Climate change influence on laying time in birds

Hypothesis

We want to investigate if the time at which flycatchers lay eggs is correlated with the increasing average temperature. We selected three species: *Elaenia gigas*, *Elaenia cristata* and *Elaenia dayi*. We recorded the difference in lay deposition (difference in month percentage, 0 at 1st of April) and the average temperature for the period of interest. We want to know if the time of deposition is significantly different between 50 years ago and last year, which have different average temperature. *Temperature A* is 50 years ago and *Tempreature B* is last year.

Workflow

1. Load the .csv file *climate-change-data-1*:

```
setwd('/home/GIT/BEHAVIOURAL-BIOLOGY-2019/climate-change') #relative to you
df <- read.csv('climate-change-data-1.csv') #feel free to use your favorite import function
```

2. Explore the data Look at the dataset

```
head(df) #shows the first 6 lines of data
```

```
##       Species LayDay Temperature
## 1    E. gigas  -0.12           A
## 2 E. cristata  -0.18           A
## 3     E. dayi  -0.21           A
## 4    E. gigas  -0.53           A
## 5 E. cristata   0.09           A
## 6     E. dayi  -0.16           A
```

```
View(df) #view the dataset in a separate tab
```

Explore the data from 50 years ago, *Temperature A*.

```
only_A <- subset(df, Temperature == "A") #subset to only look at temperature A
avg_LayDayA <- mean(only_A$LayDay) #calculate the mean lay day for temperature A
avg_LayDayA
```

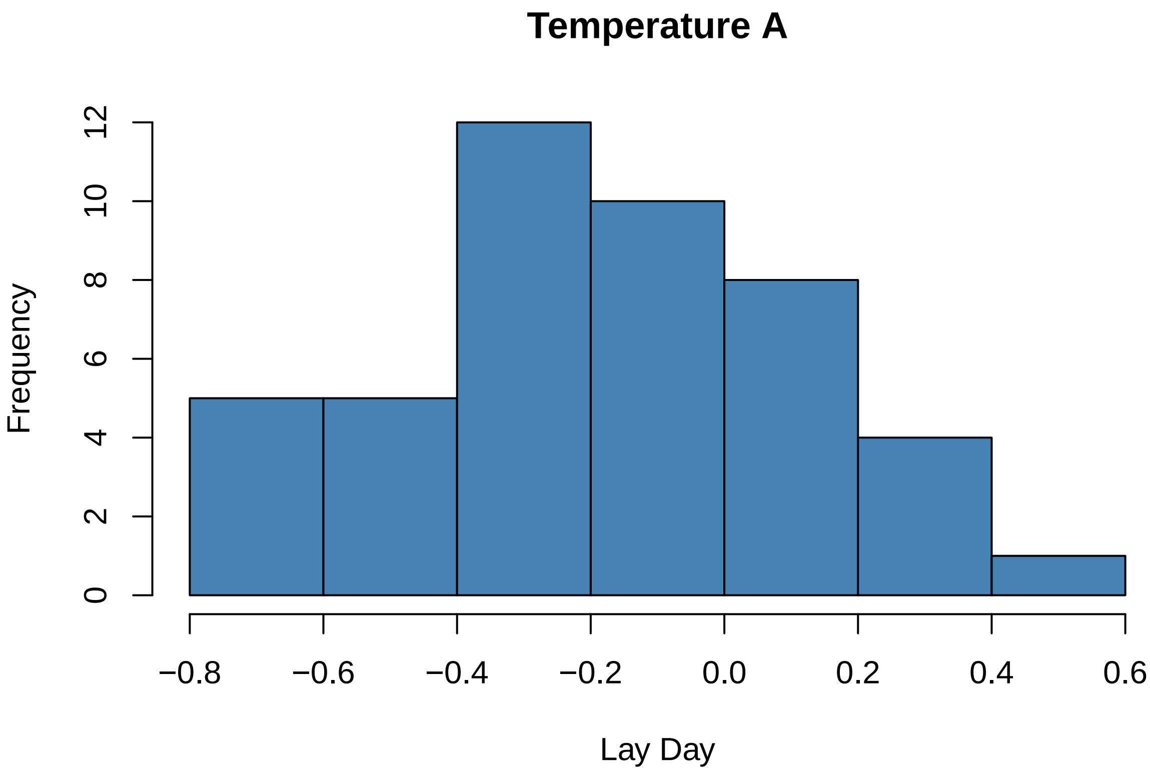
```
## [1] -0.1837778
```

```
med_LayDayA <- median(only_A$LayDay) #calculate the median lay day for temperature B
med_LayDayA
```

```
## [1] -0.19
```

We make a histogram to visually represent the distrubution of the data

```
hist(only_A$LayDay, col = 'steelblue', main = 'Temperature A', xlab = 'Lay Day')
```



We do the same for the data from the last year, *Temperature B*.

3. Test relationship between lay day and temperature

We want to see if the clutch lay day is significantly different between *Temperature A* and *Temperature B*. For this we are going to use the Student's t test.

The Student's t test compares two means and tells us if they are different from each other. The t test also shows how significant the differences are. In other words it lets us know if those differences could have happened by chance or not. When conducting a t-test you are making assumptions on the data. One of these assumptions is that residuals have a normal distribution.

First, we will do a linear model to study the relation between the lay day and temperature and later test the normality of the distribution of the residuals.

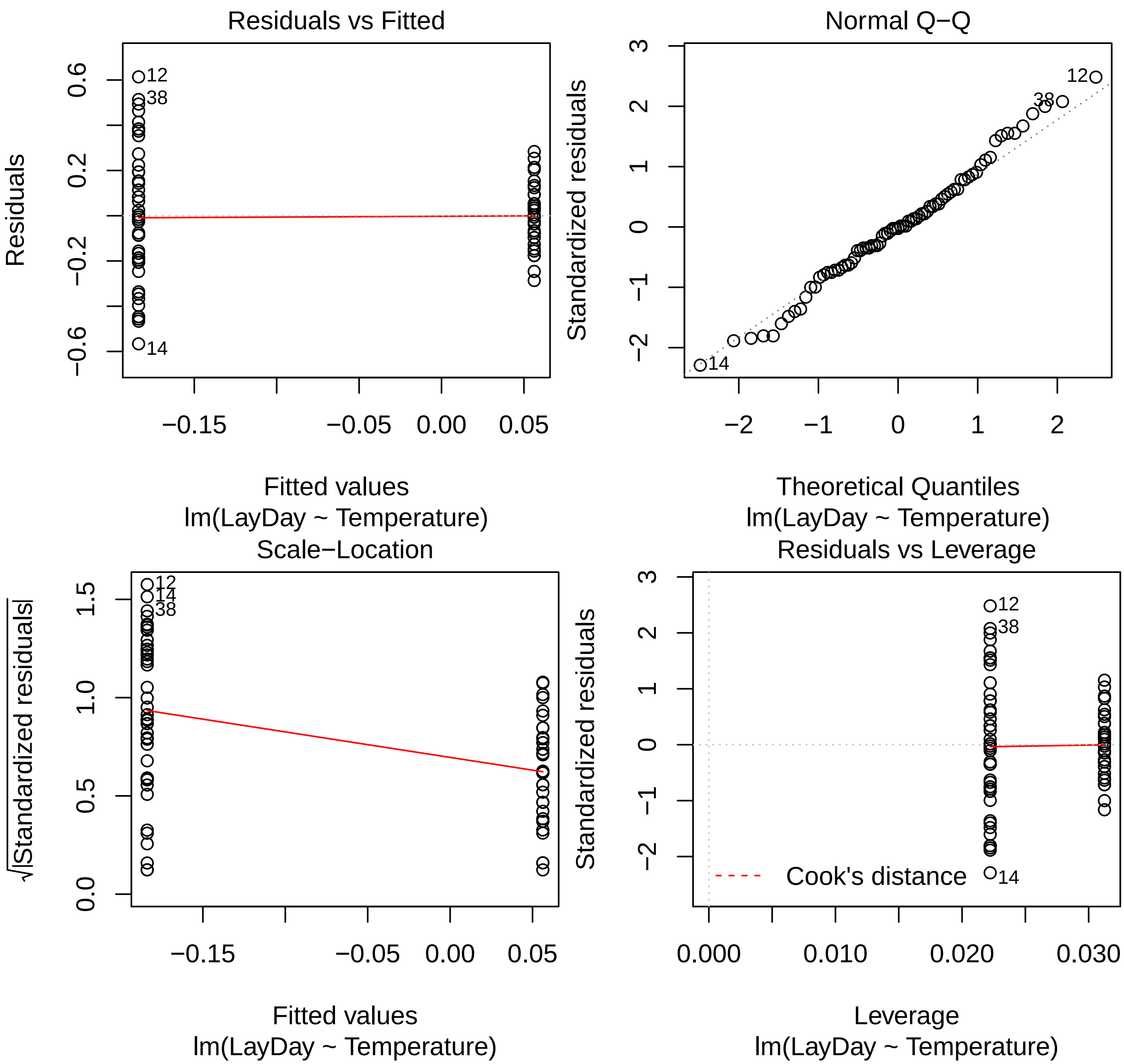
```
model1 <- lm(LayDay ~ Temperature, data = df)
summary(model1)
```

```
##
## Call:
## lm(formula = LayDay ~ Temperature, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.56622 -0.15625 -0.00622  0.14378  0.61378
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.18378    0.03727  -4.932 4.77e-06 ***
## TemperatureB  0.24003    0.05781   4.152 8.62e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.25 on 75 degrees of freedom
## Multiple R-squared:  0.1869, Adjusted R-squared:  0.1761
## F-statistic: 17.24 on 1 and 75 DF,  p-value: 8.619e-05
```

To test the assumption of normality we will do a qqplot of the residuals. The residual of an observed value is the difference between the observed value and the estimated value.

```r
plot(model1) #this command generates 4 plots, we are only going to work with Normal Q-Q
```



Now we perform the actual t test to learn if the lay day changes in relation to the temperature. A p-value of more than 0.05 means that the relation is significant.

```r
t.test(LayDay ~ Temperature, data = df)
```

```
##
##  Welch Two Sample t-test
##
## data:  LayDay by Temperature
## t = -4.6373, df = 66.161, p-value = 1.714e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.3433657 -0.1366899
## sample estimates:
## mean in group A mean in group B
##      -0.1837778       0.0562500
```

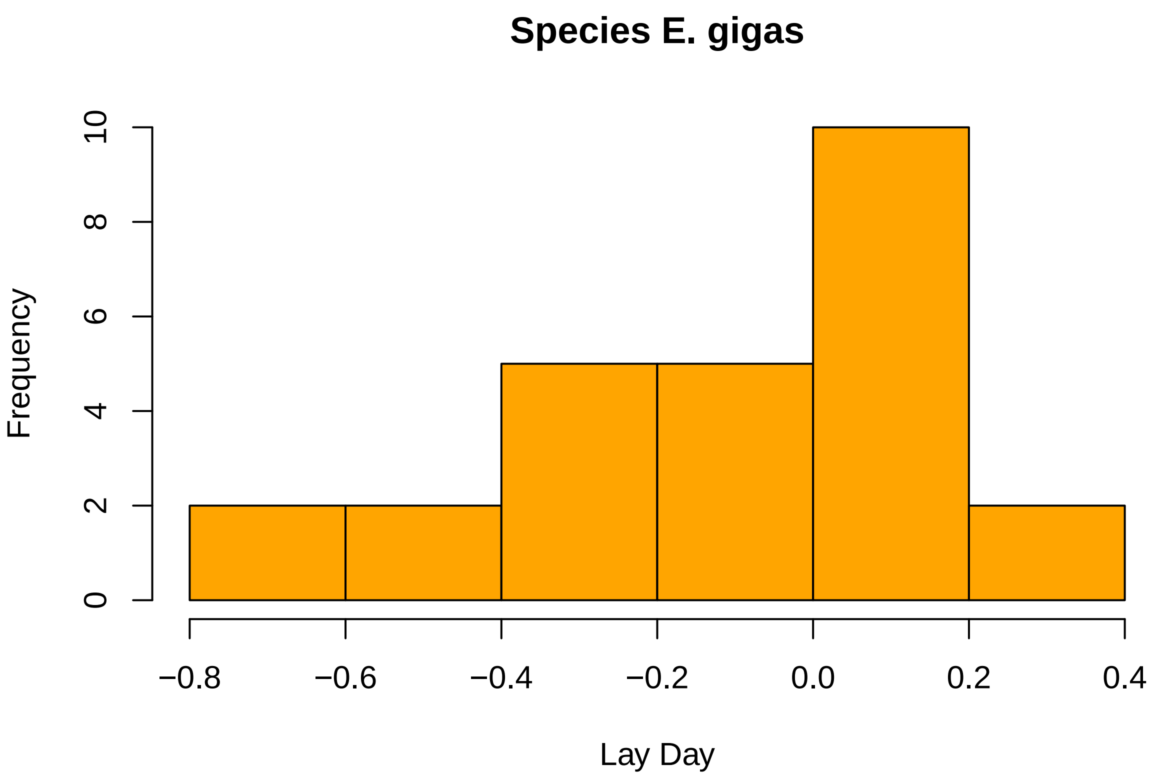
4. Test if the lay day changes significantly according to species

In the data set we have three different species, in order to investigate if these species have differences in clutch lay day we will preform an ANOVA. This analysis allows us to compare three or more groups.

First we will explore the data from Species *E. gigas*.

```r
only_gigas <- subset(df, Species == "E. gigas")
```

```r
hist(only_gigas$LayDay, col = 'orange', main = 'Species E. gigas', xlab = 'Lay Day')
```



We perfom the same exploration of the data for Species *E. cristata* and Species *E. dayi*.

We do a linear model again, but this time we want to study lay day as a function of the *Species*.

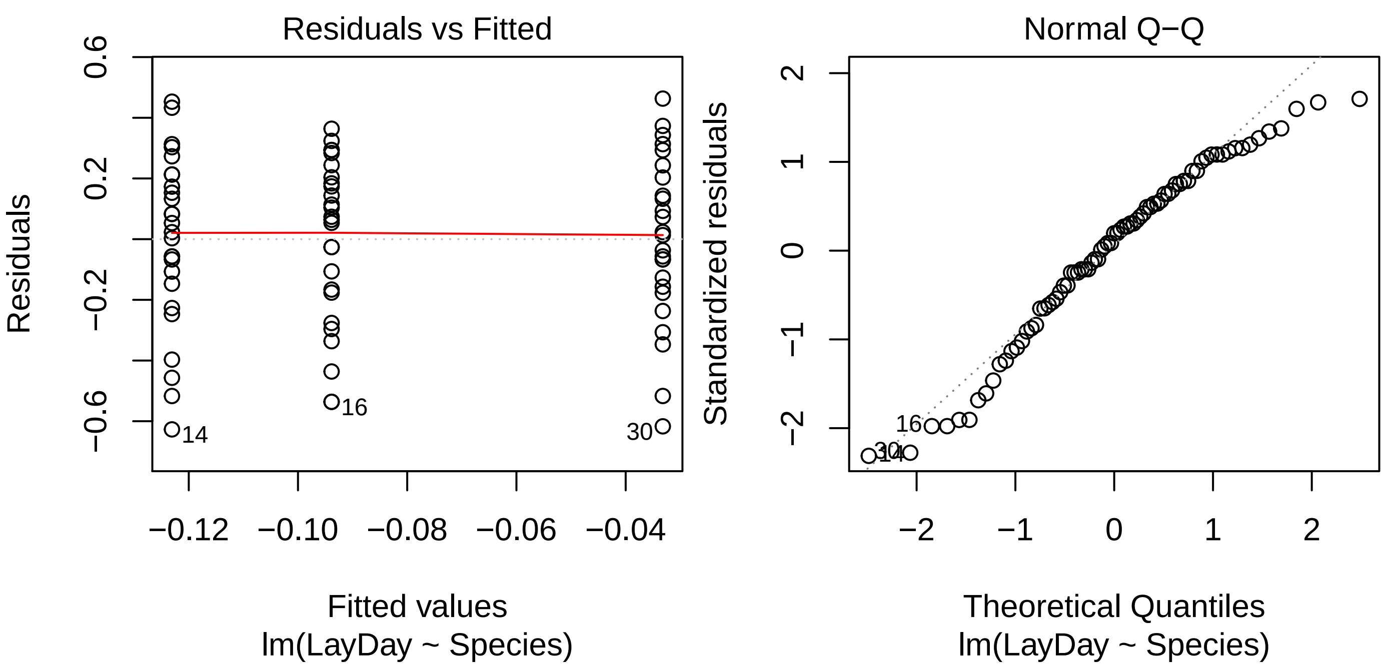
```r
model2 <- lm(LayDay ~ Species, data = df)
summary(model2)
```

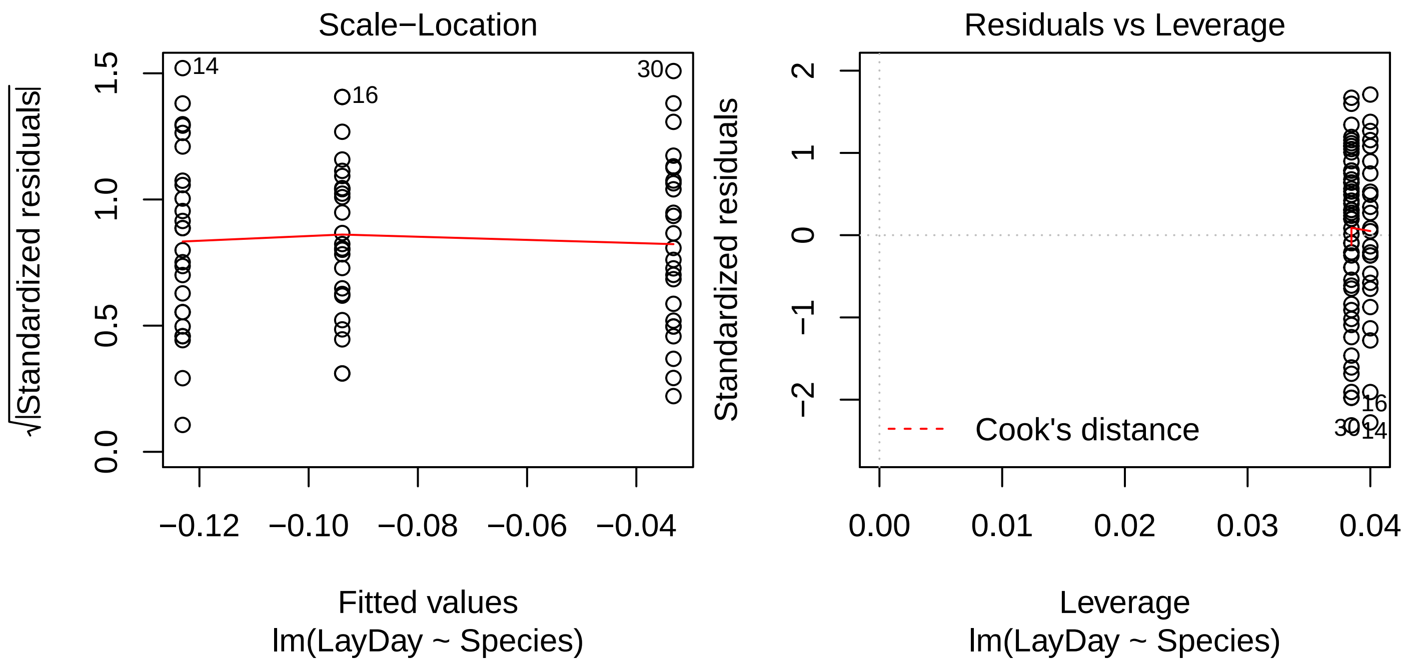
```
##
```

```
## Call:
## lm(formula = LayDay ~ Species, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.62692 -0.16615  0.05308  0.20385  0.46320
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -0.12308    0.05423  -2.270   0.0261 *
## SpeciesE. dayi  0.08988    0.07745   1.160   0.2496
## SpeciesE. gigas 0.02923    0.07669   0.381   0.7042
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2765 on 74 degrees of freedom
## Multiple R-squared:  0.01852,    Adjusted R-squared:  -0.00801
## F-statistic: 0.698 on 2 and 74 DF,  p-value: 0.5008
```

When you perform an ANOVA you are assuming that the distribution of the residuals is normal. To check this we will do a qqplot.

```
plot(model2) #this command generates 4 plots, we are only going to work with Normal Q-Q
```





Perform an ANOVA to study if the clutch's lay day is significantly different between the three species.

ANOVA Analysis of Variance (ANOVA) is a statistical method used to test general differences between two or more means. You are testing groups to see if there's a difference between them. It allows us to make comparisons between more than 2 groups.

```
anova(model2)
```

```
## Analysis of Variance Table
##
## Response: LayDay
##           Df Sum Sq  Mean Sq F value Pr(>F)
## Species    2 0.1067 0.053369   0.698 0.5008
## Residuals 74 5.6577 0.076456
```

5. Interpret the results:

- what can we say about the influence of *Species* on *LayDay*?
- what can we say about the influence of *Temperature* on *LayDay*?
- what limitations may this study have?
- give some suggestions on how to improve the study.

Exercise

We provided another dataset, *climate-change-data-2.csv*, that includes an extra species *Picumnus pumilus*. *P. pumilus* is phylogenetically distant to the flycatchers.

1. Explore the Species *P. pumilus* in the dataset, check for normality
2. Investigate if this species' lay day is influenced by temperature
3. Check if there is a significant difference between species regarding the lay day now that *P. pumilus* has been added.
4. Interpret the results:

- what can we say about the influence of *Temperature* on *LayDay* for *P. pumilus*?

- what has changed regarding influence of *Species* on *LayDay* with the addition of *P. pumilus*?
- what does this mean?

Now it is your turn to work by yourselves!