

Great Smoky Mountains Phenology Data Demo

USA-NPN

2022-12-13

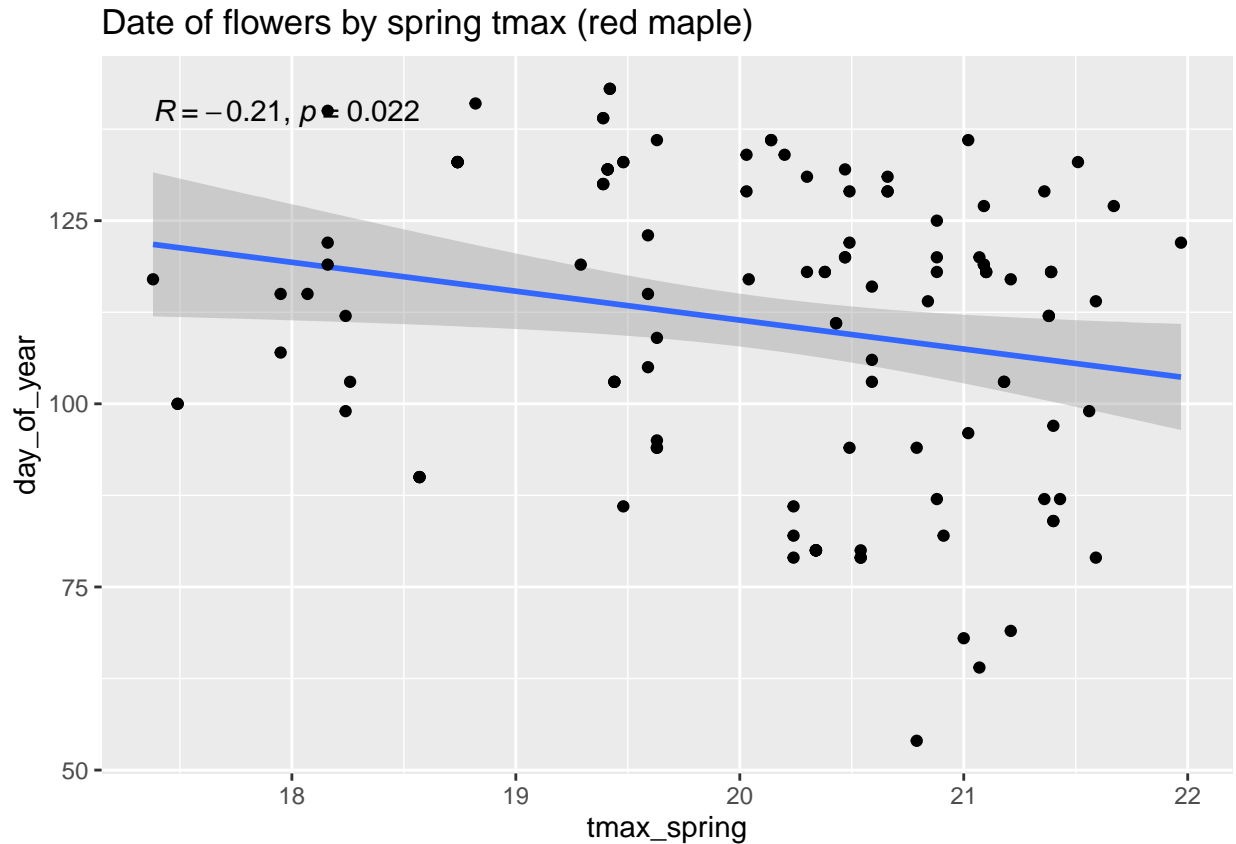
Introduction

This is a demonstration/pilot exploration of the Great Smoky Mountains phenology data, focusing on flowering in red maple. The idea is to start to lay code and data analysis framework to make reporting easy for any phenology project.

The code behind this PDF does the following:

1. Downloads status data for GRSM, 2011-2022, for red maple, flowers/flower buds phenophase, with the Daymet climate data
2. Formats data and selects the earliest yes in each year (“first flower date” or FFD), and removes any FFDs that fall after day 182 (to focus on spring flowering only)
3. Removes individual plants with fewer than five years of observations.
4. Removes first flower dates that are outliers, defined as falling outside 1.5 times the Interquartile Range for an individual plant.
5. Models the relationship between spring maximum temperatures and first flower date, and then looks for randomly distributed residuals using a Shapiro-Wilk test. As the residuals are not randomly distributed, it is likely that a non-parametric test should be used instead.
6. Models the relationship between year and first flower date (trend), and then looks for randomly distributed residuals using a Shapiro-Wilk test. As the residuals are not randomly distributed, it is likely that a non-parametric test should be used instead.

Plot 1: Relationship between spring maximum temperatures and first flower date



Model parameters

```
model <- lm(day_of_year~tmax_spring, data = df2)
summary(model)
```

```
##
## Call:
## lm(formula = day_of_year ~ tmax_spring, data = df2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -54.309 -18.835   6.019  16.823  29.284
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   190.355     34.151   5.574 1.54e-07 ***
## tmax_spring    -3.946     1.696  -2.326  0.0217 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.13 on 121 degrees of freedom
```

```
## (15 observations deleted due to missingness)
## Multiple R-squared:  0.04281,    Adjusted R-squared:  0.0349
## F-statistic: 5.411 on 1 and 121 DF,  p-value: 0.02167
```

Shapiro-Wilk test

```
#check for normality in distribution of residuals
shapiro.test(model$residuals)
```

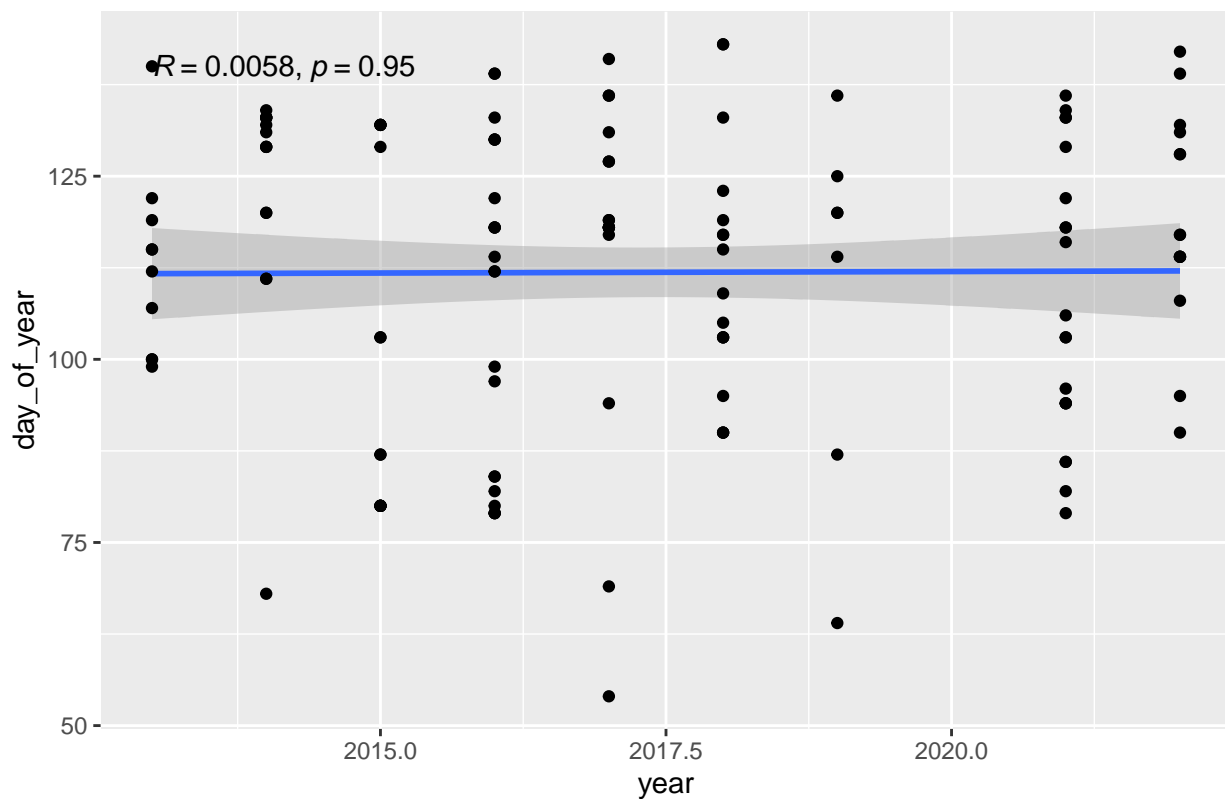
```
##
## Shapiro-Wilk normality test
##
## data:  model$residuals
## W = 0.93063, p-value = 8.317e-06
```

```
#doesn't pass this test
```

We see some evidence of relationship, with warmer spring max temps associated with earlier appearance of flowers/flower buds. The estimate is -0.04 days per degree spring maximum temperature. However, we don't see normally distributed residuals, so we likely need to look at non-parametric options for this kind of relationship.

Plot 2: Relationship between year and first flower date

Onset of flowers/buds over the period of record in Red Maple



Model parameters

```
model2 <- lm(day_of_year~year, data = df2)
summary(model2)

##
## Call:
## lm(formula = day_of_year ~ year, data = df2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -57.862 -15.725   5.015  17.261  31.097
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.925e+01  1.224e+03   0.024   0.981
## year         4.096e-02  6.069e-01   0.067   0.946
##
## Residual standard error: 20.12 on 136 degrees of freedom
## Multiple R-squared:  3.349e-05, Adjusted R-squared:  -0.007319
## F-statistic: 0.004555 on 1 and 136 DF,  p-value: 0.9463
```

Shapiro-Wilk test

```
#check for normality in distribution of residuals
shapiro.test(model2$residuals)

##
##  Shapiro-Wilk normality test
##
## data:  model2$residuals
## W = 0.94524, p-value = 2.923e-05
```

We do not see evidence of a trend in first flower date over time. Again, the residuals are not randomly distributed, pointing to applying a non-parametric test.