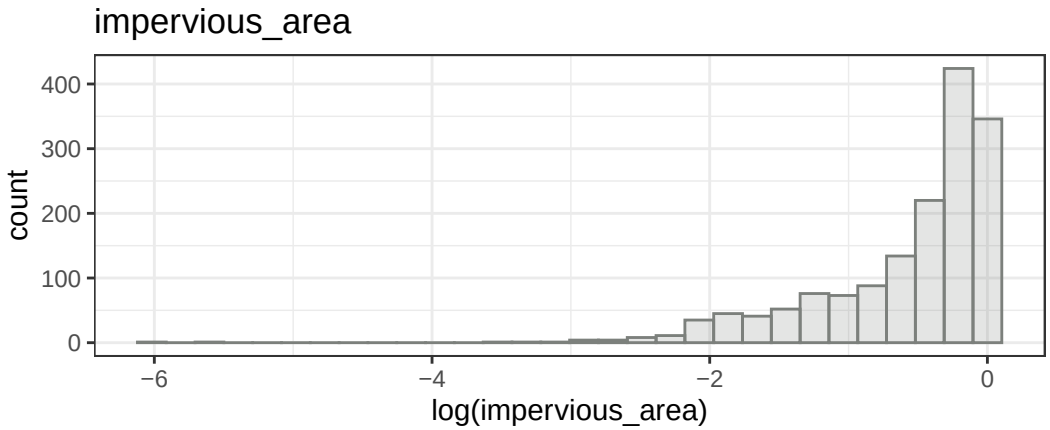
Google Earth: Tensor Flow Model

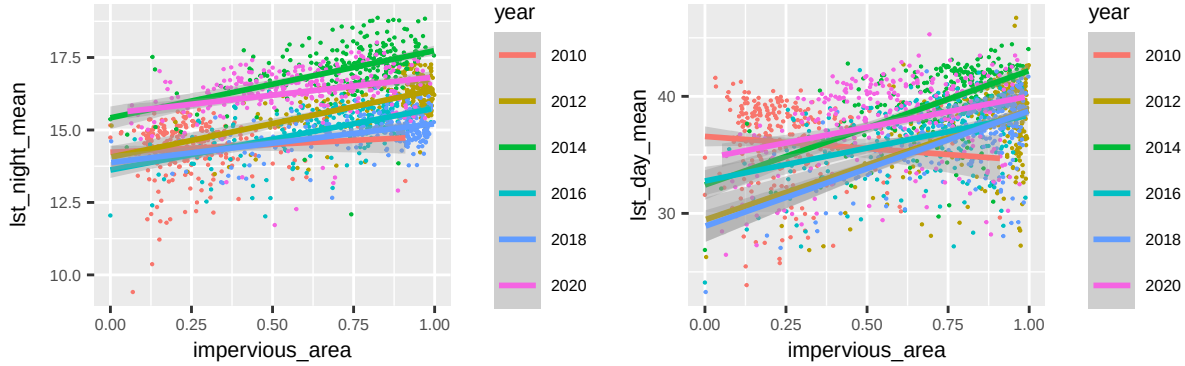
Aggregate Impervious Area (Soil + Turf + Impervious)

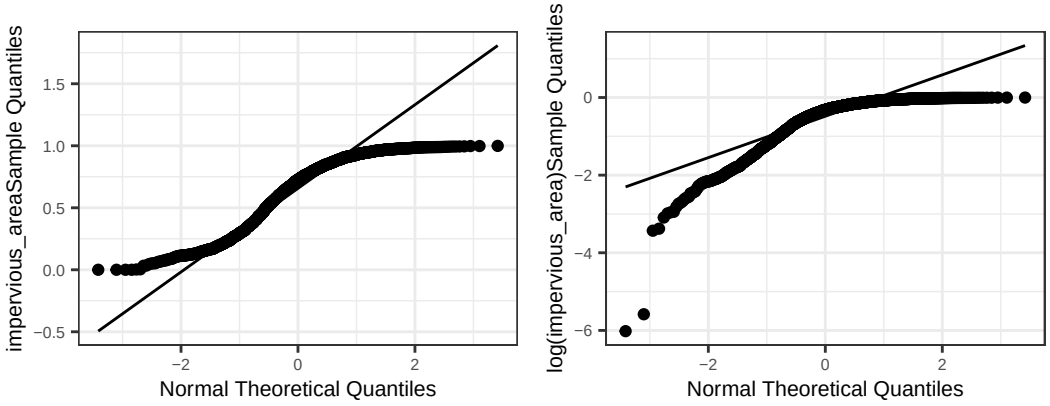


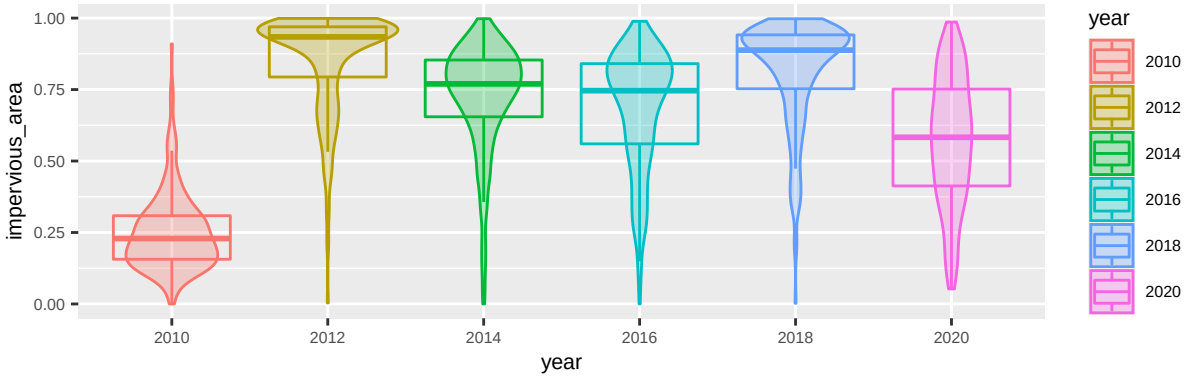
Google Earth: Tensor Flow Model

Impervious Area



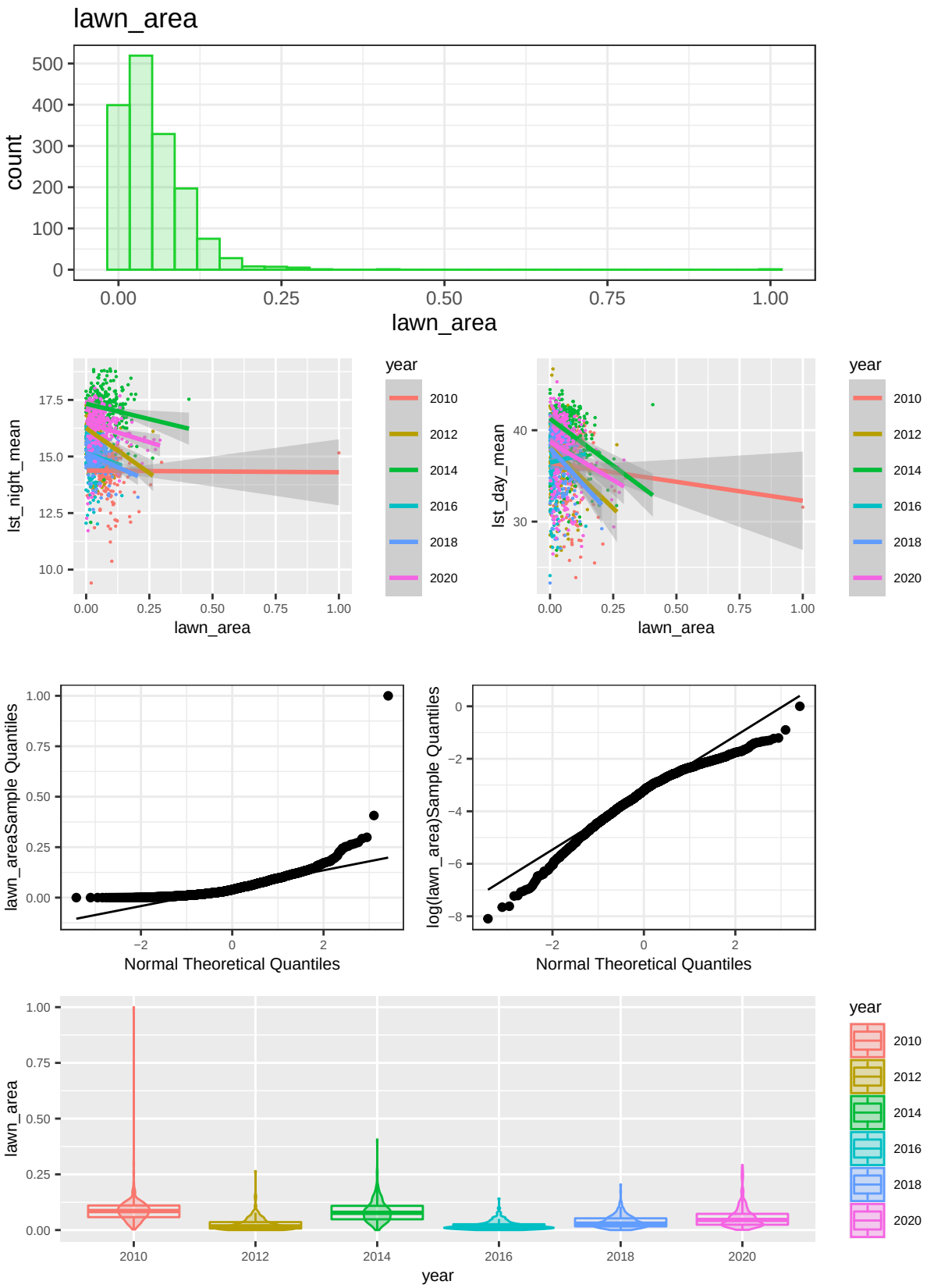






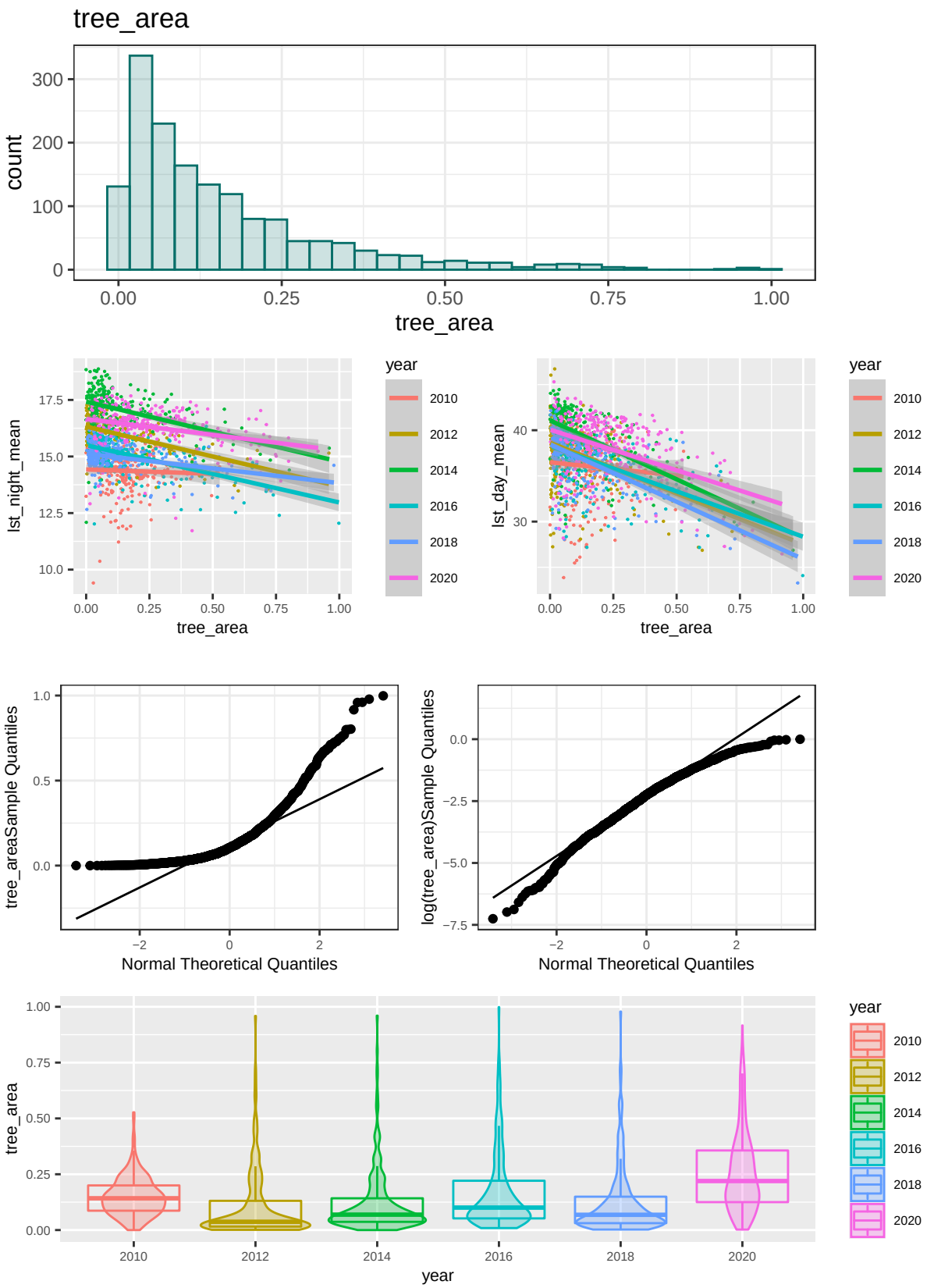
Google Earth: Tensor Flow Model

Lawn Area



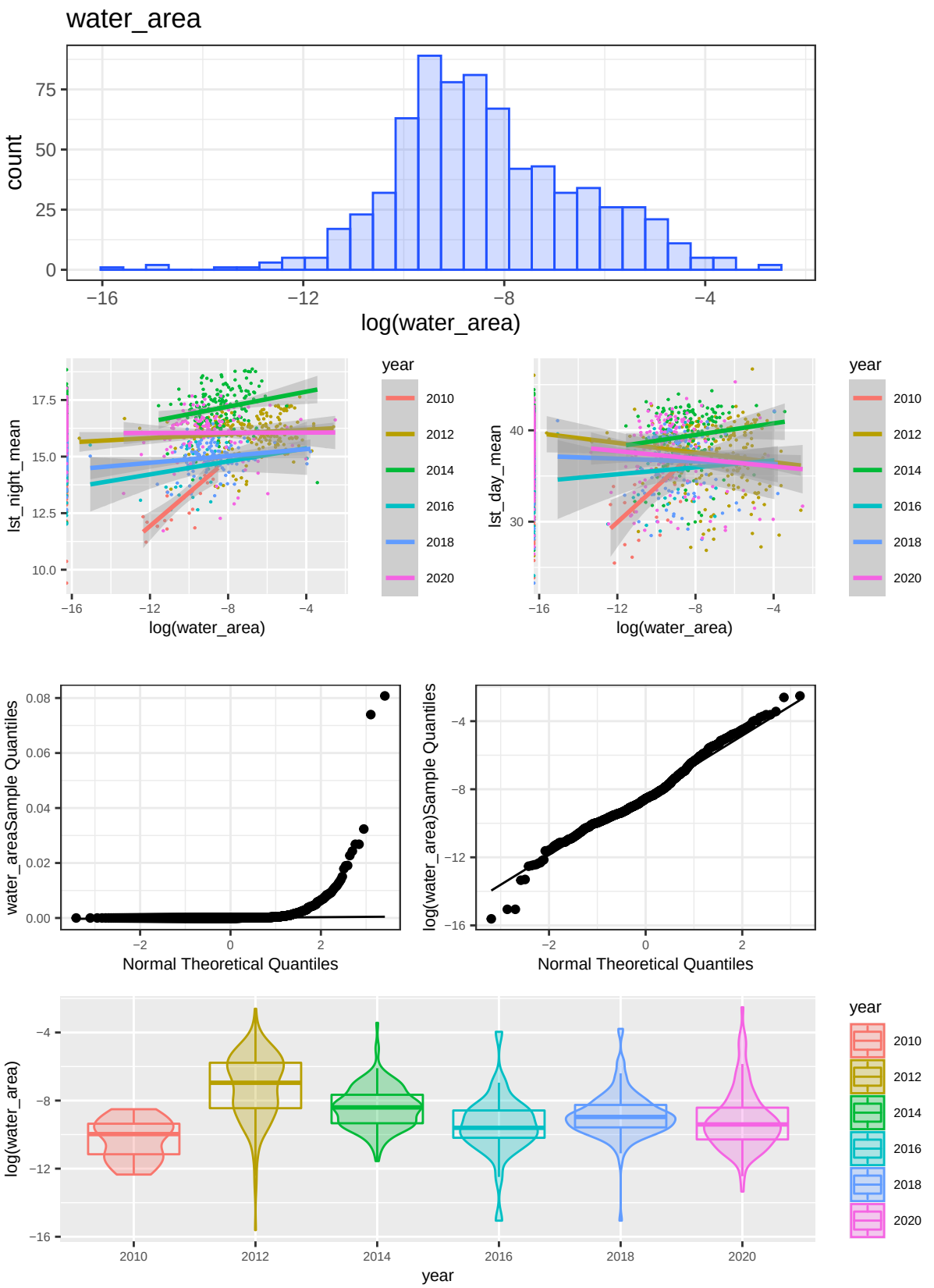
Google Earth: Tensor Flow Model

Tree Area



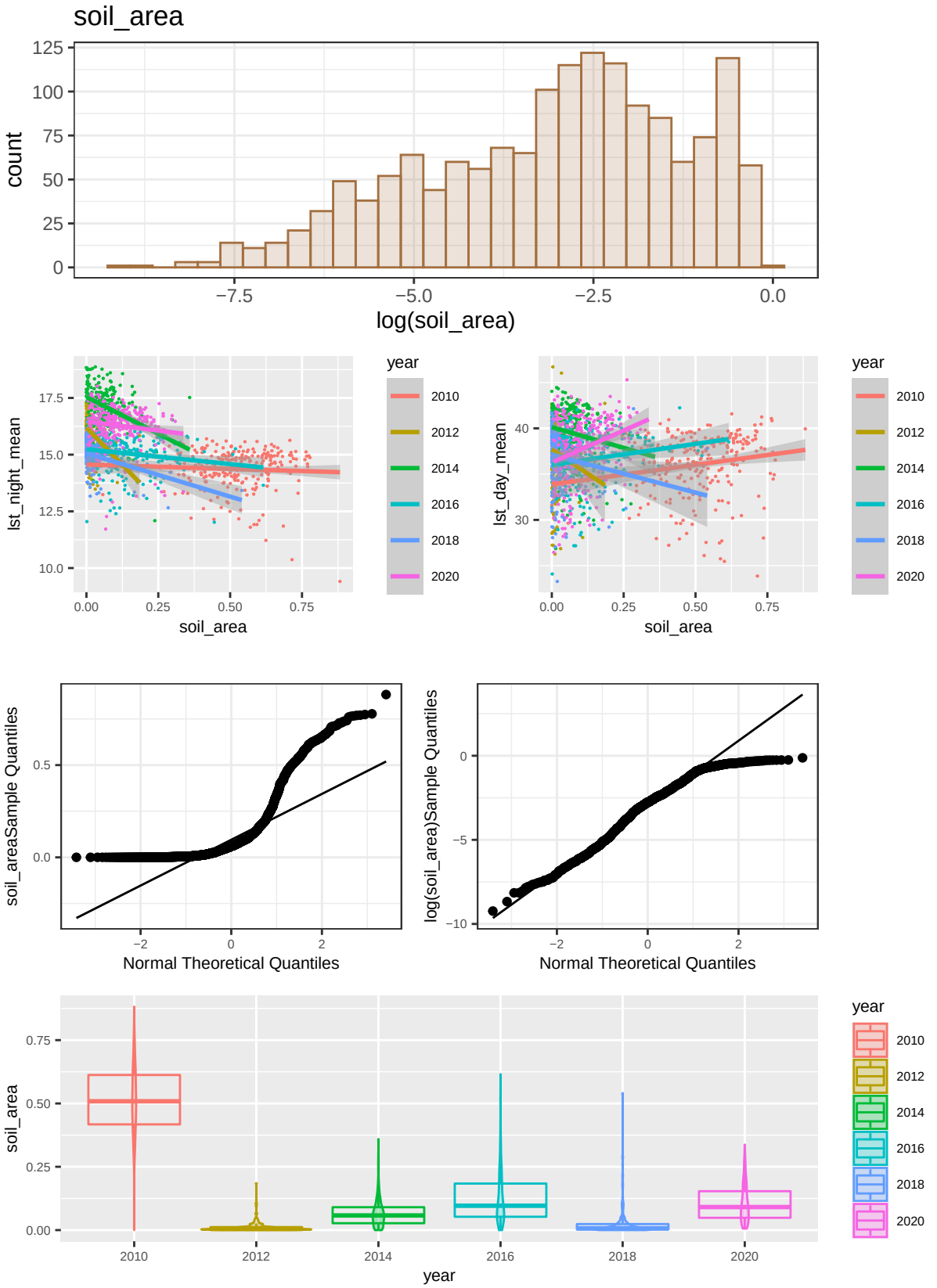
Google Earth: Tensor Flow Model

Water Area



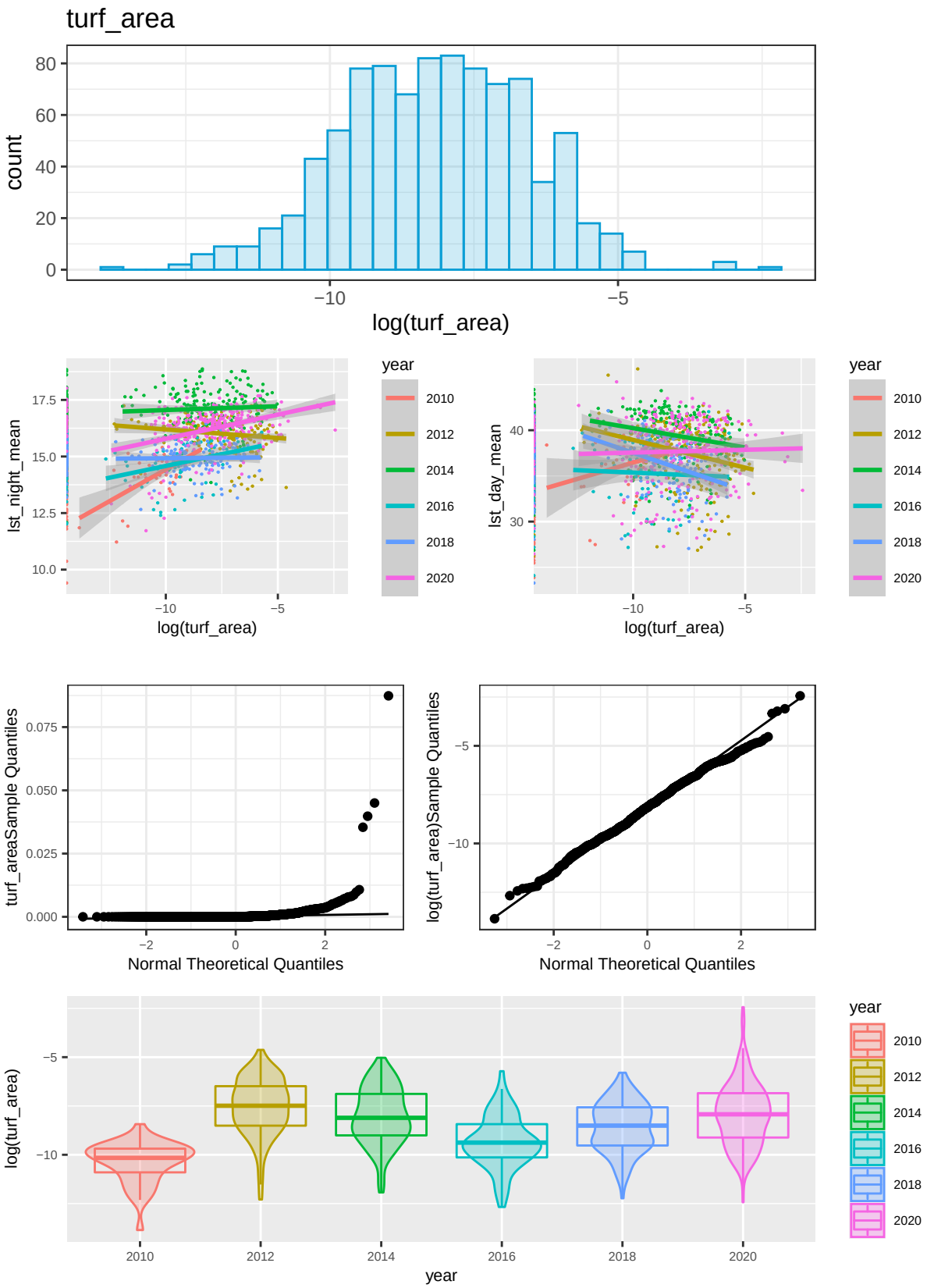
Google Earth: Tensor Flow Model

Soil Area



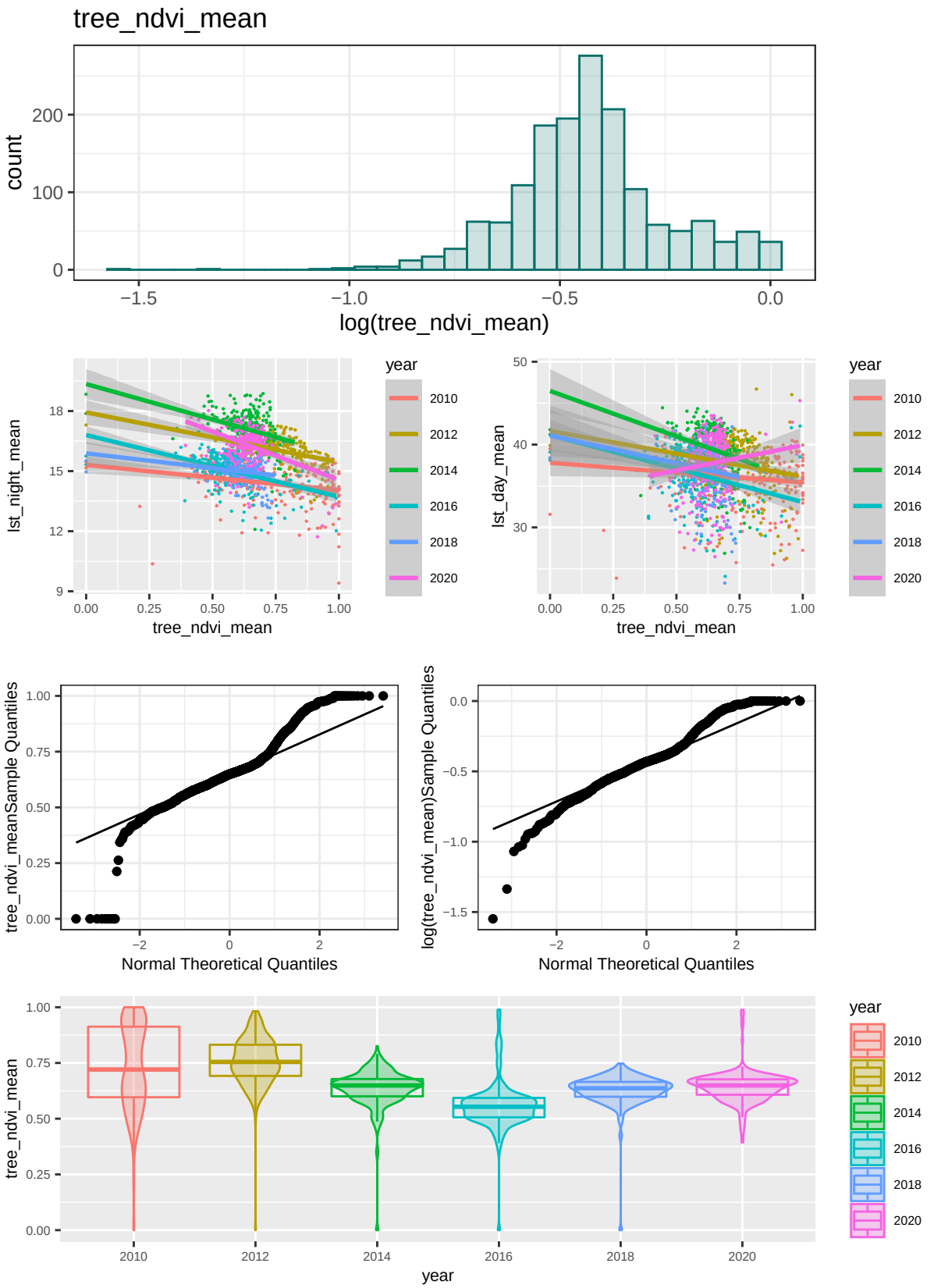
Google Earth: Tensor Flow Model

Turf Area



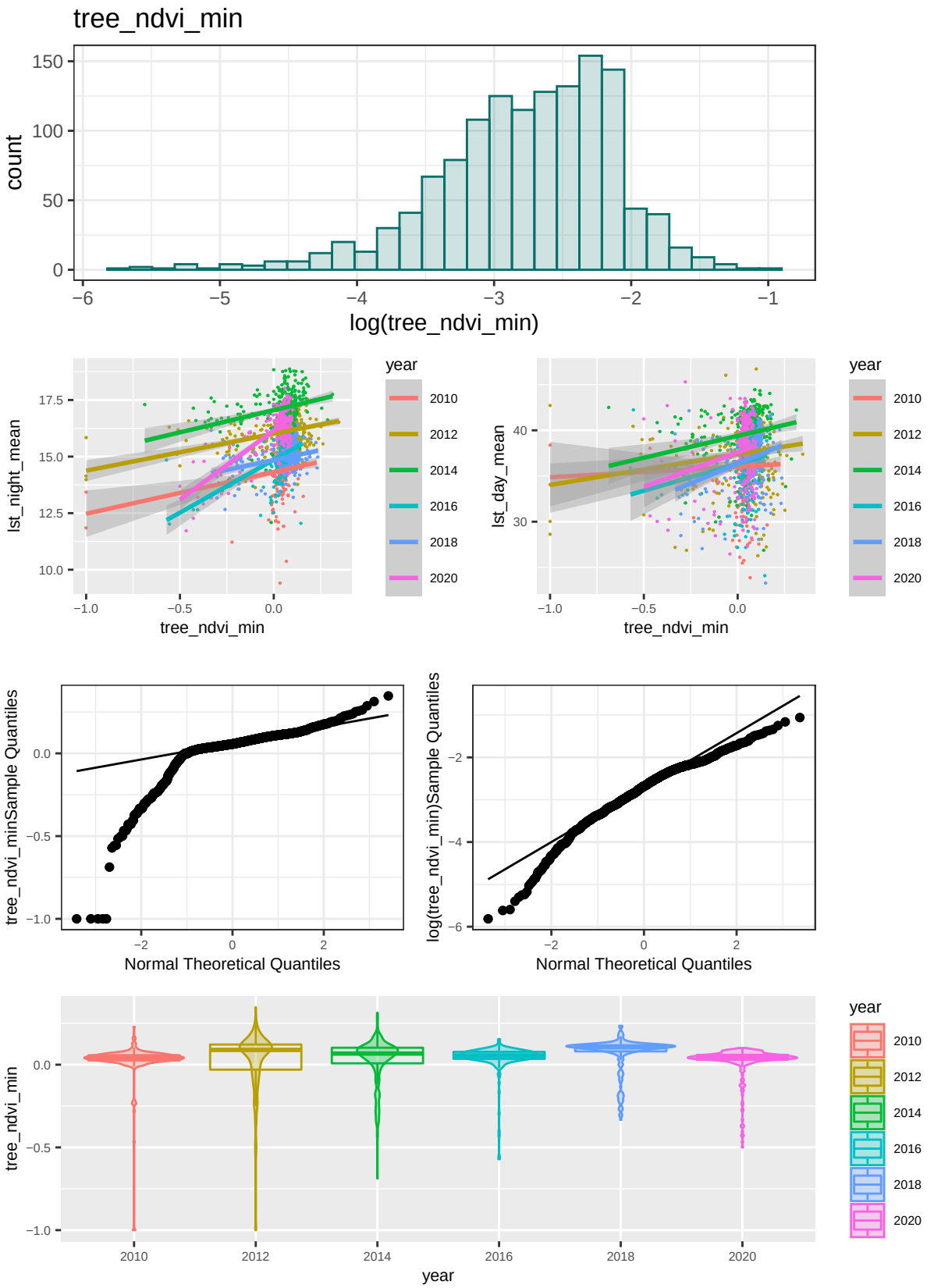
Google Earth: Tensor Flow Model

Tree NDVI Mean



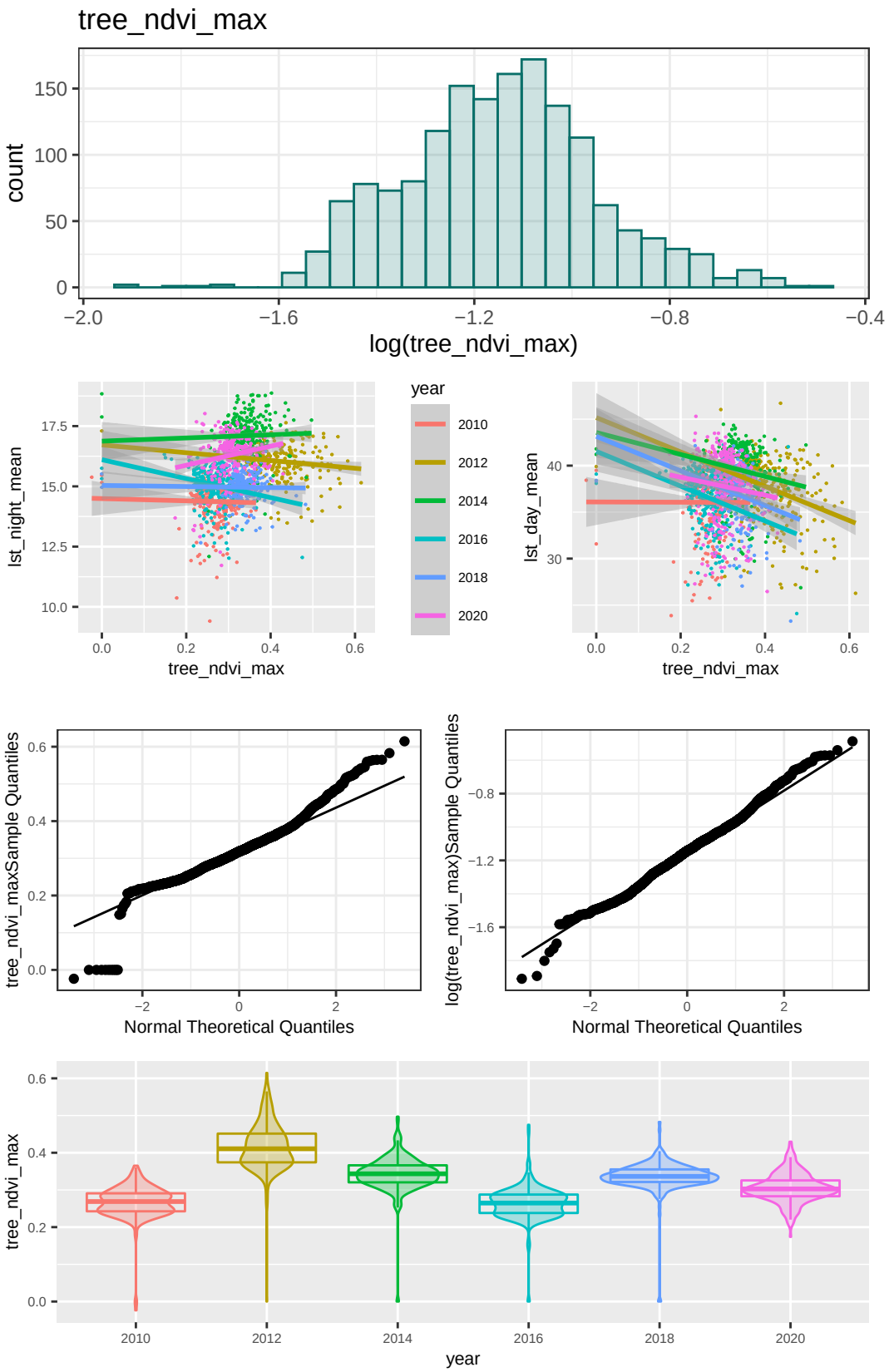
Google Earth: Tensor Flow Model

Tree NDVI Min



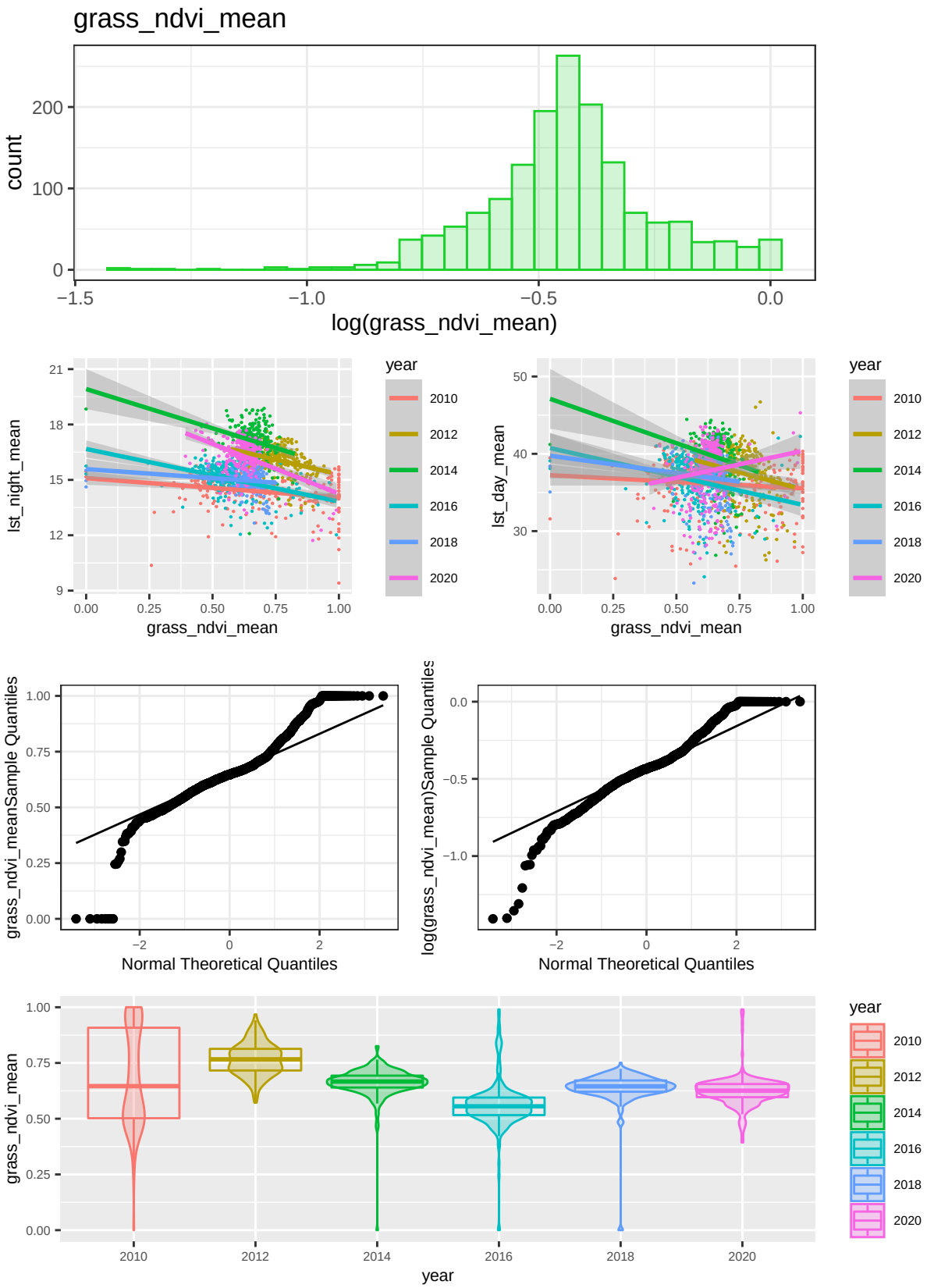
Google Earth: Tensor Flow Model

Tree NDVI Max



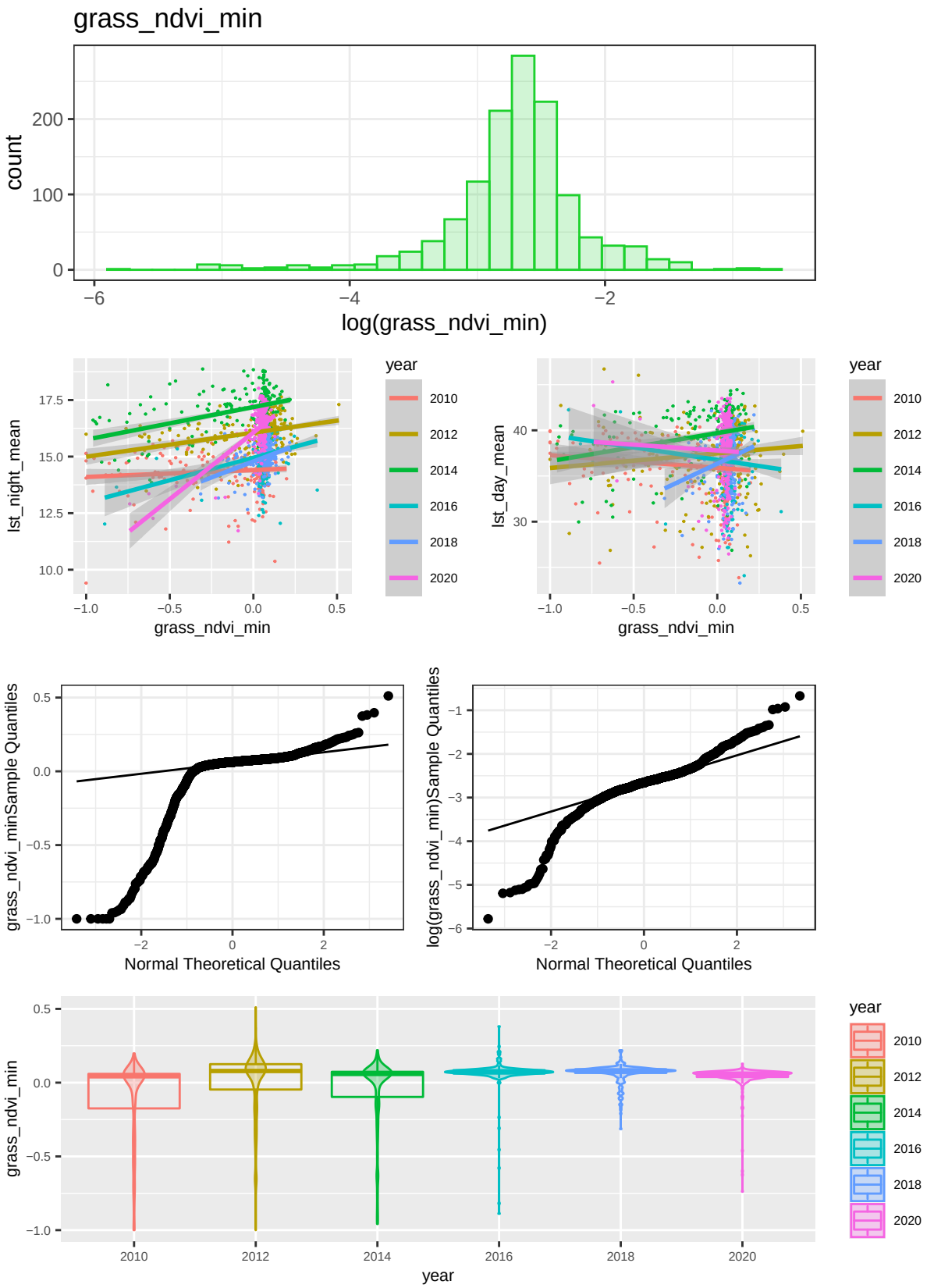
Google Earth: Tensor Flow Model

Lawn NDVI Mean



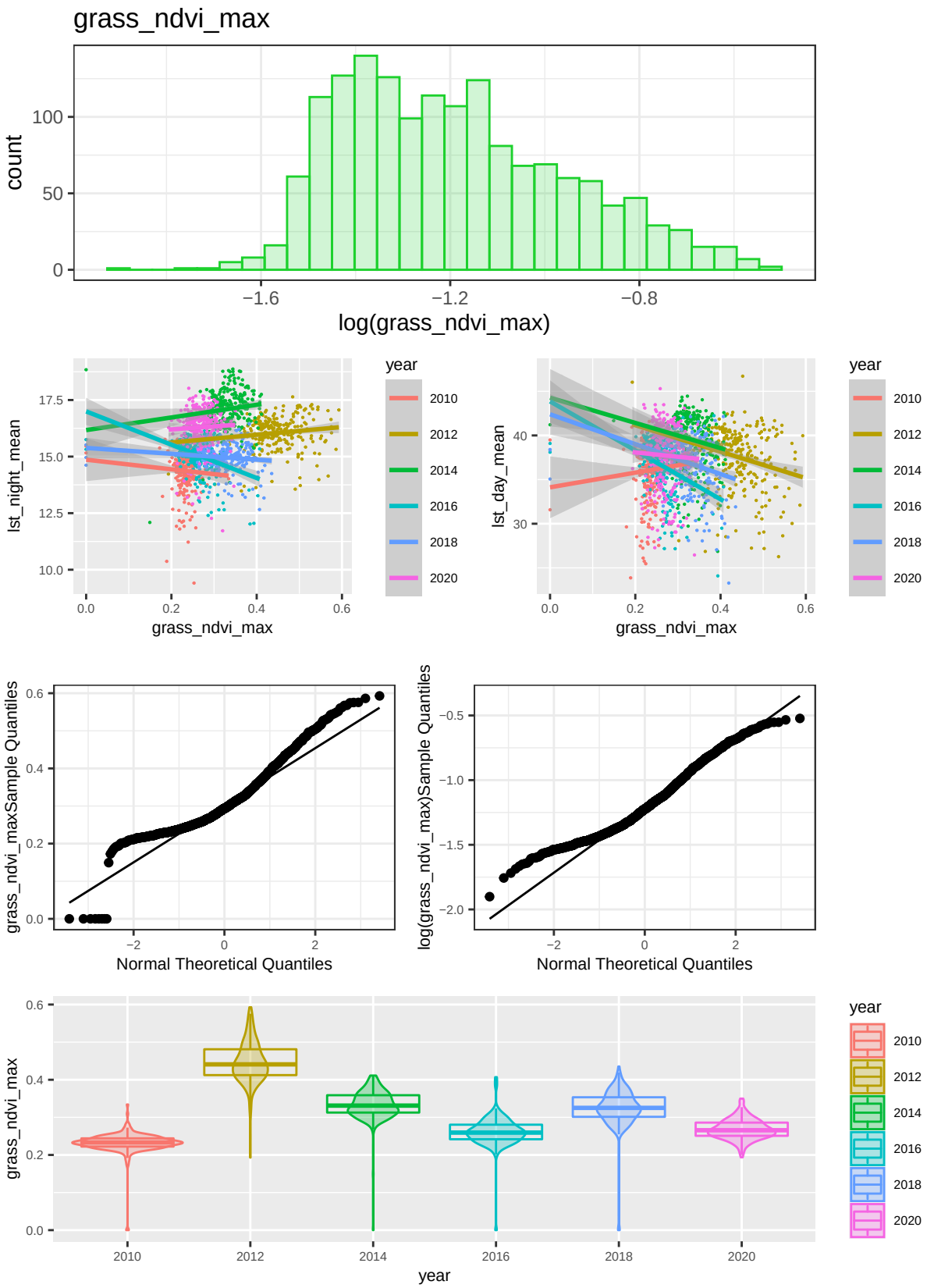
Google Earth: Tensor Flow Model

Lawn NDVI Min



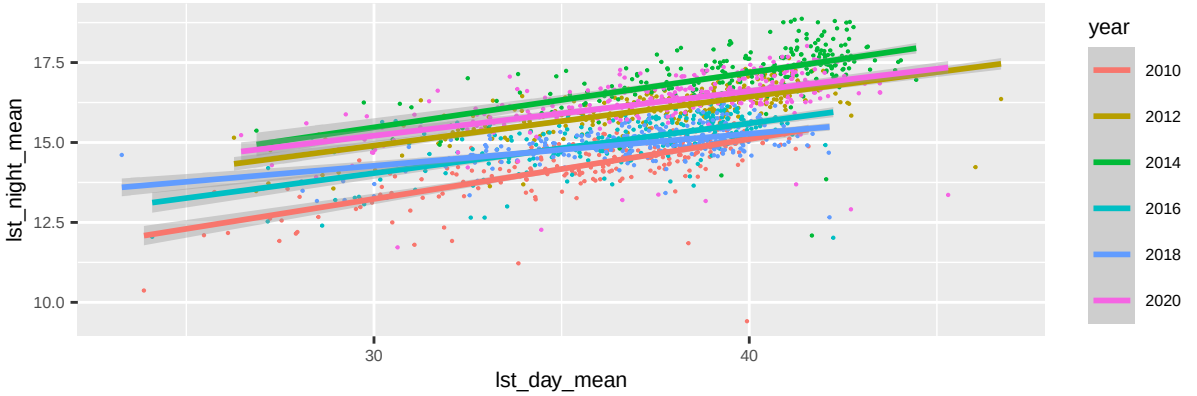
Google Earth: Tensor Flow Model

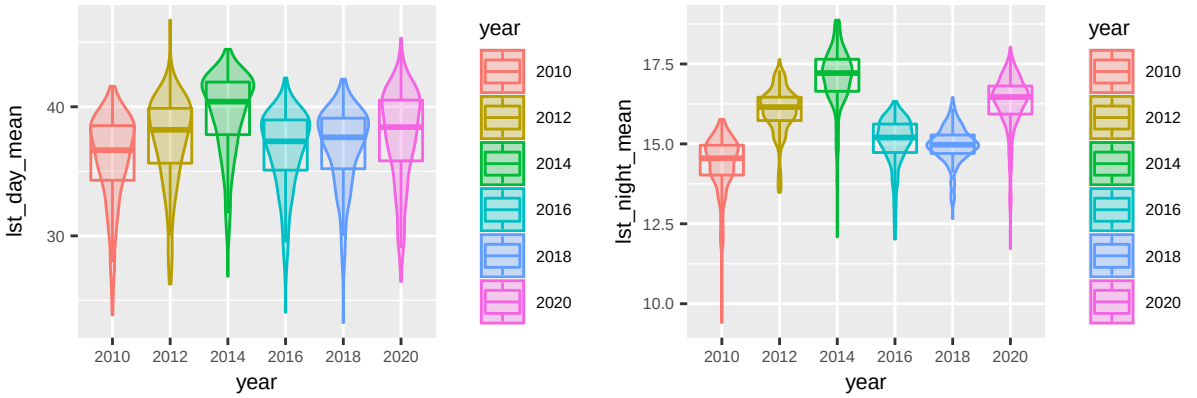
Lawn NDVI Max

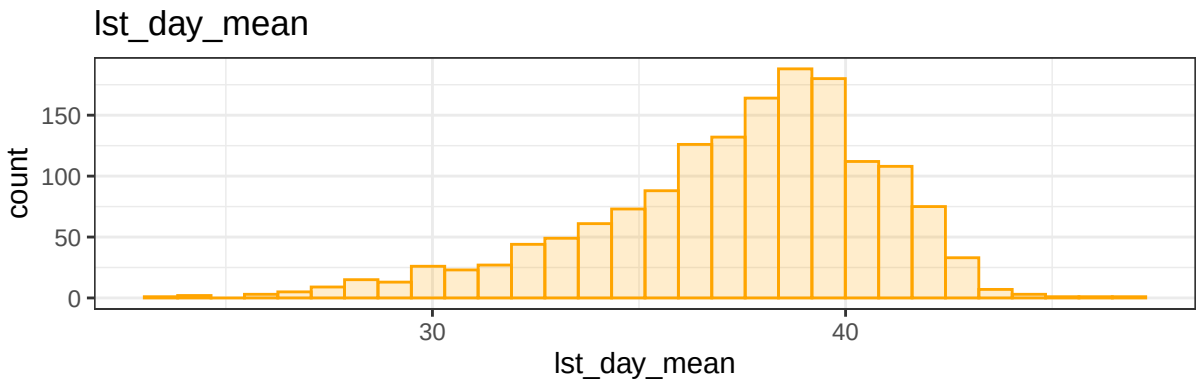


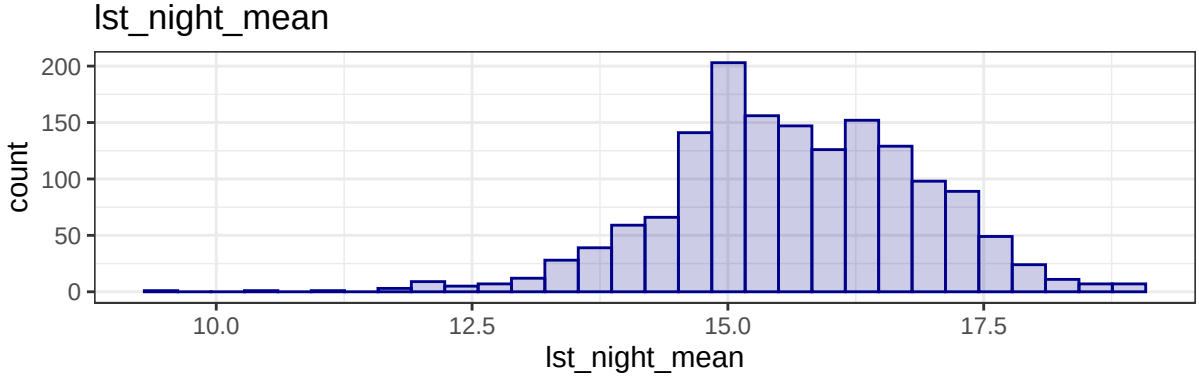
Google Earth: Tensor Flow Model

Mean Temperature









Google Earth: Tensor Flow Model

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) + tree_area + lawn_area + water_area + soil_area +
    turf_area + grass_ndvi_mean + tree_ndvi_mean,
  data = microClimatePanel)

stargazer::stargazer(OLSM1, single.row = TRUE, title = 'OLS Model')
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Tue, Jul 26, 2022 - 9:32:17 PM

Table 1: OLS Model

	Dependent variable:
	log(lst_day_mean)
factor(year)2014	3.750*** (0.019)
factor(year)2016	3.656*** (0.017)
factor(year)2018	3.677*** (0.018)
factor(year)2020	3.730*** (0.018)
tree_area	−0.294*** (0.015)
lawn_area	−0.215*** (0.061)
water_area	−2.427*** (0.759)
soil_area	0.173*** (0.032)
turf_area	−0.977 (0.652)
grass_ndvi_mean	−0.070 (0.076)
tree_ndvi_mean	0.021 (0.073)
Observations	1,057
R^2	1.000
Adjusted R^2	1.000
Residual Std. Error	0.074 (df = 1046)
F Statistic	230,008.000*** (df = 11; 1046)
Note:	*p<0.1; **p<0.05; ***p<0.01

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')

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

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Tue, Jul 26, 2022 - 9:32:18 PM

Table 2: Fixed Effects Model

	Dependent variable:
	log(lst_day_mean)
factor(year)2014	0.040^{***} (0.002)
factor(year)2016	-0.042^{***} (0.002)
factor(year)2018	-0.030^{***} (0.002)
tree_area	-0.020^{**} (0.008)
lawn_area	-0.065^{***} (0.018)
water_area	−0.138 (0.195)
soil_area	0.044^{***} (0.009)
turf_area	0.153 (0.237)
grass_ndvi_mean	0.010 (0.020)
tree_ndvi_mean	0.020 (0.018)
Observations	1,057
R^2	0.852
Adjusted R^2	0.794
F Statistic	437.403^{***} (df = 10; 759)
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 = 88.093, df1 = 287, 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 + 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 = 'random')

stargazer::stargazer(reM1, single.row = TRUE, title = 'Random Effects Model')
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Tue, Jul 26, 2022 - 9:32:18 PM

Table 3: Random Effects Model

	Dependent variable:
	log(lst_day_mean)
factor(year)2014	3.657*** (0.008)
factor(year)2016	3.575*** (0.007)
factor(year)2018	3.586*** (0.008)
factor(year)2020	3.620*** (0.008)
tree_area	−0.051*** (0.008)
lawn_area	−0.084*** (0.019)
water_area	−0.213 (0.212)
soil_area	0.051*** (0.010)
turf_area	−0.053 (0.248)
grass_ndvi_mean	0.003 (0.021)
tree_ndvi_mean	0.026 (0.019)
Observations	1,057
R^2	0.946
Adjusted R^2	0.946
F Statistic	648,701.300***
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:  log(lst_day_mean) ~ 0 + factor(year) + tree_area + lawn_area +  ...
## chisq = 46765, df = 10, 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(
  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: Tue, Jul 26, 2022 - 9:32:18 PM

```
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
```

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

```
##
##  Lagrange Multiplier Test - time effects (Breusch-Pagan) for unbalanced
##  panels
```

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

```
##
## 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
```

```
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  0.0396623  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
```

```
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
```

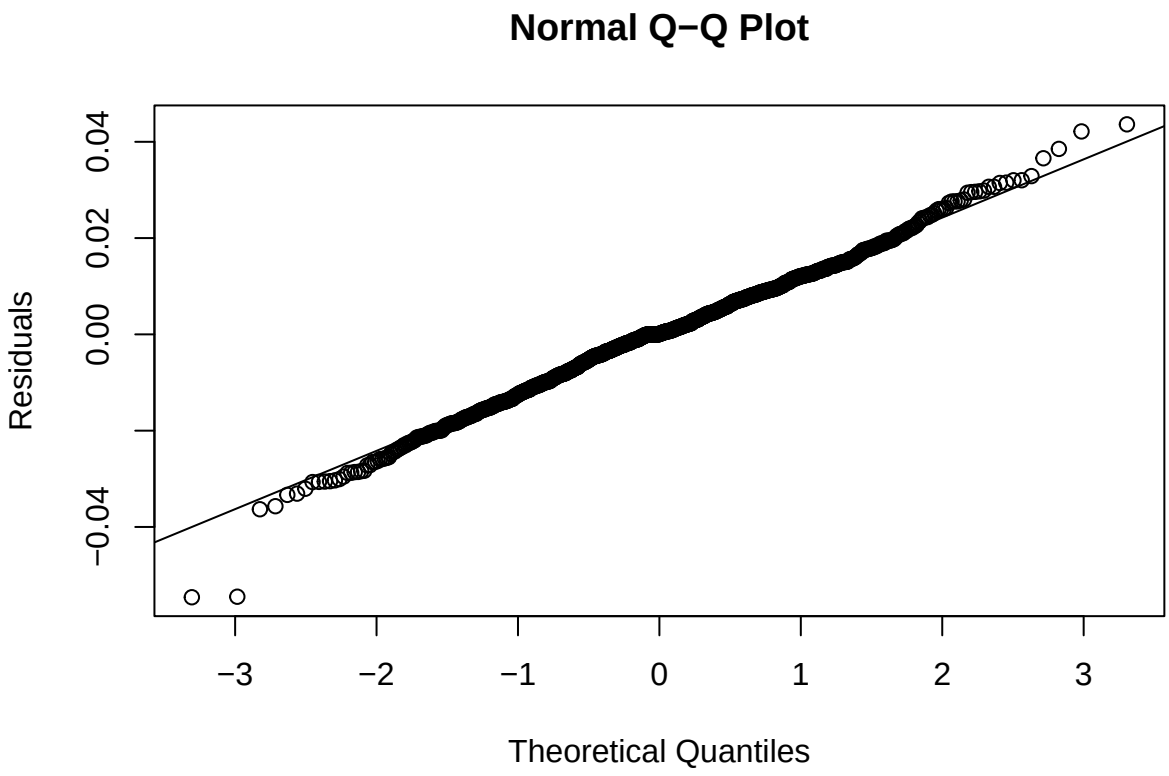
```
stargazer::stargazer(OLSM1,feM1,reM1,feM2FixedTime,
                     font.size = 'tiny',
                     title = 'Regression Models Sumamry',
                     column.labels = c("OLS", "FixedEffects",
                                       "RandomEffects",
                                       "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: Tue, Jul 26, 2022 - 9:32:21 PM

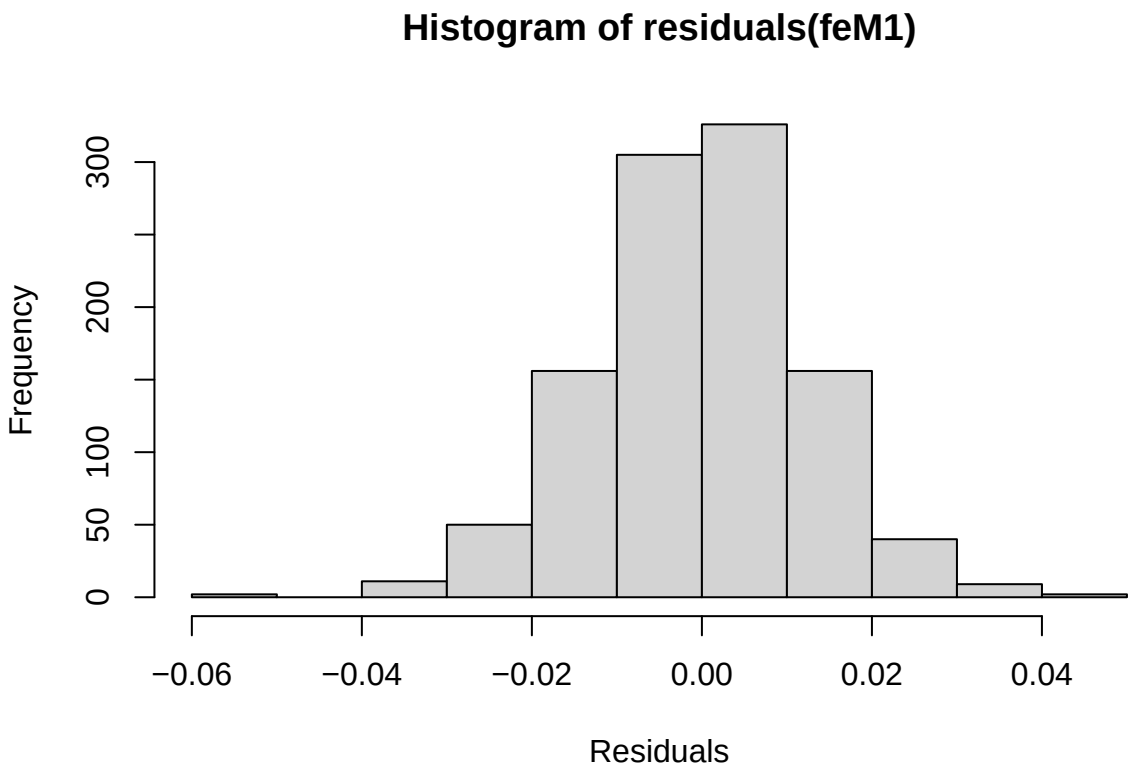
```
qqnorm(residuals(feM1), ylab = 'Residuals')
qqline(residuals(feM1))
```

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	3.750***	0.040***	3.657***	
	(0.019)	(0.002)	(0.008)	
factor(year)2016	3.656***	−0.042***	3.575***	
	(0.017)	(0.002)	(0.007)	
factor(year)2018	3.677***	−0.030***	3.586***	
	(0.018)	(0.002)	(0.008)	
factor(year)2020	3.730***		3.620***	
	(0.018)		(0.008)	
tree_area	−0.294***	−0.020**	−0.051***	0.011
	(0.015)	(0.008)	(0.008)	(0.013)
lawn_area	−0.215***	−0.065***	−0.084***	0.449***
	(0.061)	(0.018)	(0.019)	(0.028)
water_area	−2.427***	−0.138	−0.213	0.151
	(0.759)	(0.195)	(0.212)	(0.388)
soil_area	0.173***	0.044***	0.051***	0.022
	(0.032)	(0.009)	(0.010)	(0.016)
turf_area	−0.977	0.153	−0.053	0.534
	(0.652)	(0.237)	(0.248)	(0.455)
grass_ndvi_mean	−0.070	0.010	0.003	0.261***
	(0.076)	(0.020)	(0.021)	(0.035)
tree_ndvi_mean	0.021	0.020	0.026	−0.132***
	(0.073)	(0.018)	(0.019)	(0.033)
Observations	1,057	1,057	1,057	1,057
R^2	1.000	0.852	0.946	0.409
Adjusted R^2	1.000	0.794	0.946	0.181
Residual Std. Error	0.074 (df = 1046)			
F Statistic	230,008.000*** (df = 11; 1046)	437.403*** (df = 10; 759)	648,701.300***	75.386*** (df = 7; 762)
Note:				*p<0.1; **p<0.05; ***p<0.01




```
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