Leaf Color Change Date in Harvard Forest

Kevin Buck, Kyle Elliott, Sydney Higgins

2023-04-29

####Introduction#### Researcher John O'Keefe and his team starting in 1991 collected various statistics for a number of species of trees in the Prospect Hill Tract Harvard Forest in Petersham, MA. He has continued to record the statistics and compile them in an online data archive in order to provide researchers with data to investigate if certain trends are correlated with certain species. In this data archive, one particular data set has statistics regarding the mean day of the year in which each tree of various given species has 50% of their leaves change color. We will be investigating how this day (listed in Meridian days within the data set) is correlated with species type and investigate the trends for each species throughout the last 30 years.

###Data Acquisition### O'Keefe and his team's data was collected through weekly observations of percent leaf coloration and percent leaf fall started in September each year. Researchers observed when 50% of the leaves of trees of different species had changed color and started to fall, and recorded those dates. They recorded this data for 33 unique types of species from 1991 to the present day, with the most recent data coming from 2022.

####Analysis#### Upon piping our desired data set, we filtered down to the 14 species that had recorded mean meridian dates for 50% of the leaves changing colors for each of the tree species. To assess the changes in the meridian date of leaf color change in each species per year, we first created a scatter plot of the meridian date vs. year and had each species of tree which were represented by the different colors.

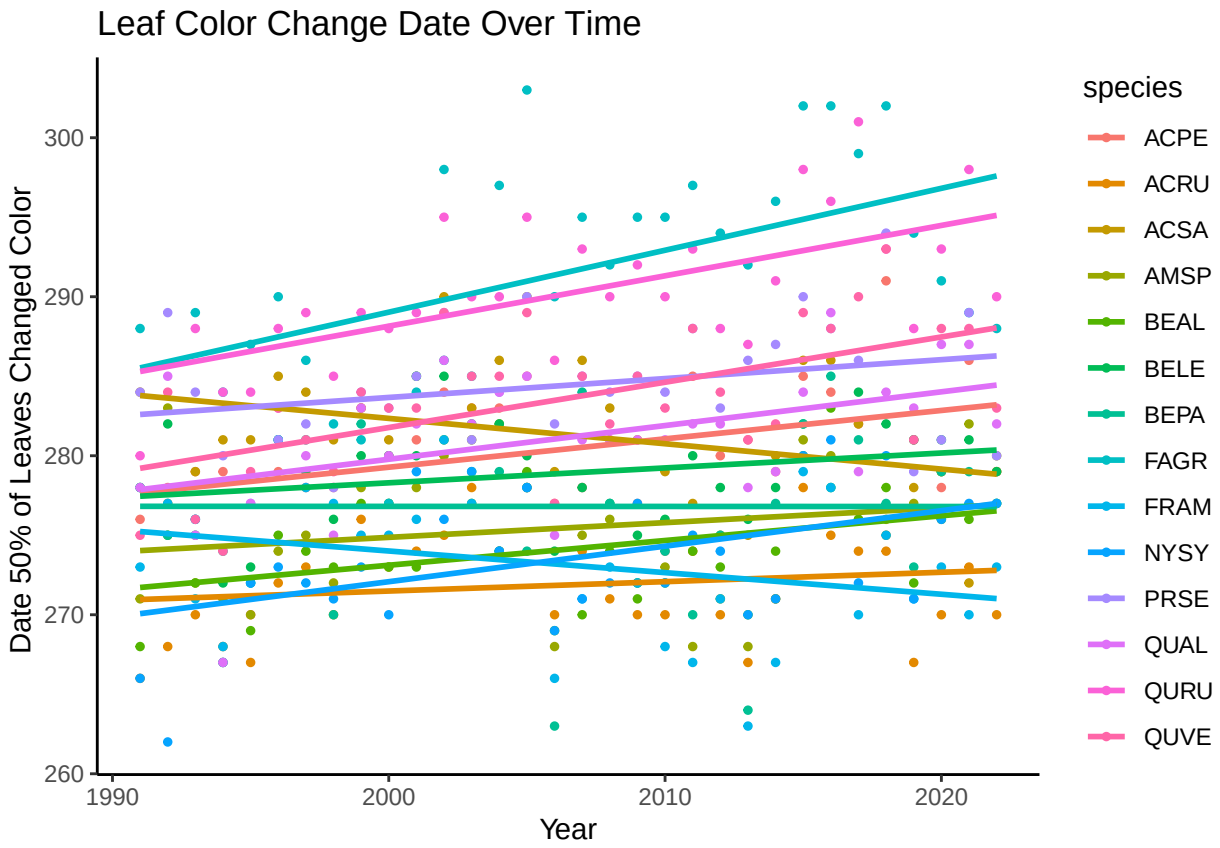
how did we choose to analyze the data with our given question Data visualization

```
pheno <- read.csv(file="hf003-08-fall-mean-spp.csv") #read in the data
pheno_trimmed <- pheno %>% #pipe in data
  #filter in species present for the whole dataset
  filter(species=="ACRU"|species=="ACSA"|species=="ACPE"|species=="AMSP"|
           species=="BEAL"|species=="BELE"|species=="BEPA"|species=="FAGR"|
           species=="FRAM"|species=="NYSY"|species=="PRSE"|species=="QUAL"|
           species=="QURU"|species=="QUVE")

pheno_trimmed %>%
  ggplot(aes(x = year, y = lc_doy)) + #set year as x axis and julian day of change as the y
  geom_point(aes(color=species, fill=species), size=1) +
  ggtitle("Leaf Color Change Date Over Time")+ #add title
  #adding in trend lines by species
  geom_smooth(aes(group=species,color=species), method = "lm", formula = y ~ x, se = F)+
  xlab("Year") + #label x axis
  ylab("Date 50% of Leaves Changed Color ") + #label y axis
  theme_classic() #changing background
```

```
## Warning: Removed 2 rows containing non-finite values (`stat_smooth()`).
```

```
## Warning: Removed 2 rows containing missing values (`geom_point()`).
```



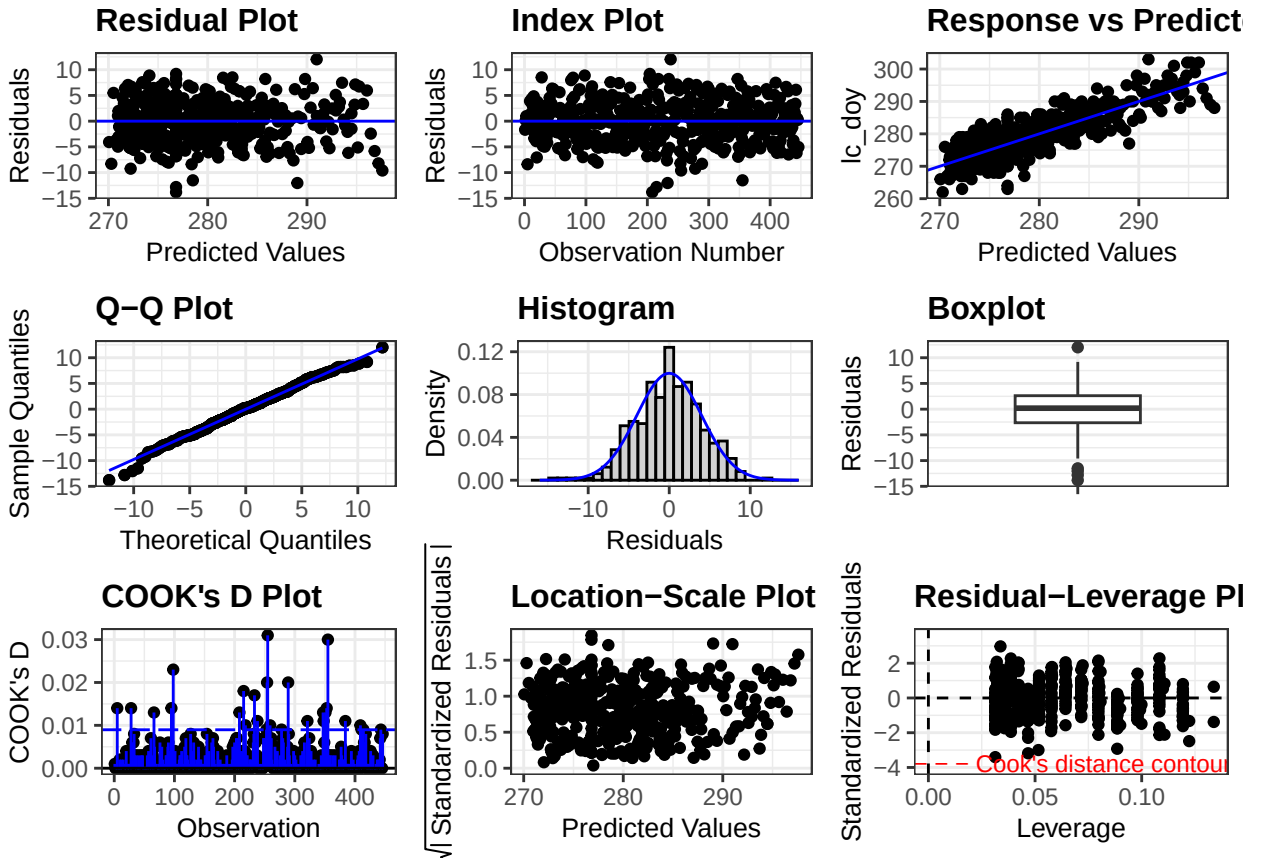
ANCOVA

```
options(contrasts = c("contr.sum","contr.ply")) #options fo later post hoc comparison
model_1 <- lm(lc_doy ~ year * species,data=pheno_trimmed) #build linear model
summary(model_1) #view linear model
```

```
##
## Call:
## lm(formula = lc_doy ~ year * species, data = pheno_trimmed)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.8127  -2.6427   0.1924   2.6173  12.0243
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      17.68169   42.57334   0.415 0.678119
## year              0.13061    0.02122   6.156 1.75e-09 ***
## species1        -95.44268  152.62876  -0.625 0.532099
## species2        135.77451  152.62876   0.890 0.374207
## species3        583.58224  152.62876   3.824 0.000152 ***
## species4         72.06813  152.62876   0.472 0.637045
## species5        -55.41739  152.62876  -0.363 0.716724
## species6         73.29907  152.62876   0.480 0.631305
## species7        259.86633  152.62876   1.703 0.089386 .
## species8       -507.42989  158.62931  -3.199 0.001485 **
## species9        528.32143  152.62876   3.461 0.000593 ***
## species10      -192.45020  152.62876  -1.261 0.208047
```

```
## species11        28.44716  152.62876   0.186 0.852236
## species12      -162.02406  152.62876  -1.062 0.289050
## species13      -363.03852  158.62931  -2.289 0.022601 *
## year:species1     0.04791    0.07607   0.630 0.529169
## year:species2    -0.07159    0.07607  -0.941 0.347136
## year:species3    -0.29007    0.07607  -3.813 0.000158 ***
## year:species4    -0.03805    0.07607  -0.500 0.617146
## year:species5     0.02481    0.07607   0.326 0.744437
## year:species6    -0.03695    0.07607  -0.486 0.627352
## year:species7    -0.13098    0.07607  -1.722 0.085826 .
## year:species8     0.25878    0.07904   3.274 0.001149 **
## year:species9    -0.26661    0.07607  -3.505 0.000506 ***
## year:species10    0.09281    0.07607   1.220 0.223100
## year:species11   -0.01184    0.07607  -0.156 0.876338
## year:species12    0.08145    0.07607   1.071 0.284901
## year:species13    0.18614    0.07904   2.355 0.018983 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.121 on 418 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.7089, Adjusted R-squared:  0.6901
## F-statistic: 37.71 on 27 and 418 DF,  p-value: < 2.2e-16
```

```
resid_panel(model_1,plot="all") #check on residuals
```



```
Anova(model_1, type = 3) #run an ANCOVA
```

```
## Anova Table (Type III tests)
##
## Response: lc_doy
##              Sum Sq  Df F value    Pr(>F)
## (Intercept)     2.9   1  0.1725    0.6781
## year          643.6   1 37.8977 1.750e-09 ***
## species       842.7  13  3.8169 7.958e-06 ***
## year:species  861.6  13  3.9027 5.367e-06 ***
## Residuals    7098.7 418
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
pwc <- pheno_trimmed %>% tukey_hsd(lc_doy ~ species) #pair wise comparisons
pwc #look at the PWC of species
```

```
## # A tibble: 91 x 9
##    term    group1 group2 null.value estimate conf.low conf.high        p.adj
##  * <chr>   <chr>  <chr>       <dbl>    <dbl>    <dbl>     <dbl>        <dbl>
##  1 species ACPE   ACRU            0   -8.56   -12.3      -4.80  0
##  2 species ACPE   ACSA            0    0.875   -2.88      4.63  1
##  3 species ACPE   AMSP            0   -4.97    -8.73     -1.21  0.000868
##  4 species ACPE   BEAL            0   -6.31   -10.1      -2.55  0.0000024
##  5 species ACPE   BELE            0   -1.53    -5.29      2.23  0.984
##  6 species ACPE   BEPA            0   -3.63    -7.38      0.133 0.0714
##  7 species ACPE   FAGR            0   11.3      7.52     15.1   0
##  8 species ACPE   FRAM            0   -7.31   -11.1      -3.55  0.0000000137
##  9 species ACPE   NYSY            0   -6.91   -10.7      -3.15  0.00000012
## 10 species ACPE   PRSE            0    4.00     0.242     7.76  0.0249
## # ... with 81 more rows, and 1 more variable: p.adj.signif <chr>
```

```
#pwc %>%
  #ggplot(aes(x = species, y = lc_doy)) + #set year as x axis and julian day of change as the y
  #geom_bar(aes(color=species, fill=species), size=2) +
  #ggtitle("Leaf Color Change Date Over Time")+ #add title
  #xlab("Species") + #label x axis
  #ylab("Date 50% of Leaves Changed Color ") + #label y axis
  #theme_classic() #changing background
```

###Results### What do graphs show Anova

###Conclusion### Conclusion from anova

###Contribution Statement### We extend our appreciation to everyone who made this project possible, specifically Mr. O'Keefe and his team at Harvard that collected the data that we were able to use as well as Dr. McLachlan and his team of TAs that helped to advise and educate us on biostatistics. This project would not have been possible without your work as well, so we thank you for making this opportunity for our group!

###Citations### O'Keefe J. 2023. Phenology of Woody Species at Harvard Forest since 1990. Harvard Forest Data Archive: HF003 (v.35). Environmental Data Initiative: https://doi.org/10.6073/pasta/eb0dd36c6ec62a918340b6bda38be832.