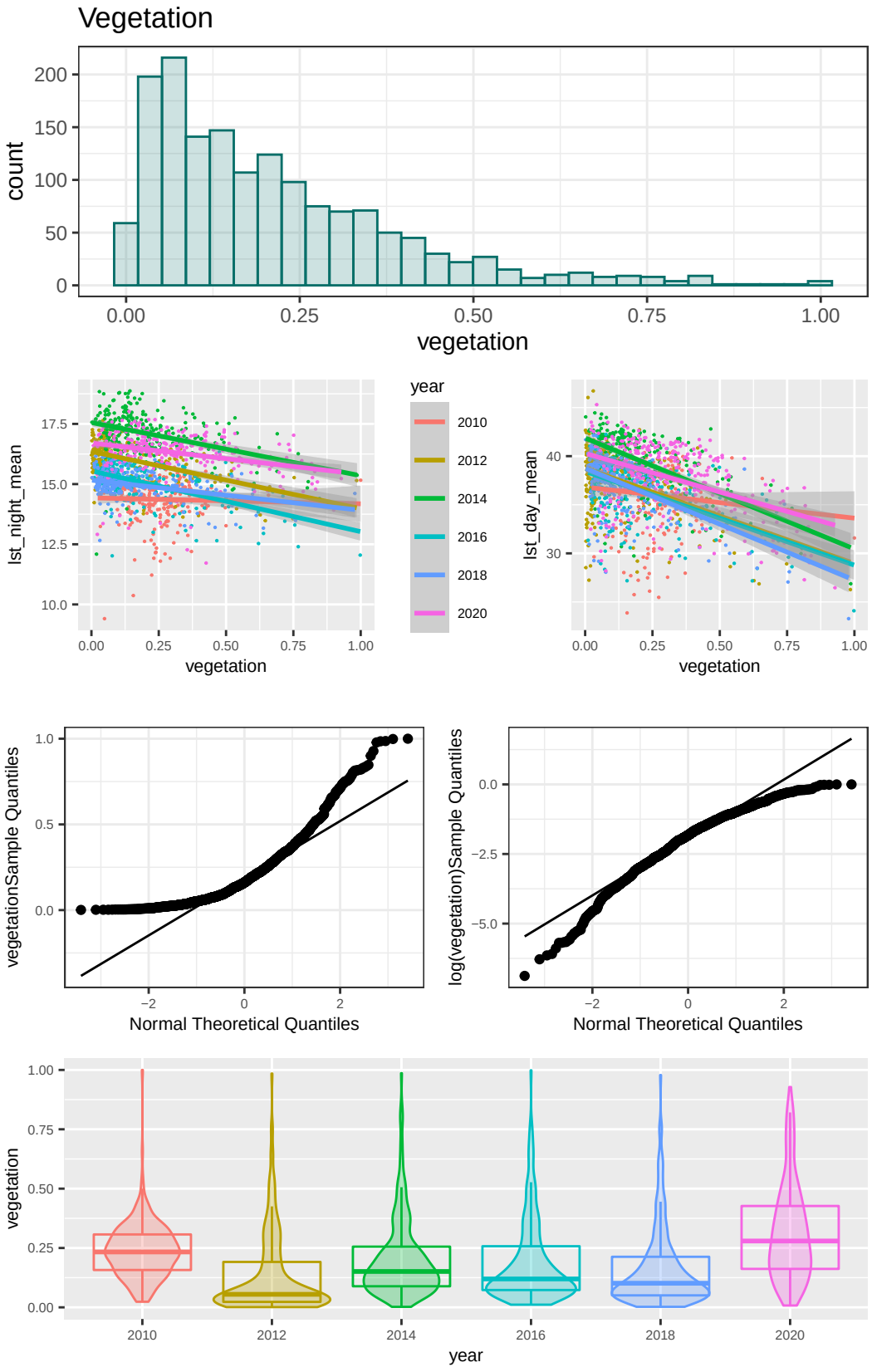
LawnProject Panel Data Analysis / Tensor Flow Data

LawnProject Team

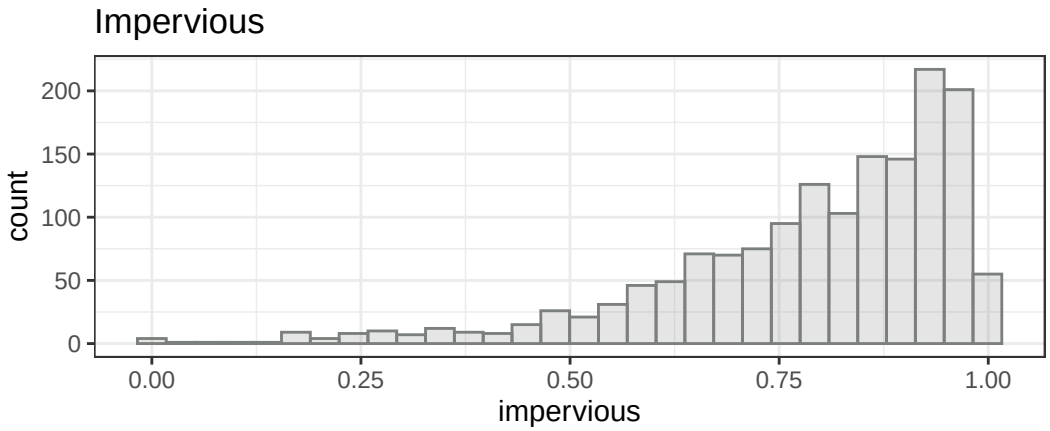
Intro

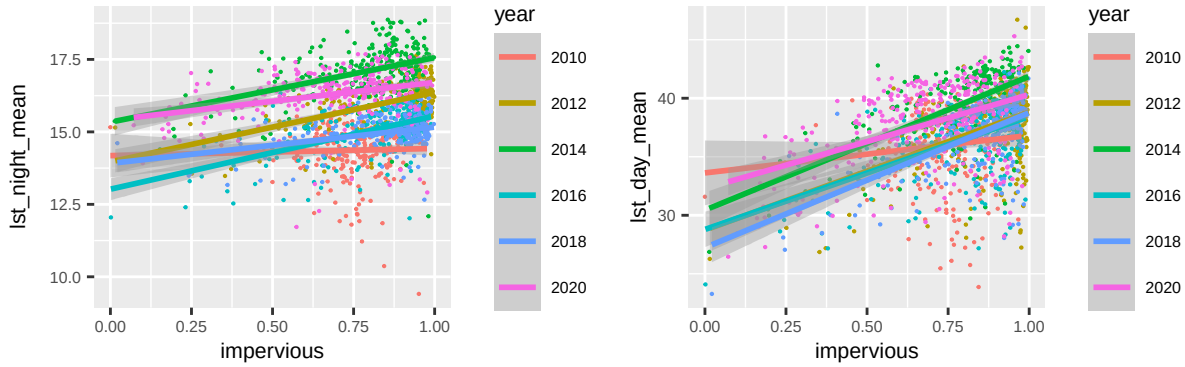
Aggregate Vegetation Area (Trees + Lawn)

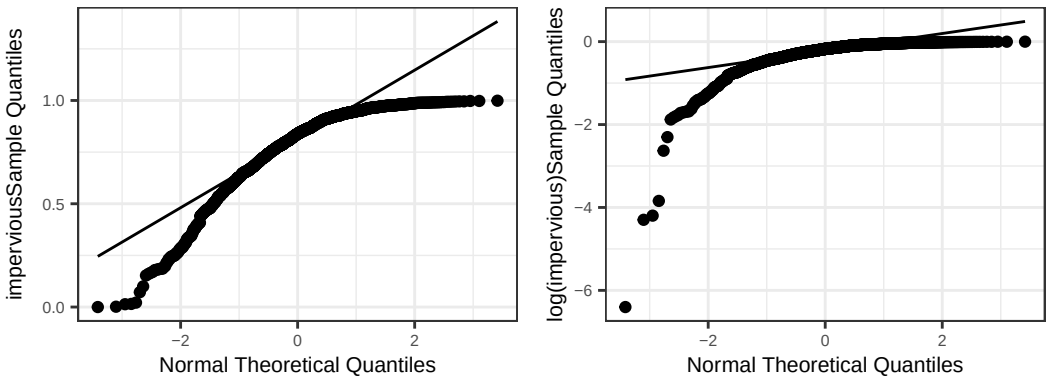


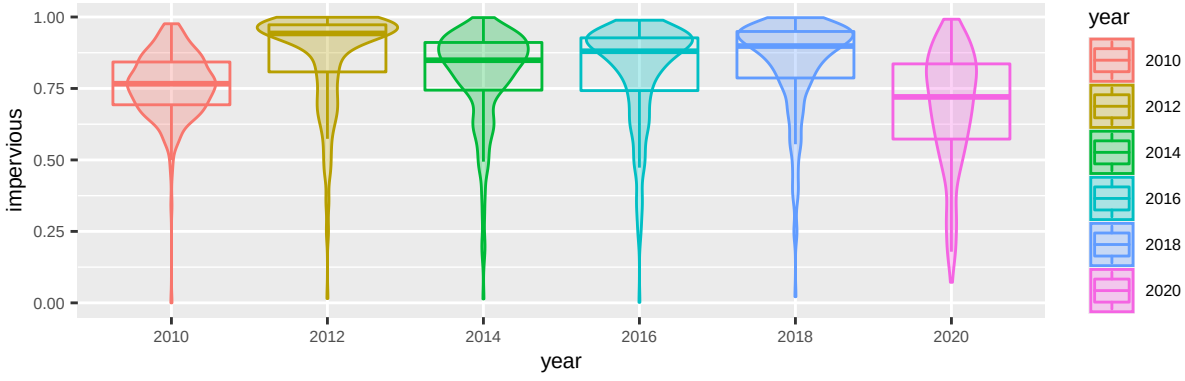
Google Earth: Tensor Flow Model

Aggregate Impervious Area (Soil + Turf + Impervious)



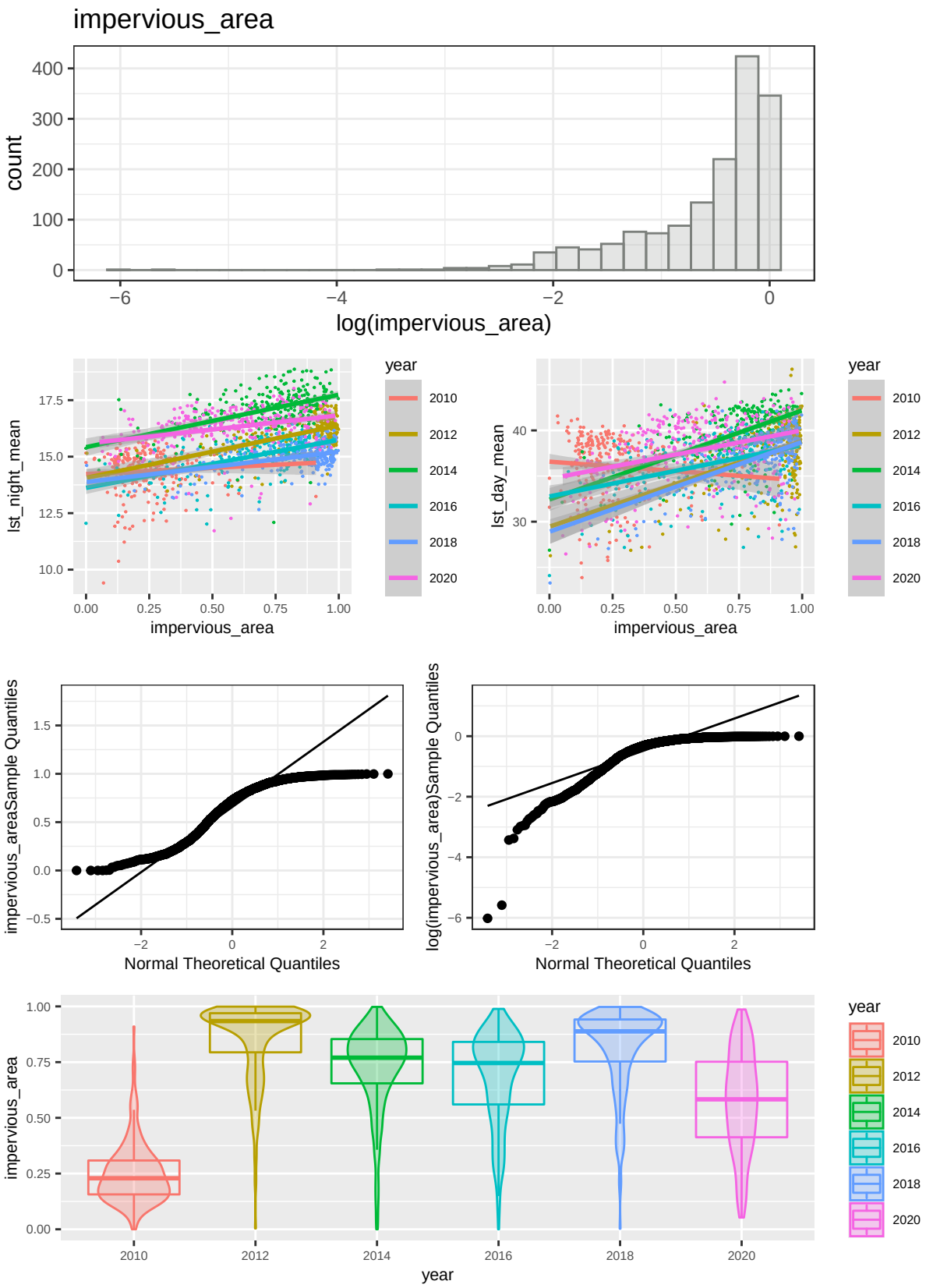






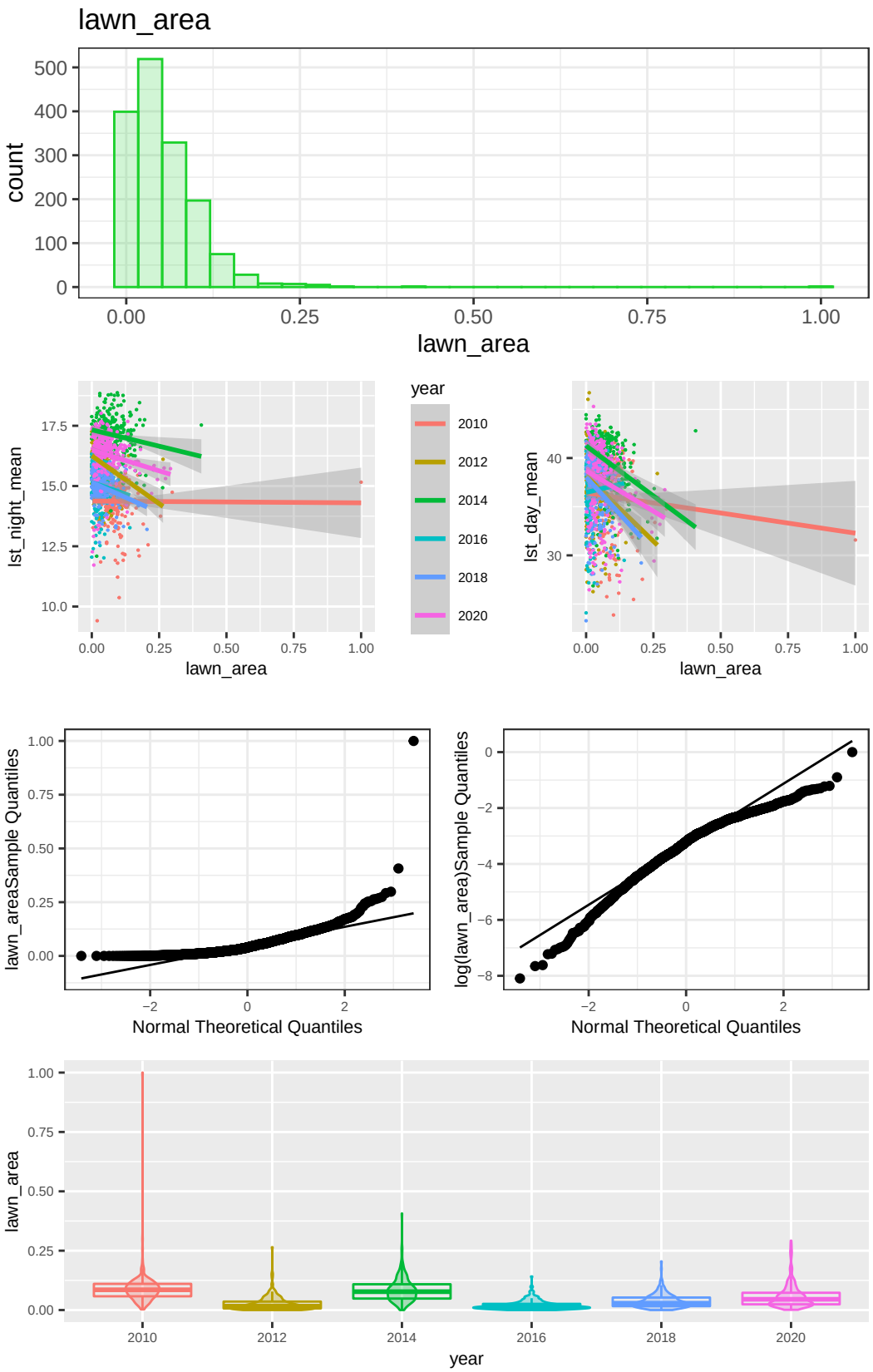
Google Earth: Tensor Flow Model

Impervious Area



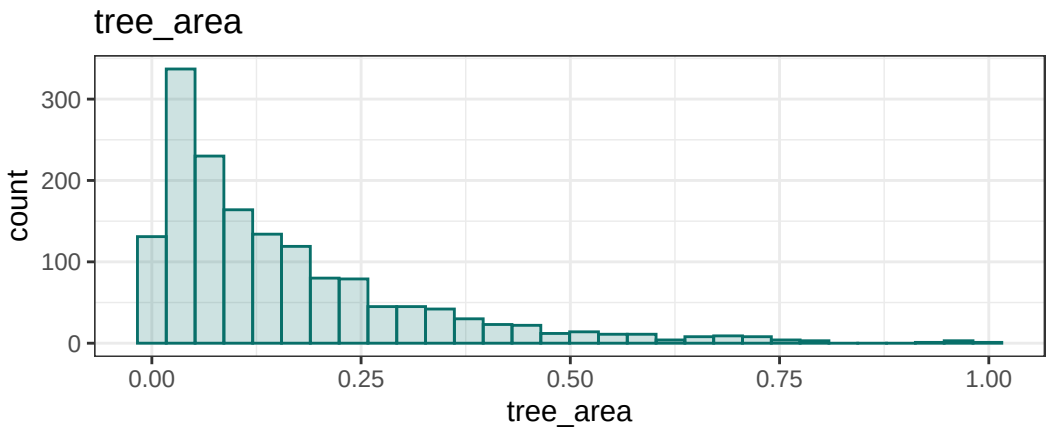
Google Earth: Tensor Flow Model

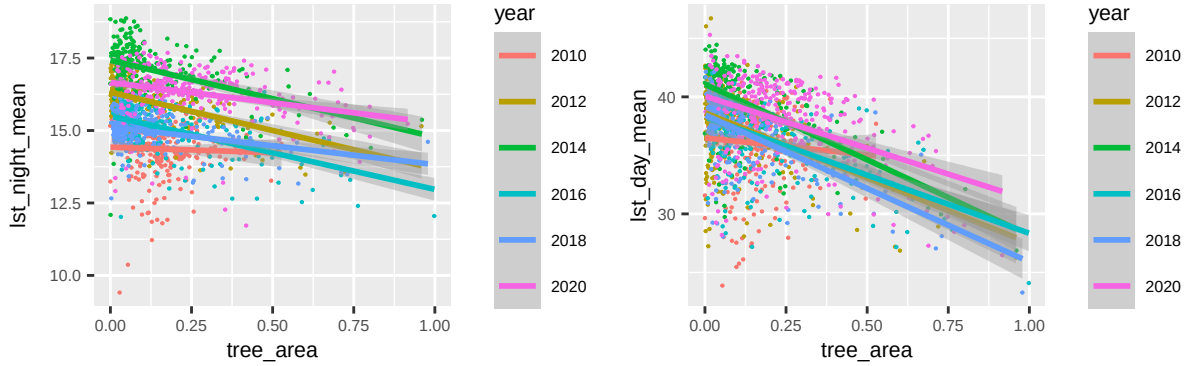
Lawn Area

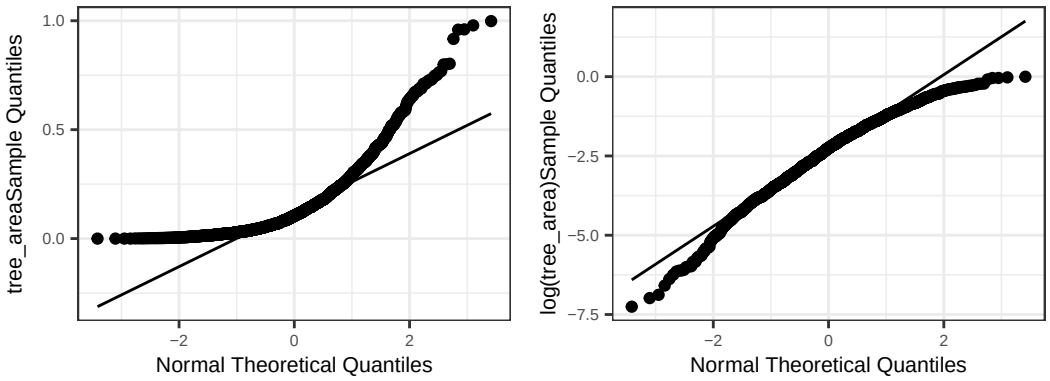


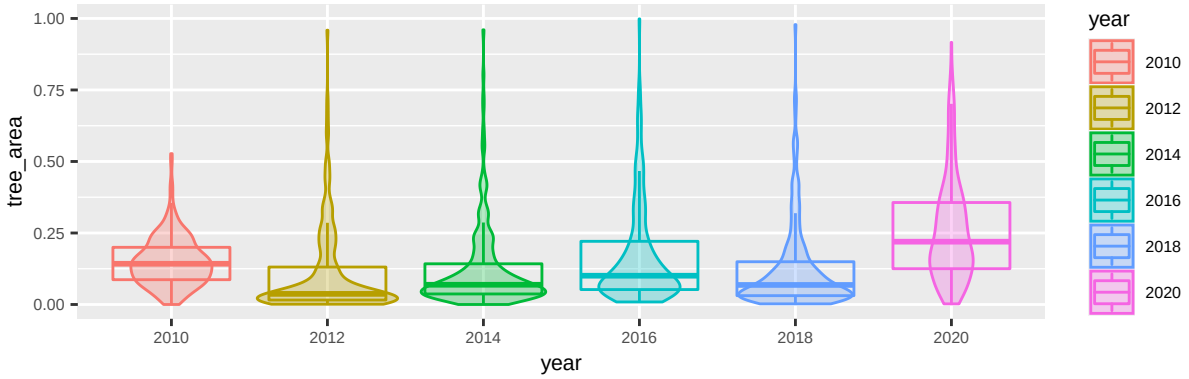
Google Earth: Tensor Flow Model

Tree Area



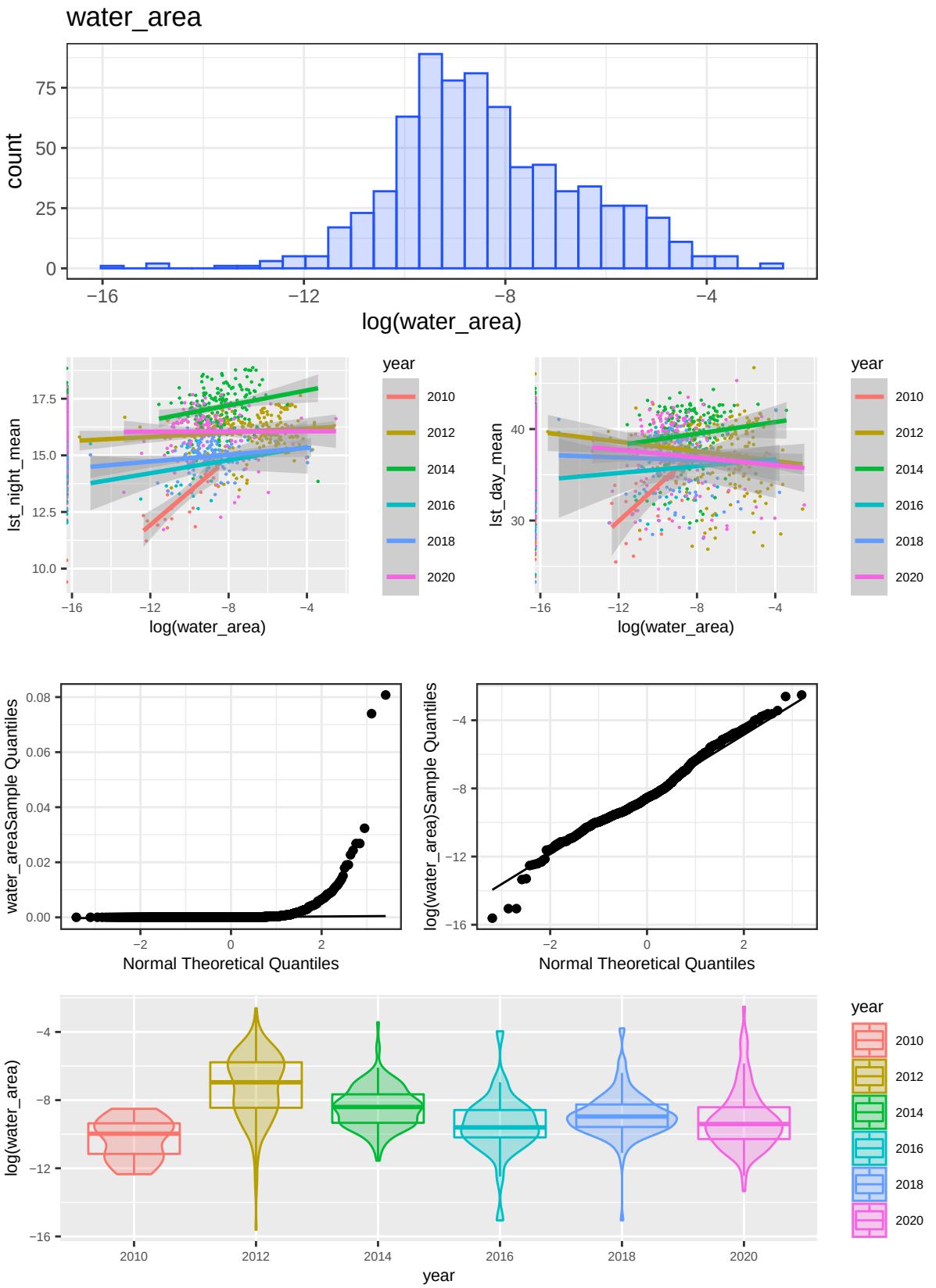






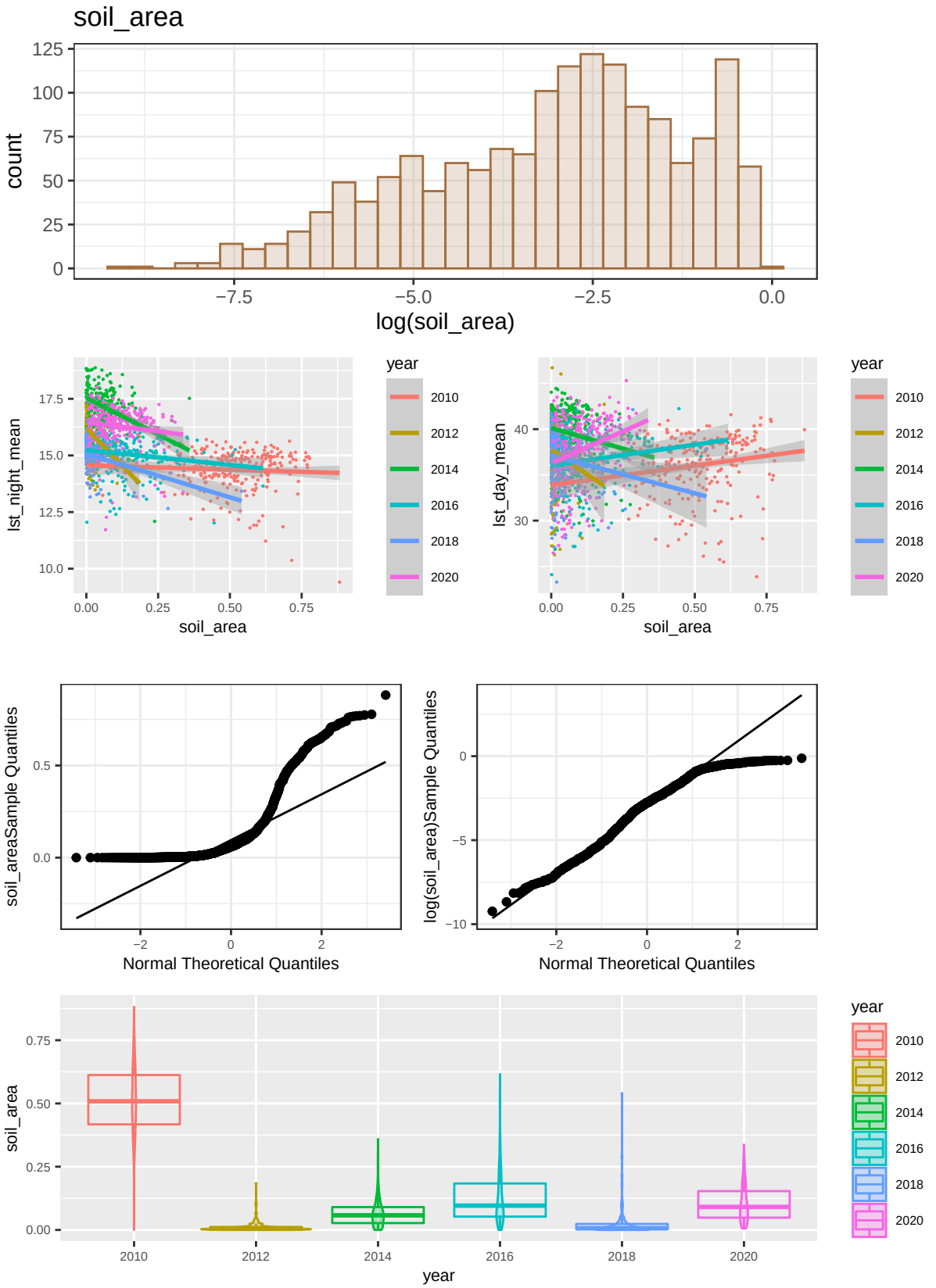
Google Earth: Tensor Flow Model

Water Area



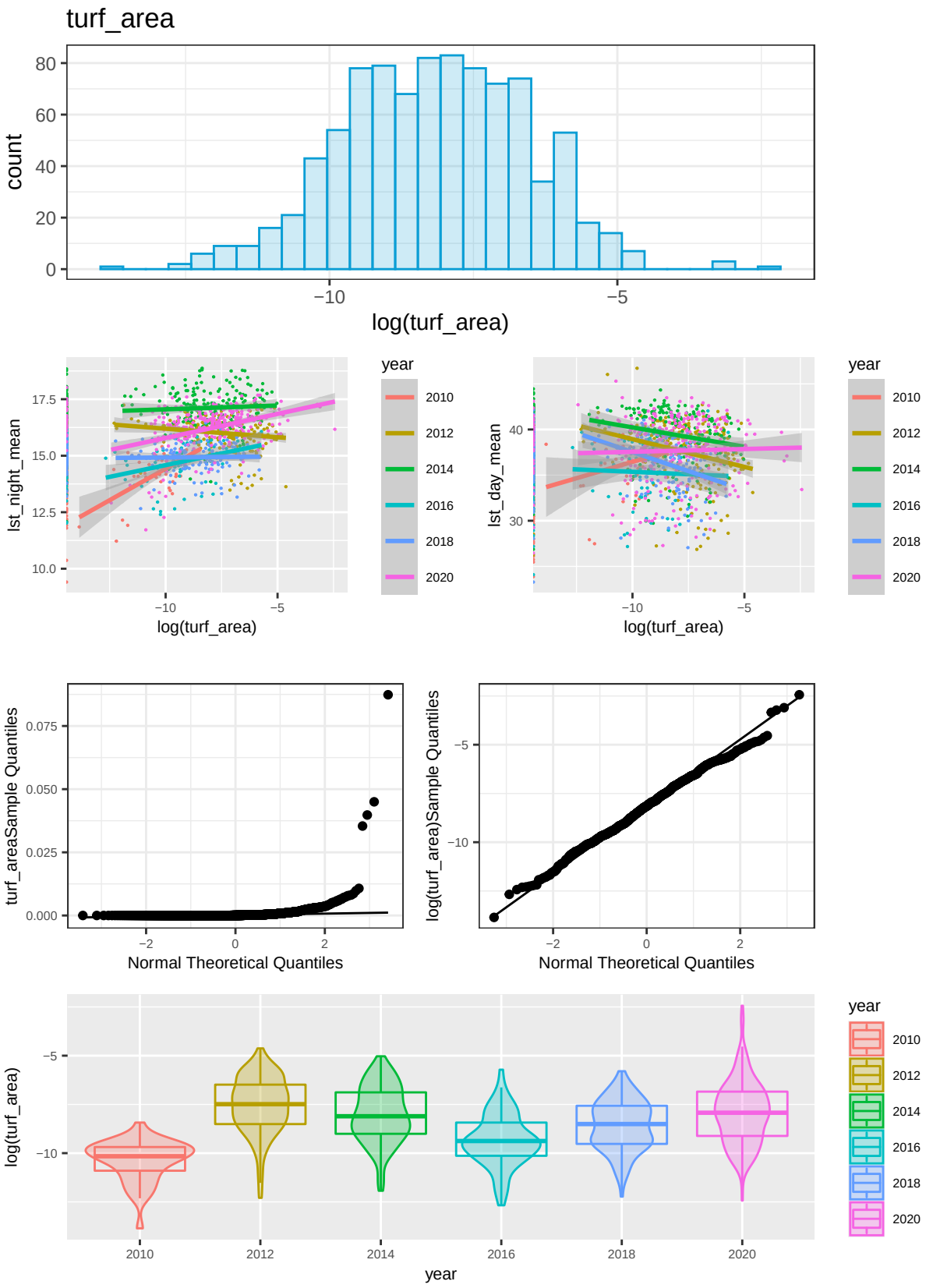
Google Earth: Tensor Flow Model

Soil Area



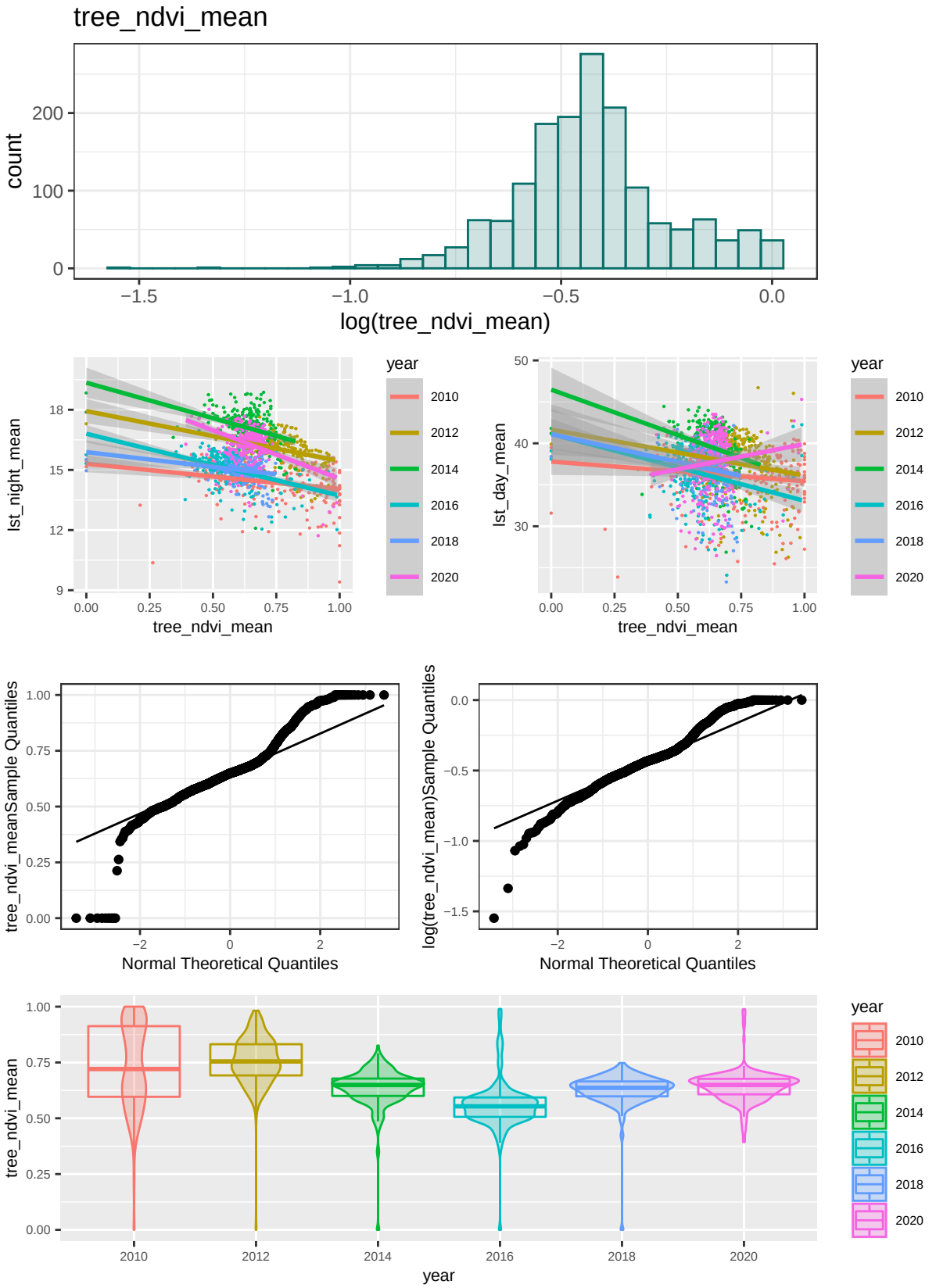
Google Earth: Tensor Flow Model

Turf Area



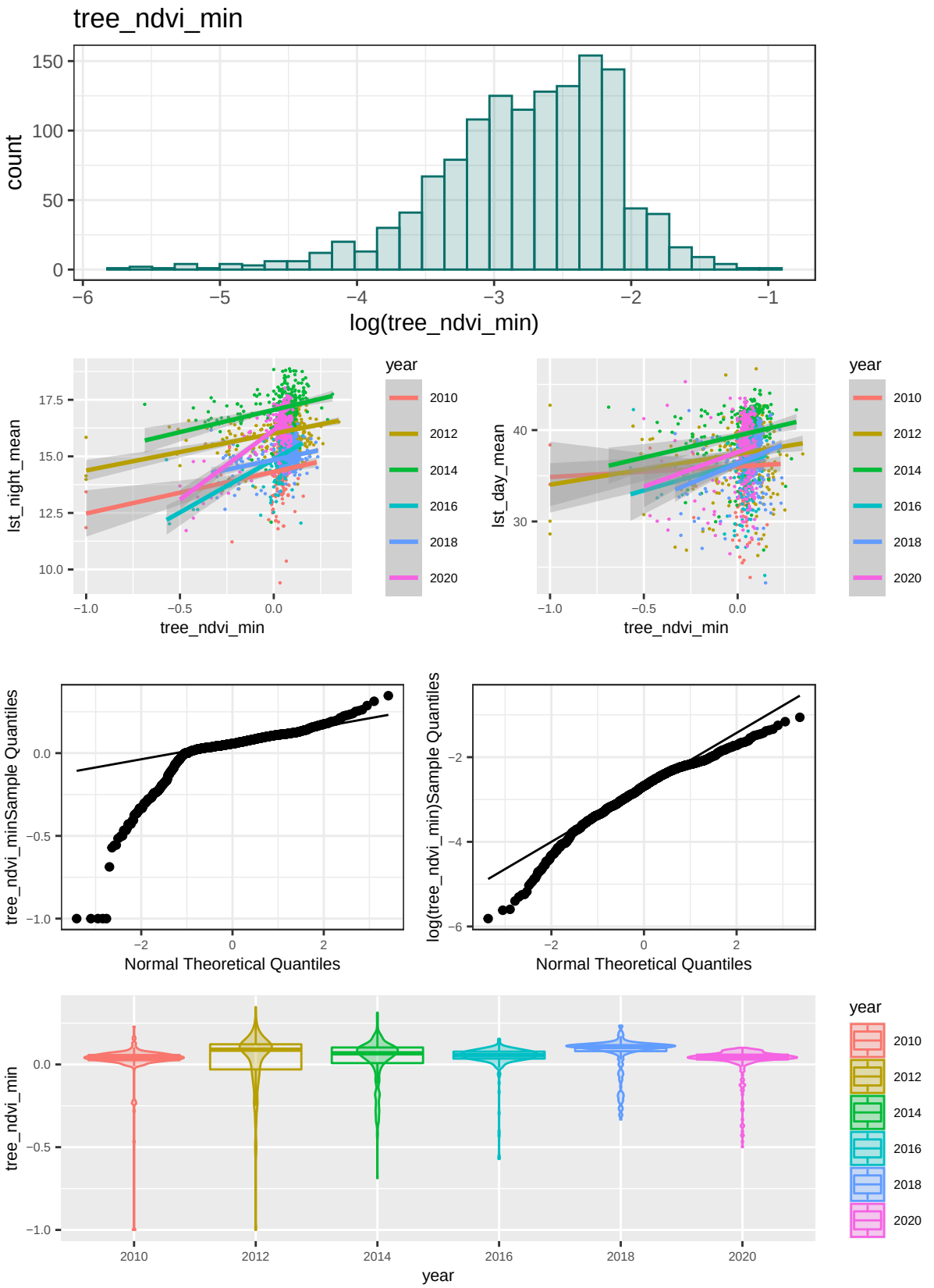
Google Earth: Tensor Flow Model

Tree NDVI Mean



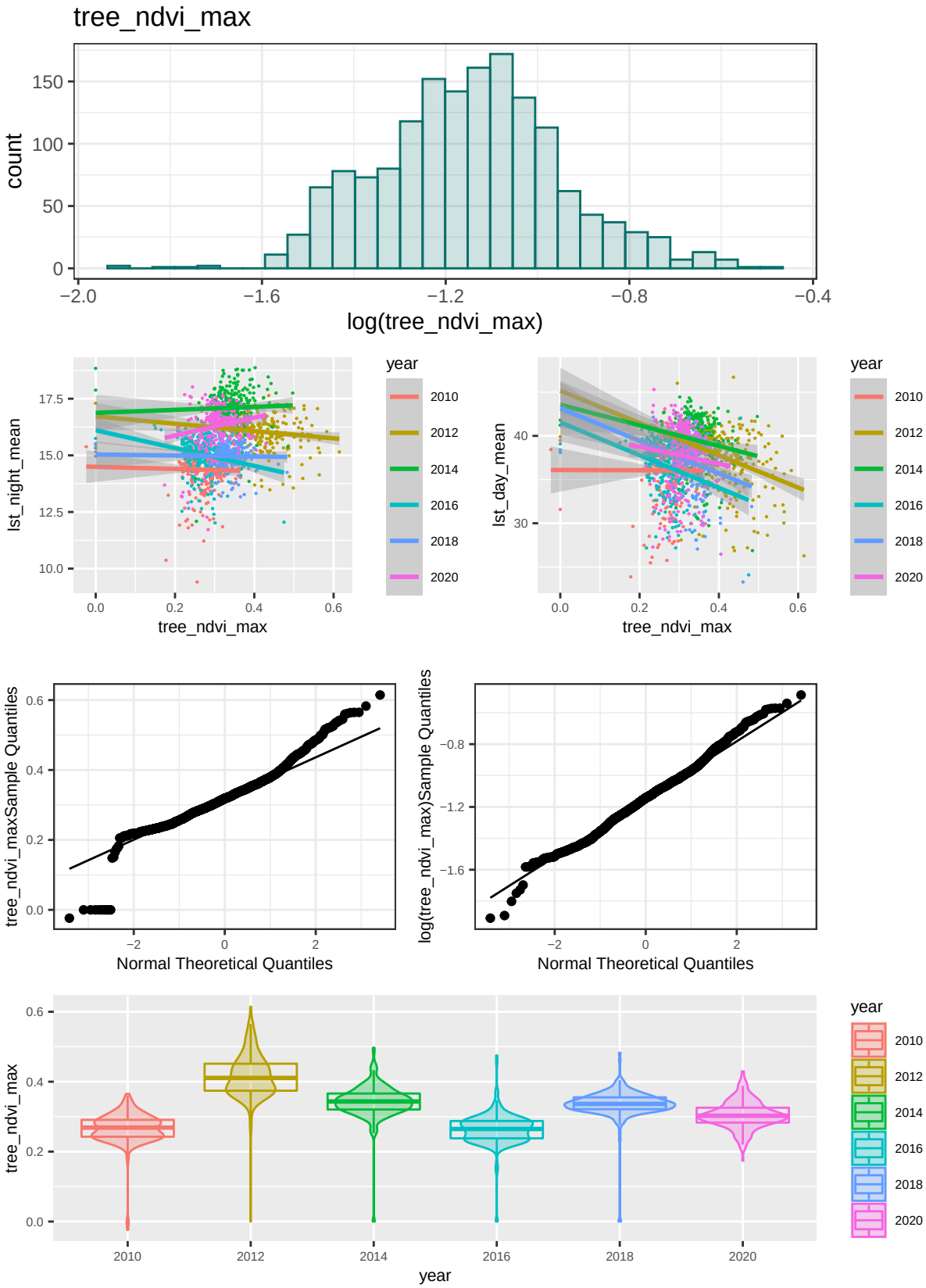
Google Earth: Tensor Flow Model

Tree NDVI Min



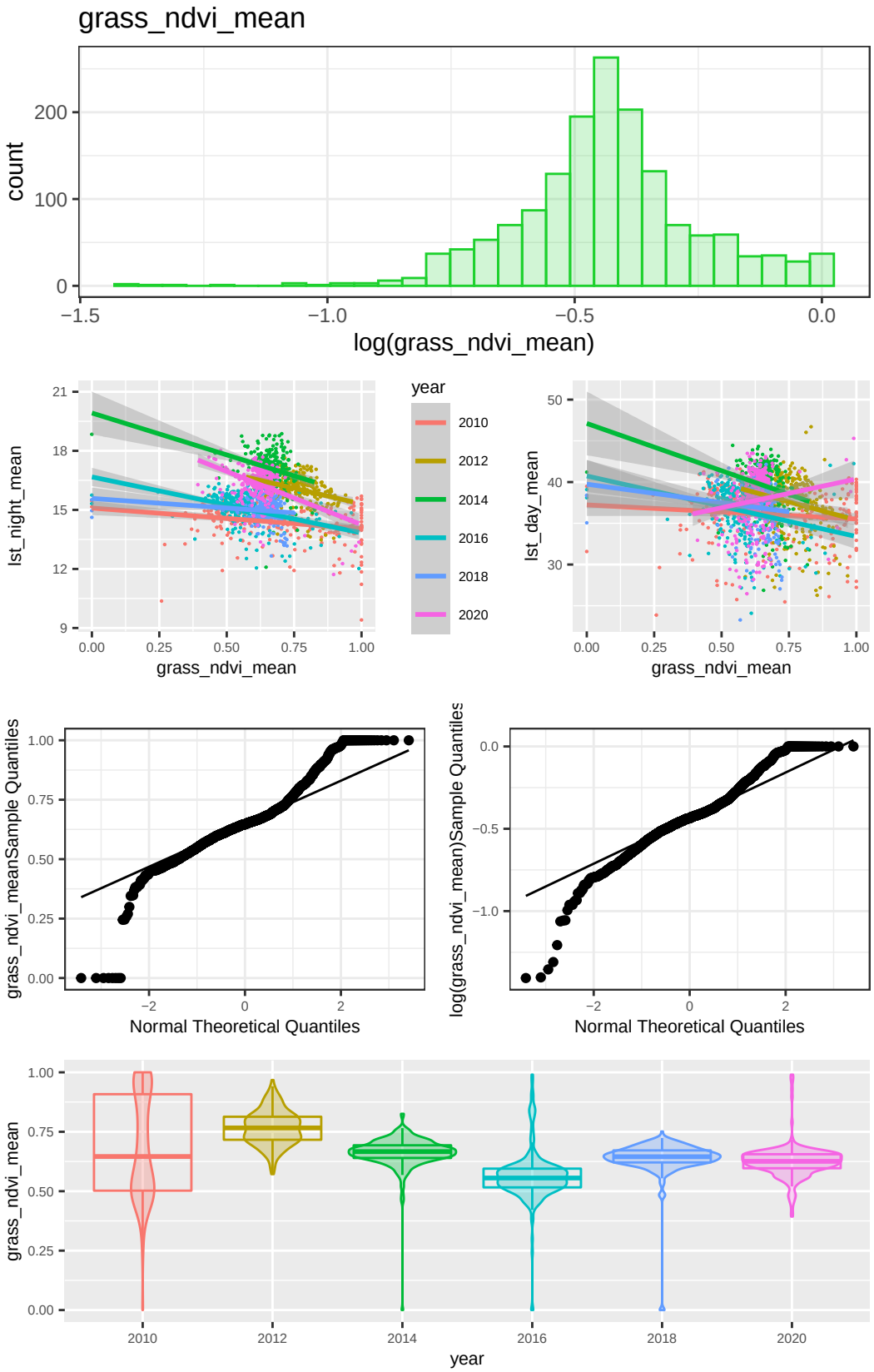
Google Earth: Tensor Flow Model

Tree NDVI Max



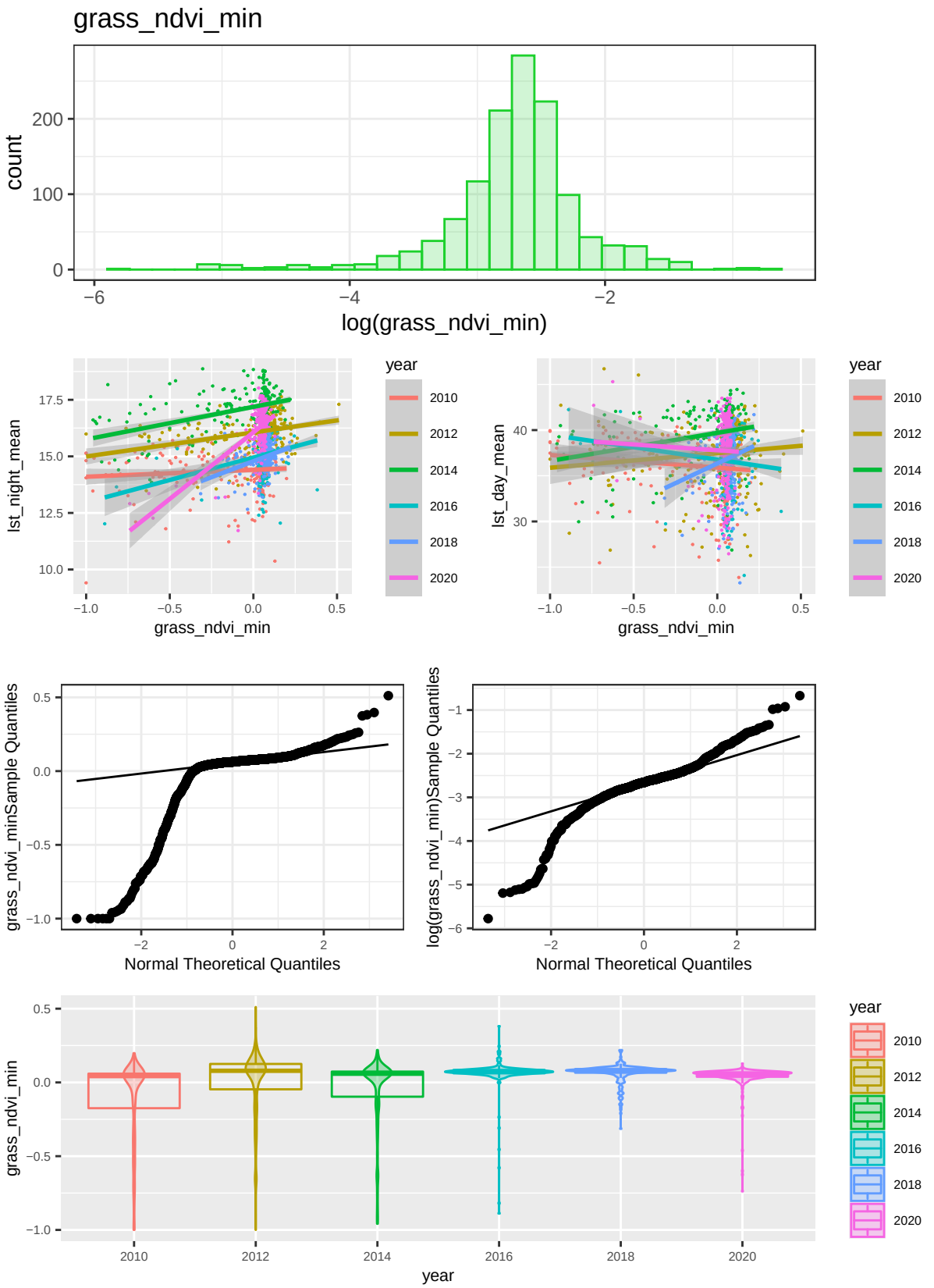
Google Earth: Tensor Flow Model

Lawn NDVI Mean



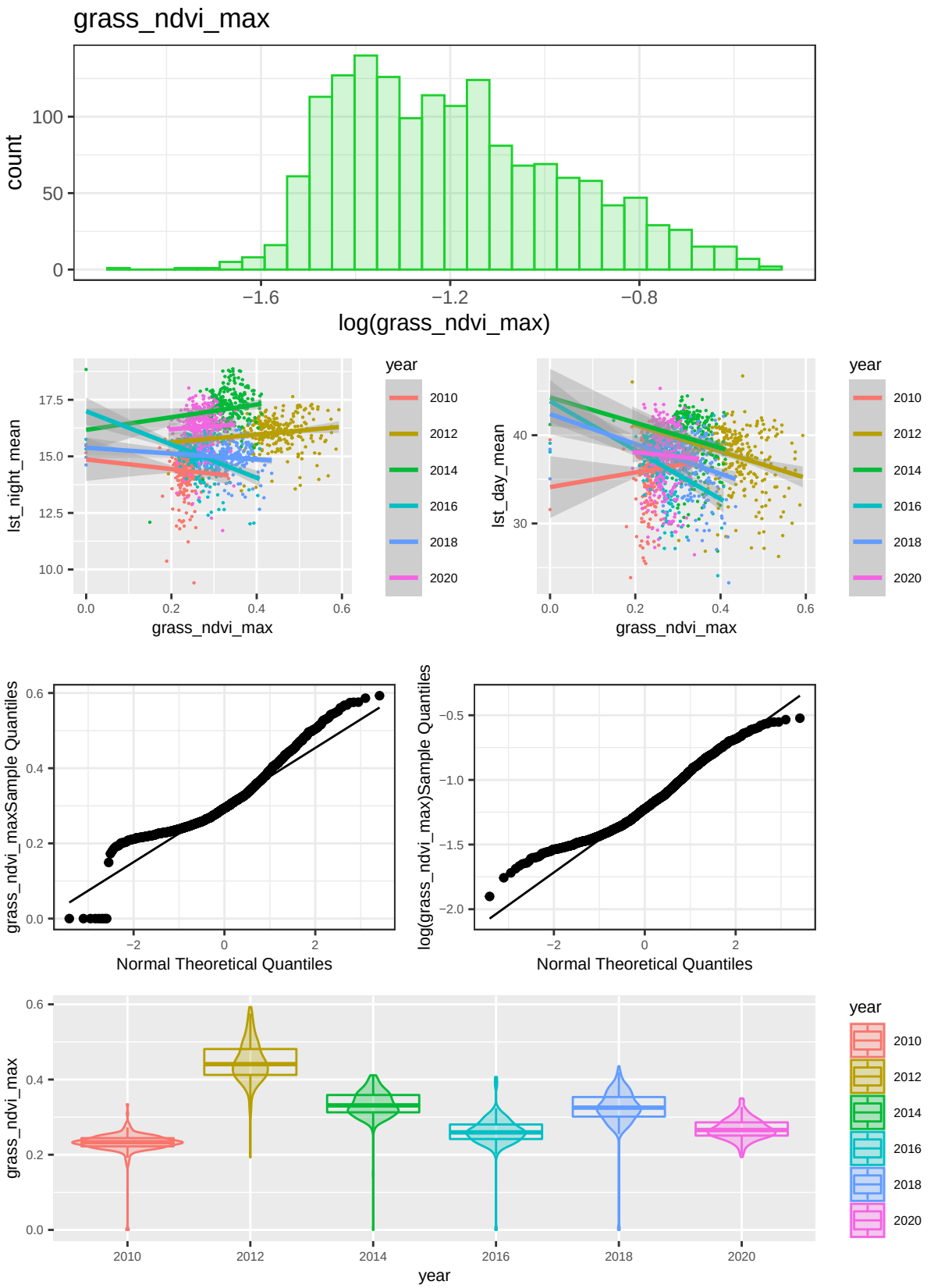
Google Earth: Tensor Flow Model

Lawn NDVI Min



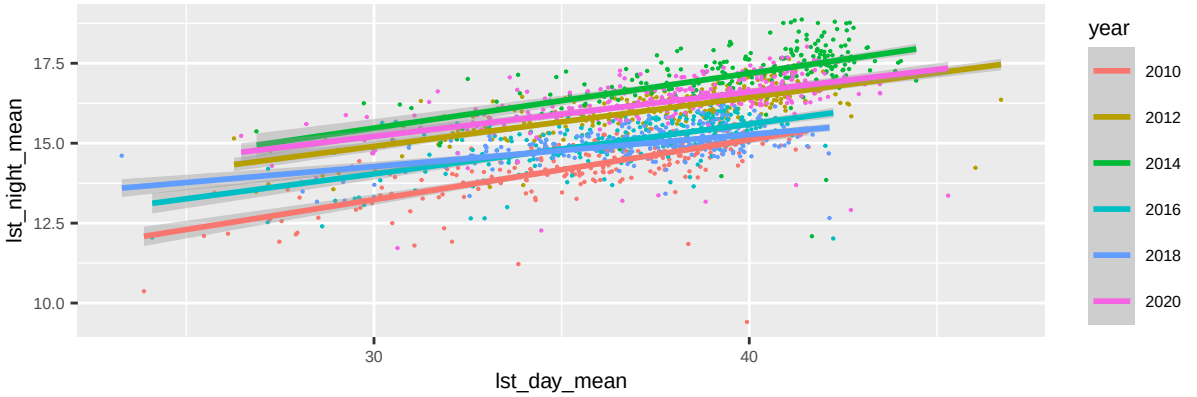
Google Earth: Tensor Flow Model

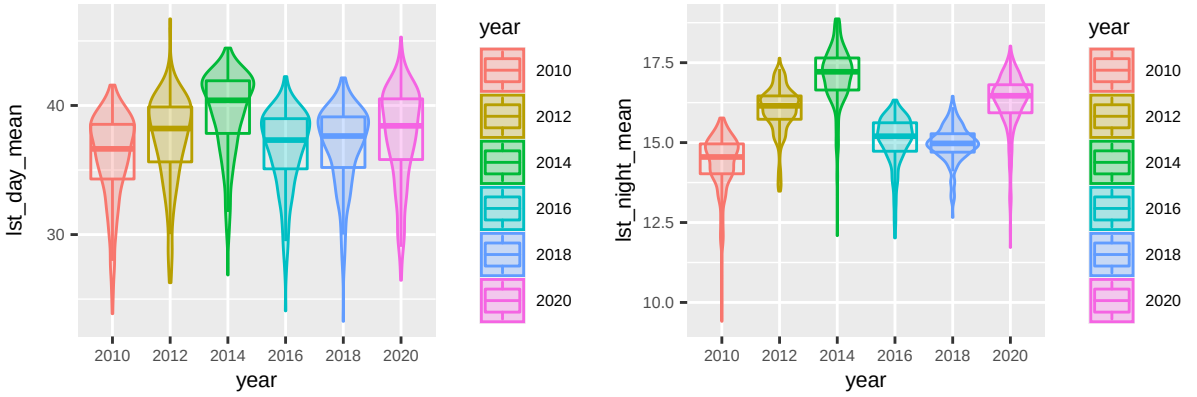
Lawn NDVI Max

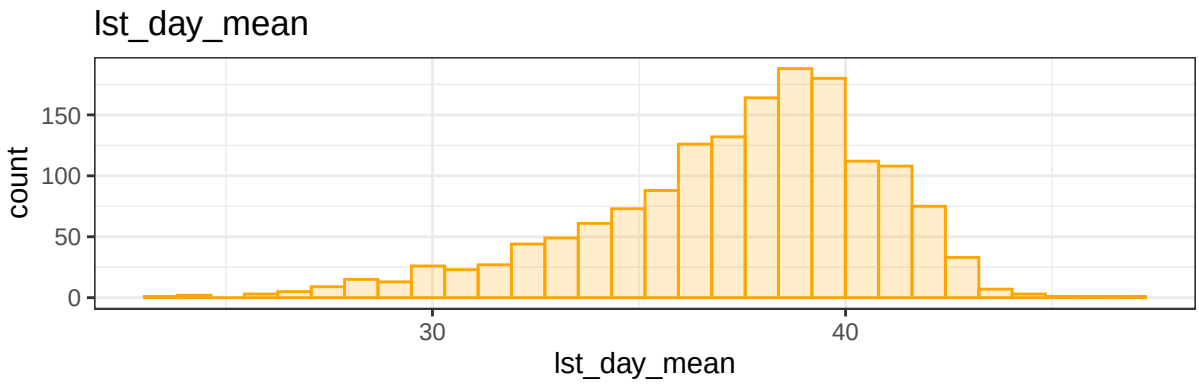


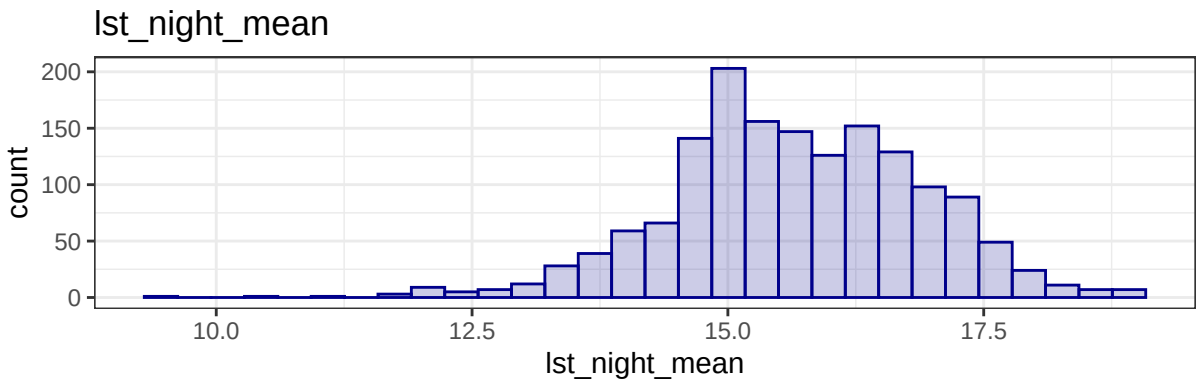
Google Earth: Tensor Flow Model

Mean Temperature









Google Earth: Tensor Flow Model

Create MicroClimate Panel Data Set

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

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: Sun, Jul 31, 2022 - 3:23:43 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')

saveRDS(feM1, "TFfixedEffectsDay.yml")

saveRDS(feM2, "TFfixedEffectsNight.yml")
```

Table 1: OLS Model

	Dependent variable:	
	lst_day_mean OLSModel Day Temp	lst_night_mean OLSModel Night Temp
	(1)	(2)
factor(year)2014	41.381*** (0.933)	19.129*** (0.244)
factor(year)2016	38.161*** (0.872)	16.958*** (0.228)
factor(year)2018	38.422*** (0.946)	16.923*** (0.247)
factor(year)2020	40.618*** (0.880)	18.445*** (0.230)
tree_area	−10.280*** (0.569)	−1.680*** (0.149)
lawn_area	−6.351*** (2.233)	−0.682 (0.584)
log(water_area)	−0.036*** (0.013)	−0.004 (0.003)
log(soil_area)	0.017 (0.033)	−0.007 (0.009)
log(turf_area)	0.002 (0.014)	0.012*** (0.004)
grass_ndvi_mean	−2.834 (2.761)	−1.569** (0.722)
tree_ndvi_mean	2.178 (2.676)	−1.049 (0.699)
Observations	1,057	1,057
R^2	0.995	0.998
Adjusted R^2	0.995	0.998
Residual Std. Error (df = 1046)	2.694	0.704
F Statistic (df = 11; 1046)	18,866.470***	48,978.310***
Note:		*p<0.1; **p<0.05; ***p<0.01

```
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: Sun, Jul 31, 2022 - 3:23:43 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	1.446*** (0.070)	0.616*** (0.040)
factor(year)2016	−1.489*** (0.071)	−1.303*** (0.041)
factor(year)2018	−1.267*** (0.072)	−1.460*** (0.041)
tree_area	−0.470 (0.286)	−0.332** (0.164)
lawn_area	−2.424*** (0.657)	0.720* (0.377)
log(water_area)	0.002 (0.004)	0.004* (0.002)
log(soil_area)	0.001 (0.009)	0.003 (0.005)
log(turf_area)	0.002 (0.004)	0.005** (0.002)
grass_ndvi_mean	0.952 (0.729)	0.377 (0.418)
tree_ndvi_mean	0.301 (0.671)	−0.491 (0.385)
Observations	1,057	1,057
R^2	0.855	0.917
Adjusted R^2	0.799	0.884
F Statistic (df = 10; 759)	448.774***	836.577***
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 = 83.424, 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(
  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 = 'random')

reM2 <- 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 = '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: Sun, Jul 31, 2022 - 3:23:45 PM

Table 3: Random Effects Model

	Dependent variable:	
	lst_day_mean RandomEffects Day Temp	lst_night_mean RandomEffects Night Temp
	(1)	(2)
factor(year)2014	38.921*** (0.323)	17.696*** (0.162)
factor(year)2016	35.971*** (0.308)	15.706*** (0.153)
factor(year)2018	36.173*** (0.324)	15.579*** (0.163)
factor(year)2020	37.566*** (0.313)	17.087*** (0.155)
tree_area	−1.572*** (0.296)	−0.977*** (0.147)
lawn_area	−3.059*** (0.703)	0.318 (0.393)
log(water_area)	−0.0001 (0.004)	0.001 (0.002)
log(soil_area)	0.0004 (0.010)	0.0002 (0.006)
log(turf_area)	0.001 (0.004)	0.006** (0.002)
grass_ndvi_mean	0.598 (0.784)	−0.367 (0.448)
tree_ndvi_mean	0.592 (0.723)	−0.421 (0.417)
Observations	1,057	1,057
R^2	0.820	0.898
Adjusted R^2	0.819	0.897
F Statistic	55,980.430***	165,588.100***
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 = 1888.8, 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(
  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')

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: Sun, Jul 31, 2022 - 3:23:46 PM

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

```
##
##  F test for individual effects
##
## data:  lst_day_mean ~ 0 + factor(year) + tree_area + lawn_area + log(water_area) +  ...
## F = NaN, df1 = 0, df2 = 759, p-value = NA
## 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 = 2.003, df = 1, p-value = 0.157
## alternative hypothesis: significant effects
```

Table 4:

	Dependent variable:
	lst_day_mean
factor(year)2014	1.446*** (0.070)
factor(year)2016	−1.489*** (0.071)
factor(year)2018	−1.267*** (0.072)
tree_area	−0.470 (0.286)
lawn_area	−2.424*** (0.657)
log(water_area)	0.002 (0.004)
log(soil_area)	0.001 (0.009)
log(turf_area)	0.002 (0.004)
grass_ndvi_mean	0.952 (0.729)
tree_ndvi_mean	0.301 (0.671)
Observations	1,057
R^2	0.855
Adjusted R^2	0.799
F Statistic	448.774*** (df = 10; 759)
Note:	*p<0.1; **p<0.05; ***p<0.01

```
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 = 63.96, df = 1, p-value = 1.269e-15
## alternative hypothesis: serial correlation in idiosyncratic errors
```

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

```
##
## t test of coefficients:
##
##                      Estimate  Std. Error  t value  Pr(>|t|)
## factor(year)2014  1.44554180  0.07522295  19.2168 < 2.2e-16 ***
## factor(year)2016 -1.48852424  0.07962260 -18.6947 < 2.2e-16 ***
## factor(year)2018 -1.26725732  0.07330603 -17.2872 < 2.2e-16 ***
## tree_area        -0.47047382  0.29350404  -1.6030 0.1093606
## lawn_area        -2.42438974  0.66414662  -3.6504 0.0002798 ***
## log(water_area)   0.00228291  0.00333302   0.6849 0.4935934
## log(soil_area)    0.00088539  0.00732982   0.1208 0.9038876
## log(turf_area)    0.00170148  0.00368336   0.4619 0.6442589
## grass_ndvi_mean   0.95245403  0.62772823   1.5173 0.1296065
## tree_ndvi_mean    0.30131761  0.59217533   0.5088 0.6110180
## ---
```

```
## 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  1.44554180  0.07522295  19.2168 < 2.2e-16 ***
## factor(year)2016 -1.48852424  0.07962260 -18.6947 < 2.2e-16 ***
## factor(year)2018 -1.26725732  0.07330603 -17.2872 < 2.2e-16 ***
## tree_area        -0.47047382  0.29350404  -1.6030 0.1093606
## lawn_area        -2.42438974  0.66414662  -3.6504 0.0002798 ***
## log(water_area)   0.00228291  0.00333302   0.6849 0.4935934
## log(soil_area)    0.00088539  0.00732982   0.1208 0.9038876
## log(turf_area)    0.00170148  0.00368336   0.4619 0.6442589
## grass_ndvi_mean   0.95245403  0.62772823   1.5173 0.1296065
## tree_ndvi_mean    0.30131761  0.59217533   0.5088 0.6110180
## ---
## 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       0.07522295       0.07962260       0.07330603 0.2935040 0.6641466
## HC1       0.07558133       0.08000194       0.07365528 0.2949023 0.6673107
## HC2       0.07598188       0.08023038       0.07392496 0.2976493 0.6722953
## HC3       0.07676684       0.08084975       0.07456335 0.3019512 0.6806513
## HC4       0.07748397       0.08102597       0.07502642 0.3080608 0.6897522
##     log(water_area) log(soil_area) log(turf_area) grass_ndvi_mean
## HC0     0.003333023    0.007329821    0.003683361       0.6277282
## HC1     0.003348902    0.007364742    0.003700909       0.6307188
## HC2     0.003352529    0.007491388    0.003705880       0.6345154
## HC3     0.003372337    0.007659104    0.003728814       0.6415432
## HC4     0.003369895    0.007972571    0.003727963       0.6488436
##     tree_ndvi_mean
## HC0      0.5921753
## HC1      0.5949966
## HC2      0.5974213
## HC3      0.6028150
## HC4      0.6068934
```

```
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  1.44554180  0.07522295  19.2168 < 2.2e-16 ***
## factor(year)2016 -1.48852424  0.07962260 -18.6947 < 2.2e-16 ***
## factor(year)2018 -1.26725732  0.07330603 -17.2872 < 2.2e-16 ***
## tree_area        -0.47047382  0.29350404  -1.6030 0.1093606
## lawn_area        -2.42438974  0.66414662  -3.6504 0.0002798 ***
## log(water_area)   0.00228291  0.00333302   0.6849 0.4935934
## log(soil_area)    0.00088539  0.00732982   0.1208 0.9038876
## log(turf_area)    0.00170148  0.00368336   0.4619 0.6442589
## grass_ndvi_mean   0.95245403  0.62772823   1.5173 0.1296065
## tree_ndvi_mean    0.30131761  0.59217533   0.5088 0.6110180
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
print(cInterval)
```

```
##                          2.5 %       97.5 %
## factor(year)2014  1.297872041  1.593211562
## factor(year)2016 -1.644830921 -1.332217550
## factor(year)2018 -1.411163981 -1.123350653
## tree_area        -1.046649950  0.105702315
## lawn_area        -3.728172253 -1.120607230
## log(water_area)  -0.004260132  0.008825946
## log(soil_area)   -0.013503746  0.015274516
## log(turf_area)   -0.005529305  0.008932266
## grass_ndvi_mean  -0.279835746  2.184743802
## tree_ndvi_mean   -0.861178472  1.463813687
```

```
totalRobust <- coeftest(feM2, vcov = vcovHC(feM1, type = 'HC0'))
cInterval <- coefci(feM2, vcov. = vcovHC(feM1, type = 'HC0'))
```

```
print(totalRobust)
```

```
##
## t test of coefficients:
##
##                      Estimate Std. Error  t value  Pr(>|t|)
## factor(year)2014  0.6156506  0.0752230   8.1843 1.154e-15 ***
## factor(year)2016 -1.3029103  0.0796226 -16.3636 < 2.2e-16 ***
## factor(year)2018 -1.4595569  0.0733060 -19.9105 < 2.2e-16 ***
## tree_area        -0.3315523  0.2935040  -1.1296    0.2590
## lawn_area         0.7199097  0.6641466   1.0840    0.2787
## log(water_area)   0.0035256  0.0033330   1.0578    0.2905
## log(soil_area)    0.0034717  0.0073298   0.4736    0.6359
## log(turf_area)    0.0049625  0.0036834   1.3473    0.1783
## grass_ndvi_mean   0.3768448  0.6277282   0.6003    0.5485
## tree_ndvi_mean   -0.4906334  0.5921753  -0.8285    0.4076
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
print(cInterval)
```

```
##                          2.5 %      97.5 %
## factor(year)2014  0.467980799  0.76332032
## factor(year)2016 -1.459217017 -1.14660365
## factor(year)2018 -1.603463567 -1.31565024
## tree_area        -0.907728472  0.24462379
## lawn_area        -0.583872835  2.02369219
## log(water_area)  -0.003017404  0.01006867
## log(soil_area)   -0.010917399  0.01786086
## log(turf_area)   -0.002268249  0.01219332
## grass_ndvi_mean  -0.855445012  1.60913454
## tree_ndvi_mean   -1.653129456  0.67186270
```

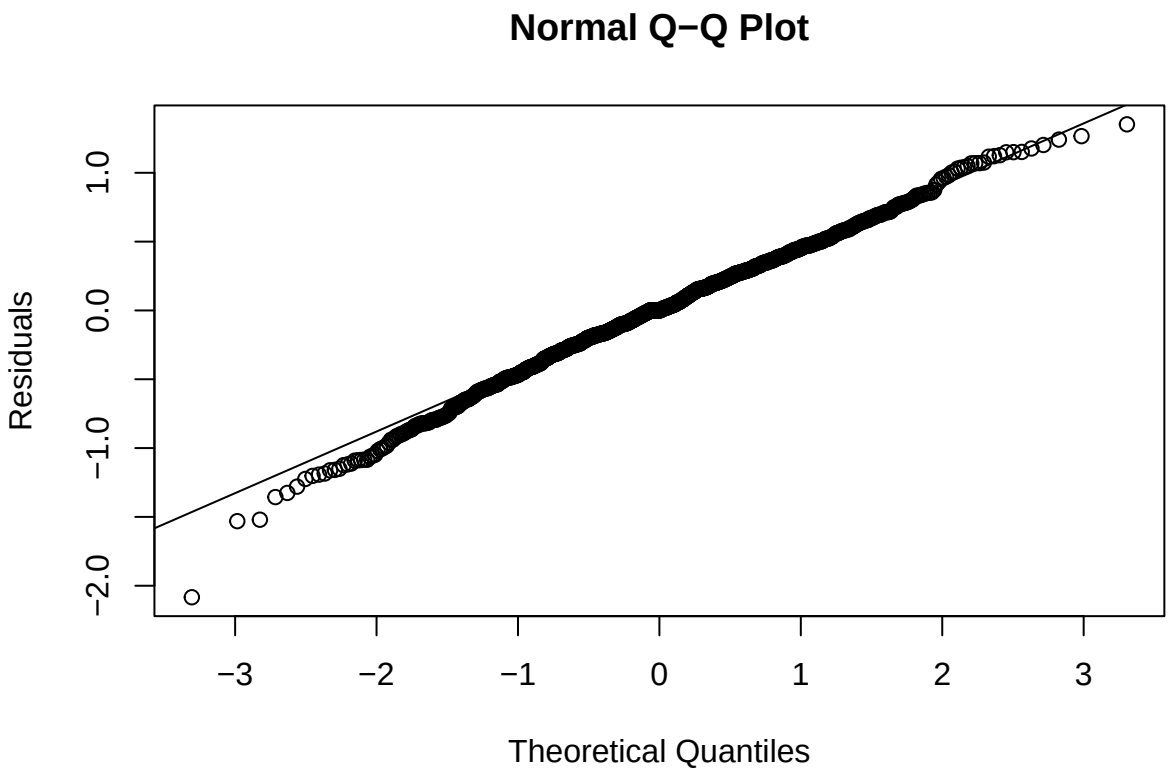
```
stargazer::stargazer(OLSM1,feM1,reM1,feM2FixedTime,
                     font.size = 'tiny',
                     title = 'Regression Models Sumamry',
                     column.labels = c("OLS", "FixedEffects",
                                       "RandomEffects",
                                       "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: Sun, Jul 31, 2022 - 3:23:48 PM

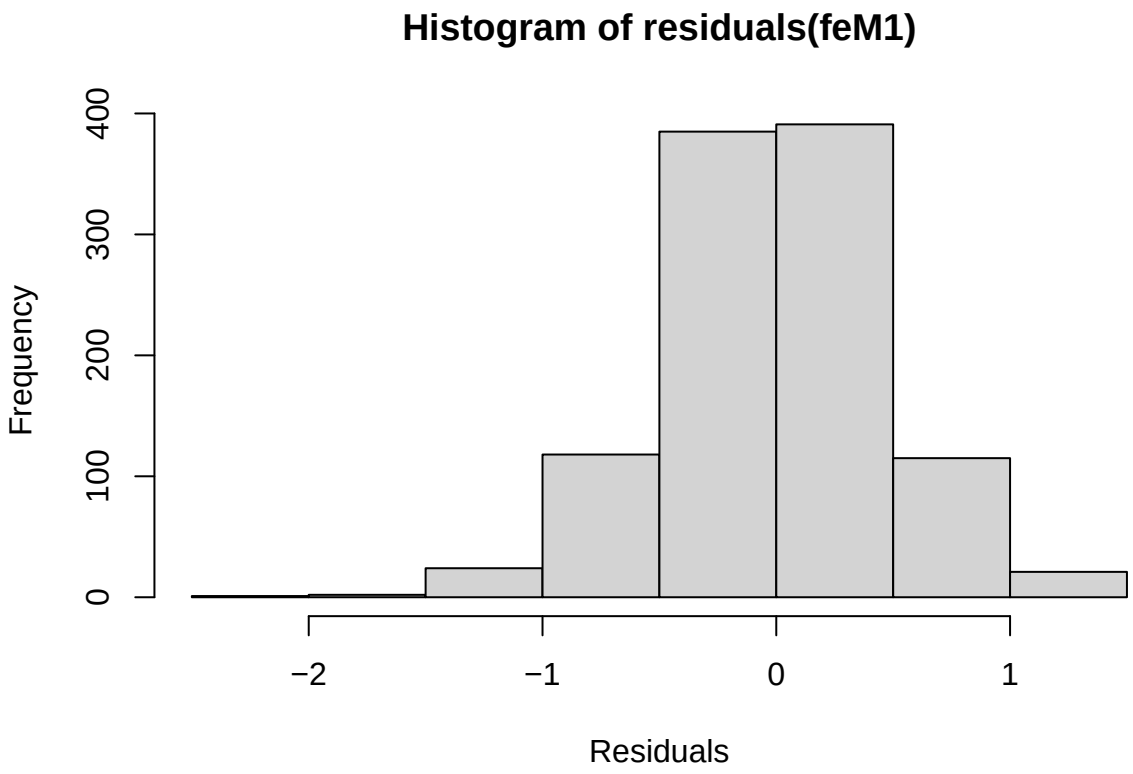
```
qqnorm(residuals(feM1), ylab = 'Residuals')
qqline(residuals(feM1))
```

Table 5: Regression Models Sumamry

	Dependent variable:			
	lst_day_mean			
	OLS		*panel linear*	
	OLS	FixedEffects	RandomEffects	FixedEffectsFixedTime
	(1)	(2)	(3)	(4)
factor(year)2014	41.381***	1.446***	38.921***	1.446***
	(0.933)	(0.070)	(0.323)	(0.070)
factor(year)2016	38.161***	−1.489***	35.971***	−1.489***
	(0.872)	(0.071)	(0.308)	(0.071)
factor(year)2018	38.422***	−1.267***	36.173***	−1.267***
	(0.946)	(0.072)	(0.324)	(0.072)
factor(year)2020	40.618***		37.566***	
	(0.880)		(0.313)	
tree_area	−10.280***	−0.470	−1.572***	−0.470
	(0.569)	(0.286)	(0.296)	(0.286)
lawn_area	−6.351***	−2.424***	−3.059***	−2.424***
	(2.233)	(0.657)	(0.703)	(0.657)
log(water_area)	−0.036***	0.002	−0.0001	0.002
	(0.013)	(0.004)	(0.004)	(0.004)
log(soil_area)	0.017	0.001	0.0004	0.001
	(0.033)	(0.009)	(0.010)	(0.009)
log(turf_area)	0.002	0.002	0.001	0.002
	(0.014)	(0.004)	(0.004)	(0.004)
grass_ndvi_mean	−2.834	0.952	0.598	0.952
	(2.761)	(0.729)	(0.784)	(0.729)
tree_ndvi_mean	2.178	0.301	0.592	0.301
	(2.676)	(0.671)	(0.723)	(0.671)
Observations	1,057	1,057	1,057	1,057
R^2	0.995	0.855	0.820	0.855
Adjusted R^2	0.995	0.799	0.819	0.799
Residual Std. Error	2.694 (df = 1046)			
F Statistic	18,866.470*** (df = 11; 1046)	448.774*** (df = 10; 759)	55,980.430***	448.774*** (df = 10; 759)
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
−2
−1
0
1
Residuals