LauraTraderscode_testedonPSME—Canopy.R

edeegan

2023-11-08

```
# Add packages here
library(knitr)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggplot2)
library(car)
```

```
## Loading required package: carData
##
## Attaching package: 'car'
##
## The following object is masked from 'package:dplyr':
##
##     recode
##
## The following object is masked from 'package:purrr':
##
##     some
```

```
library(dplyr)
library(car)
library(broom)
library(emmeans)
library(lmerTest)
```

```
## Loading required package: lme4
## Loading required package: Matrix
##
## Attaching package: 'Matrix'
```

```
##
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
##
##
## Attaching package: 'lmerTest'
##
## The following object is masked from 'package:lme4':
##
##     lmer
##
## The following object is masked from 'package:stats':
##
##     step
```

```
library(lme4)
library(pbkrtest)

canopy <- read_csv("C:/Users/edeegan/OneDrive - DOI/Fire_project/Fire_project/PMSE_data/PSME_Cover - Sp
```

```
## New names:
## Rows: 948 Columns: 24
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (14): MacroPlot Name, Monitoring Status, Date, Status, Comment, UV1, UV2... dbl
## (5): Index, Cover, Height, UV3, MinCovLevel lgl (5): SizeCl, AgeCl, Area,
## Visited, ...24
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `` -> `...24`
```

```
canopy<-as_tibble(canopy)
canopy=rename(canopy, plot=`MacroPlot Name`)

canopy=canopy %>% separate(Date, c("month", "day", "year"), "/")
canopy=canopy %>% separate(year, c("year"), " ")
```

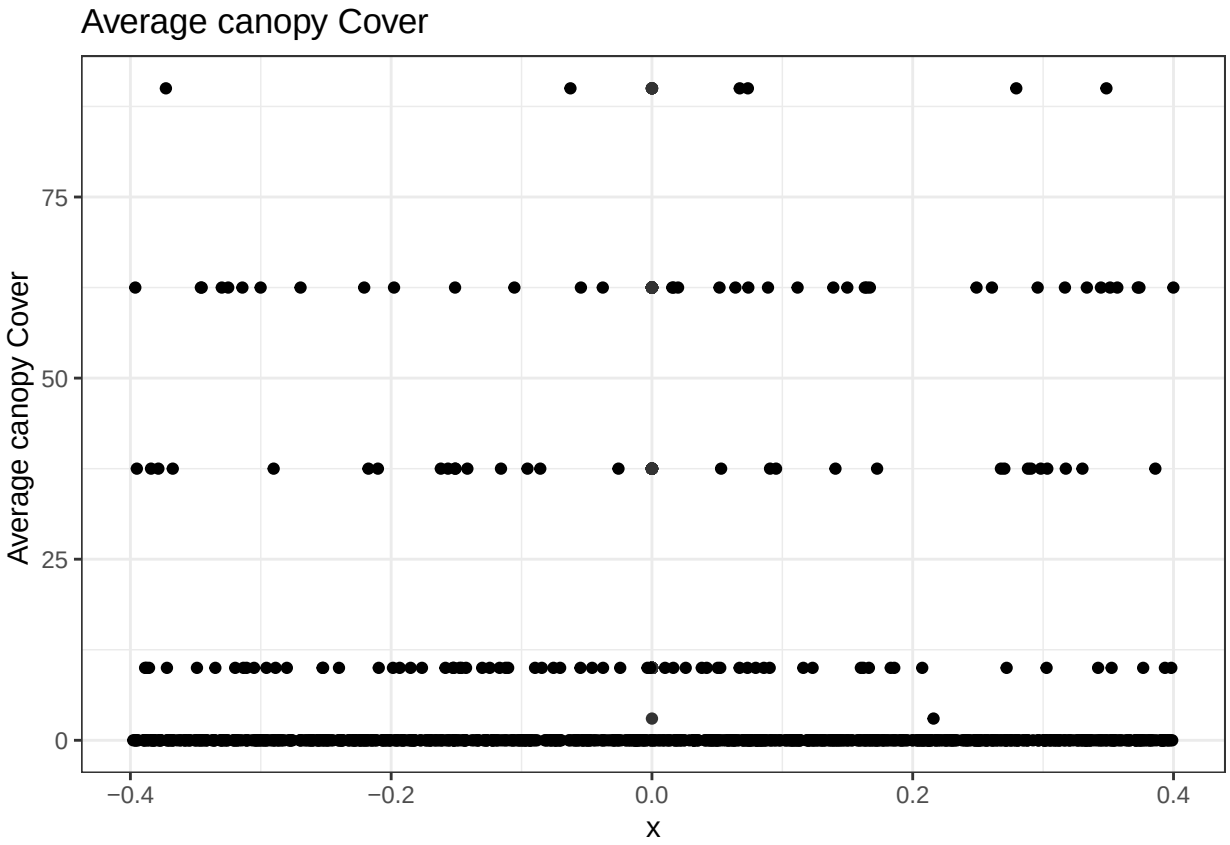
```
## Warning: Expected 1 pieces. Additional pieces discarded in 948 rows [1, 2, 3, 4, 5, 6,
## 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

```
#Fill in the missing plot names, year, and/or species and insert a "zero" for Cover

canopy_fill <- canopy %>%
  complete(plot,year,fill=list(Cover=0))

#Convert "plot" and "year" to factors with mutate. In R, factors are variables that take on a limited n
#of different values; often referred to as categorical variables (e.g. not continuous). Categorical var
#enter into statistical models differently than continuous variables, so storing data as factors ensure
#modeling functions will treat such data correctly.

canopy_factor1 <- canopy_fill %>%
```

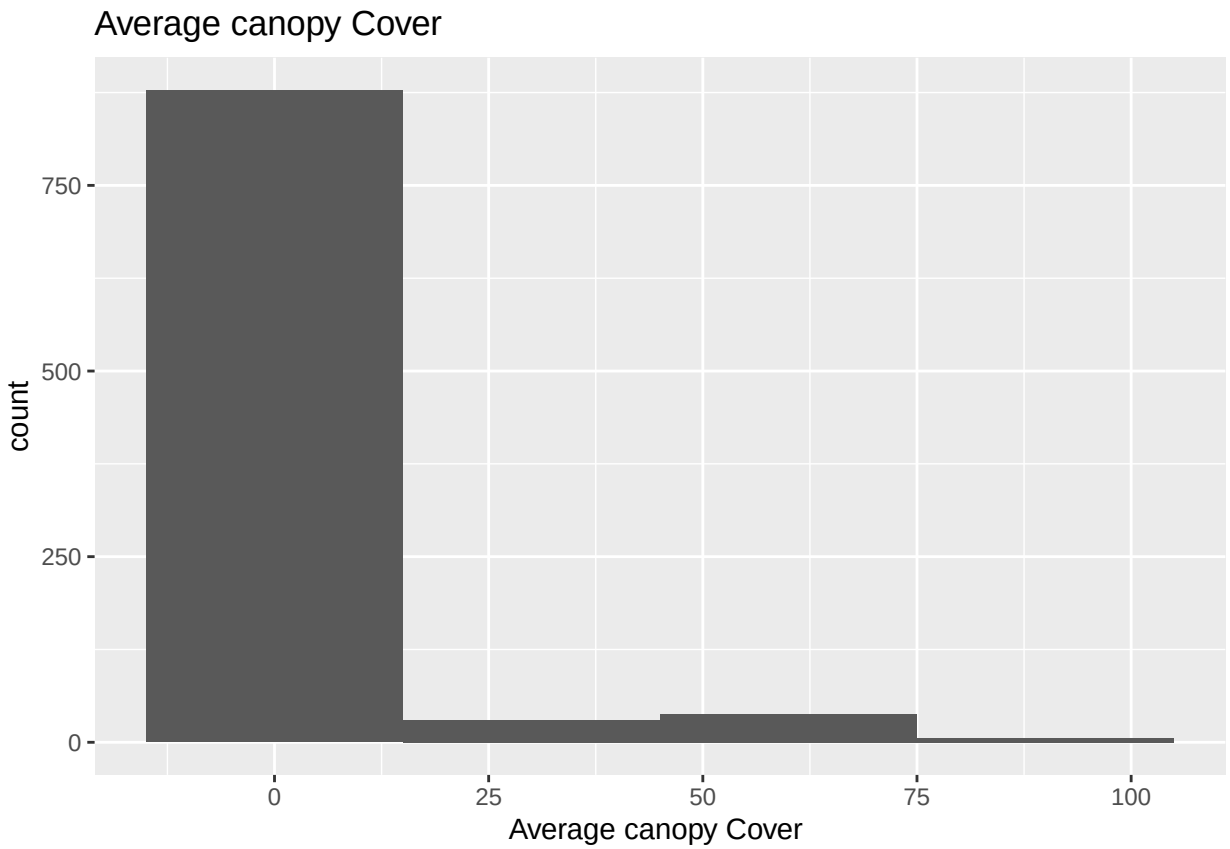
```
  mutate(plot=factor(plot))

canopy_factor2 <- canopy_factor1 %>%
  mutate(year=factor(year))


#Calculate summary statistics. Create a table with sample size, minimum value, maximum value, mean,
#standard deviation, and standard error. Use kable for an improved table.

canopy_factor2_summ <- canopy_factor2 %>%
  group_by(`Year`=year) %>%
  summarise(n=n(),
            `Min.`=min(Cover),
            `Max.`=max(Cover),
            `Mean`=mean(Cover),
            `SD`=sd(Cover),
            `SE`=sd(Cover)/sqrt(n))

kable(canopy_factor2_summ, booktabs=T, digits=3)
```

Year	n	Min.	Max.	Mean	SD	SE
1997	57	0	0.0	0.000	0.000	0.000
2001	199	0	0.0	0.000	0.000	0.000
2004	281	0	90.0	5.205	16.389	0.978
2008	200	0	90.0	6.263	16.904	1.195
2013	102	0	90.0	10.098	21.364	2.115
2023	113	0	62.5	8.743	17.287	1.626

```
#Create a boxplot. A boxplot displays the distribution of a dataset based on its five number summary of
#points: minimum, 1st quartile (25th percentile), median, third quartile (75th percentile), and maximum
#boxplot can be used to show the symmetry, skew, variance, and outliers of a dataset.

ggplot() +
  theme_bw() +
  geom_boxplot(aes(y=Cover),
               data=canopy_factor2) +
  geom_jitter(aes(y=Cover, x=0),
              height=0,
              data=canopy_factor2) +
  ylab("Average canopy Cover ") +
  ggtitle("Average canopy Cover ")
```




```
#Create a histogram. A histogram is an approximate representation of the distribution of continuous dat
#Histograms can be used to identify patterns in data, such as the shape of the distribution (e.g. norma
# distributed, skewed), the spread of the data, and outliers. The height of each bar represents the fre
#(Cover) of data points within the corresponding bin (x-axis).


ggplot() +
  geom_histogram(aes(x=Cover),
                 bins=4,
                 data=canopy_factor2) +
  xlab("Average canopy Cover ") +
  ggtitle("Average canopy Cover ")
```




```
#Fit a mixed model for a repeated measures design using lmer and indicate "plot" as a random effect. Ca
#summary on the model. We have repeated measurements on individual plots (experimental units) and those
#measurements will be correlated (not independent). A mixed model acCovers for the correlated responses
#We indicate "plot" as a random effect to acCover for the correlation between measurements that arise f
#the same plot. "Plot" has random variation and is not of primary interest in this analysis.

canopy_lmer <- lmer(Cover~year + (1|plot),
                    data=canopy_factor2)
summary(canopy_lmer)
```

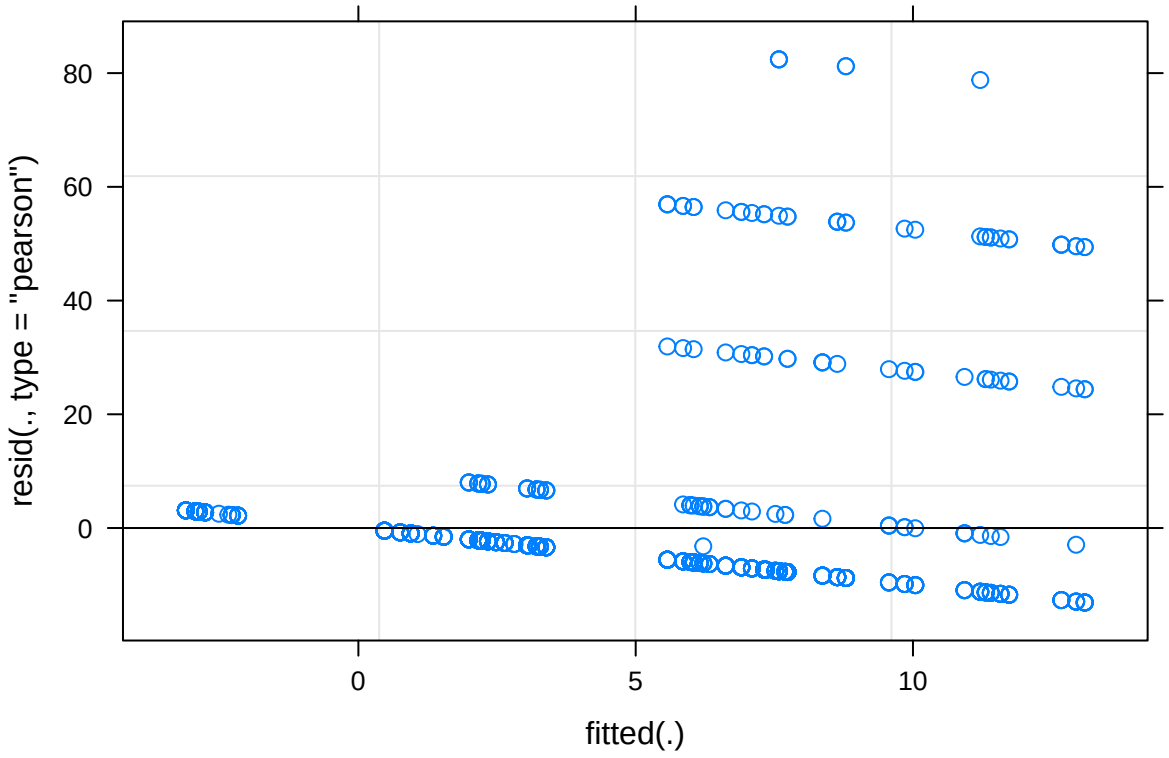
```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Cover ~ year + (1 | plot)
##    Data: canopy_factor2
##
## REML criterion at convergence: 7816.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -0.8892 -0.4967 -0.2066 -0.0319  5.5957
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  plot     (Intercept)   7.725   2.779
##  Residual             216.932  14.729
## Number of obs: 952, groups:  plot, 10
```

```
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)   0.3761     2.1743 161.5905   0.173 0.862898
## year2001     -0.5953     2.2471 946.0000  -0.265 0.791126
## year2004      4.5059     2.1741 945.9922   2.073 0.038488 *
## year2008      5.5572     2.2492 945.9404   2.471 0.013658 *
## year2013      9.8645     2.4645 945.4296   4.003 6.75e-05 ***
## year2023      8.5013     2.4250 945.8299   3.506 0.000477 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) yr2001 yr2004 yr2008 yr2013
## year2001 -0.809
## year2004 -0.836  0.809
## year2008 -0.810  0.784  0.810
## year2013 -0.737  0.712  0.736  0.713
## year2023 -0.750  0.726  0.750  0.726  0.661
```

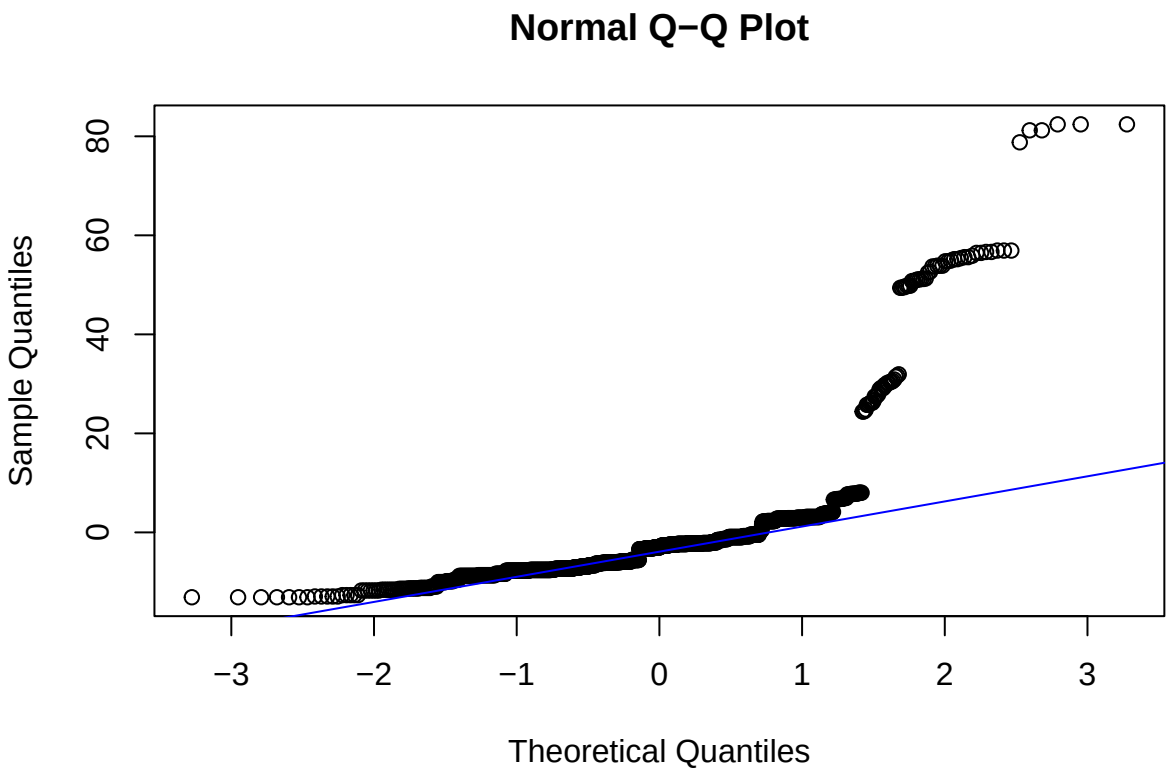
```
#Call anova (Analysis of Variance) on the model. Question: "Is there a difference in mean canopy Cover
#between years?"
#Null hypothesis (HO): There is no difference in mean canopy Cover between years.
#Alternative hypothesis (HA): There is a difference in mean canopy Cover between years.
#A small p-value (less than alpha 0.05) will reject the null hypothesis that there is no difference in
#canopy Cover between years. You can then conclude that there is evidence of a difference in mean canop
#Cover between years.

anova(canopy_lmer, ddf="Kenward-Roger")
```

```
## Type III Analysis of Variance Table with Kenward-Roger's method
##      Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## year  11174  2234.9     5 941.44  10.302 1.243e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Create a plot of standardized residuals vs. fitted values for the model to assess the assumption of co
#variance. Fitted values are the values predicted by the model. Residuals are the differences between t
#observed values (data) and the corresponding fitted values. This plot displays the fitted values of th
#along the x-axis and the residuals of the fitted values along the y-axis. If the spread of the residua
#equal at each level of the fitted values, the constant variance assumption is met. The residuals shoul
#scattered randomly about zero, with no obvious pattern emerging

plot(canopy_lmer)
```




```
#Create a Quantile-Quantile (QQ) plot to assess normality of the residuals (normal distribution). A QQ
#compares two probability distributions by plotting their quantiles against each other: the quantiles o
#sample data versus the theoretical quantile values from a normal distribution (or what we would expect
# a normal distribution). Data points should fall on a fairly straight line to indicate linearity. If d
#deviate largely from a straight line, it suggests that the two data sets do not have the same distribu

qqnorm(residuals(canopy_lmer))
qqline(residuals(canopy_lmer), col = "blue")
```




```
#Create an emmeans object and conduct Tukey-adjusted pairwise comparisons ("contrasts") between years.
#These contrasts provide estimates of the pairwise differences in average live canopy Cover between yea
#Include confidence intervals and p-values that have been adjusted for multiple comparisons.
#Multiple comparison problem: each time you run a hypothesis test, there is a small chance you will obt
#a "false" significant result (you will reject the null hypothesis when it is actually true, also calle
#Error Rate). If you run multiple tests, the number of "false positives" increases with each test, so t
#this Type I Error Rate, the p-values can be adjusted (Tukey is one method and used here) to be more
#conservative (less false positives).
#Call tidy to create an emmeans table. Copy tidy code and include kable for an improved table. Use the
#table column headers in the first tidy table to create the headers in the second (and final) tidy tabl

canopy_lmer_emm <- emmeans(canopy_lmer, ~year)
contrast(canopy_lmer_emm, "pairwise" , infer=TRUE, conf.int=TRUE)
```

```
##  contrast           estimate   SE  df lower.CL upper.CL t.ratio p.value
##  year1997 - year2001    0.595 2.25 946    -5.84    7.026   0.264  0.9998
##  year1997 - year2004   -4.506 2.18 946   -10.73    1.716  -2.068  0.3051
##  year1997 - year2008   -5.557 2.25 946   -11.99    0.880  -2.465  0.1356
##  year1997 - year2013   -9.865 2.47 945   -16.91   -2.816  -3.996  0.0010
##  year1997 - year2023   -8.501 2.43 946   -15.44   -1.564  -3.499  0.0065
##  year2001 - year2004   -5.101 1.37 939    -9.01   -1.194  -3.729  0.0028
##  year2001 - year2008   -6.152 1.48 939   -10.37   -1.931  -4.162  0.0005
##  year2001 - year2013  -10.460 1.80 940   -15.60   -5.323  -5.815  <.0001
##  year2001 - year2023   -9.097 1.74 939   -14.06   -4.132  -5.232  <.0001
##  year2004 - year2008   -1.051 1.36 938    -4.95    2.845  -0.770  0.9724
```

```
##  year2004 - year2013   -5.359 1.71 940   -10.24   -0.481  -3.137  0.0217
##  year2004 - year2023   -3.995 1.64 939    -8.69    0.699  -2.430  0.1469
##  year2008 - year2013   -4.307 1.80 939    -9.44    0.821  -2.399  0.1577
##  year2008 - year2023   -2.944 1.74 939    -7.90    2.015  -1.695  0.5352
##  year2013 - year2023    1.363 2.01 938    -4.39    7.112   0.677  0.9844
##
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 6 estimates
## P value adjustment: tukey method for comparing a family of 6 estimates
```

```
tidy(contrast(canopy_lmer_emm, "pairwise", infer=TRUE), conf.int=TRUE)
```

```
## # A tibble: 15 x 10
##    term  contrast         null.value estimate std.error    df conf.low conf.high
##    <chr> <chr>                 <dbl>    <dbl>     <dbl> <dbl>    <dbl>     <dbl>
##  1 year  year1997 - year~          0    0.595      2.25  946.    -5.84     7.03
##  2 year  year1997 - year~          0   -4.51       2.18  946.   -10.7      1.72
##  3 year  year1997 - year~          0   -5.56       2.25  946.   -12.0      0.880
##  4 year  year1997 - year~          0   -9.86       2.47  945.   -16.9     -2.82
##  5 year  year1997 - year~          0   -8.50       2.43  946.   -15.4     -1.56
##  6 year  year2001 - year~          0   -5.10       1.37  939.    -9.01    -1.19
##  7 year  year2001 - year~          0   -6.15       1.48  939.   -10.4     -1.93
##  8 year  year2001 - year~          0  -10.5        1.80  940.   -15.6     -5.32
##  9 year  year2001 - year~          0   -9.10       1.74  939.   -14.1     -4.13
## 10 year  year2004 - year~          0   -1.05       1.36  938.    -4.95     2.84
## 11 year  year2004 - year~          0   -5.36       1.71  940.   -10.2     -0.481
## 12 year  year2004 - year~          0   -4.00       1.64  939.    -8.69     0.699
## 13 year  year2008 - year~          0   -4.31       1.80  939.    -9.44     0.821
## 14 year  year2008 - year~          0   -2.94       1.74  939.    -7.90     2.01
## 15 year  year2013 - year~          0    1.36       2.01  938.    -4.39     7.11
## # i 2 more variables: statistic <dbl>, adj.p.value <dbl>
```

```
tidy(contrast(canopy_lmer_emm, "pairwise", infer=TRUE), conf.int=TRUE) %>%
  select(Contrast=contrast, Estimate=estimate,
         SE=std.error, df=df, `CI-low`=conf.low, `CI-high`=conf.high,
         `P-value`='adj.p.value') %>%
  kable(digits=c(2, 2, 2, 2, 2, 2, 6), booktabs=T)
```

Contrast	Estimate	SE	df	CI-low	CI-high	P-value
year1997 - year2001	0.60	2.25	946.00	-5.84	7.03	0.999825
year1997 - year2004	-4.51	2.18	945.99	-10.73	1.72	0.305059
year1997 - year2008	-5.56	2.25	945.94	-11.99	0.88	0.135629
year1997 - year2013	-9.86	2.47	945.44	-16.91	-2.82	0.000977
year1997 - year2023	-8.50	2.43	945.83	-15.44	-1.56	0.006469
year2001 - year2004	-5.10	1.37	939.35	-9.01	-1.19	0.002794
year2001 - year2008	-6.15	1.48	939.03	-10.37	-1.93	0.000492
year2001 - year2013	-10.46	1.80	939.57	-15.60	-5.32	0.000000
year2001 - year2023	-9.10	1.74	939.02	-14.06	-4.13	0.000003
year2004 - year2008	-1.05	1.36	938.30	-4.95	2.84	0.972408
year2004 - year2013	-5.36	1.71	939.96	-10.24	-0.48	0.021676

Contrast	Estimate	SE	df	CI-low	CI-high	P-value
year2004 - year2023	-4.00	1.64	939.01	-8.69	0.70	0.146907
year2008 - year2013	-4.31	1.80	938.98	-9.44	0.82	0.157664
year2008 - year2023	-2.94	1.74	938.85	-7.90	2.01	0.535158
year2013 - year2023	1.36	2.01	937.72	-4.39	7.11	0.984438

```
#Eva's timeline graph
canopy_factor2_summ %>%
ggplot(aes(x=Year, y=Mean, size=Mean))+geom_point()+theme_classic()+ylab("Mean Canopy Cover")
```