Univolutine butterflies READ ME

2023-05-03

This project is about male and female butterflies taken over the years of 1880 - 1973 from museums. This data was alongside the average temperature and rainfall in the month of June.

The first thing that I decided to do was create a script to tidy the data. This included a variety of code to make sure that the data had no missing values, checked for no NA values, cleaned column names, and individually looked at the different columns to check that the data overall didn't have missing values. I finished this off using the quick summary function to do an overall check on the data. The packages that I used for this were: library(tidyverse), library(here), and library(janitor).

Next, I created an exploratory data analysis script. I visualized the data by looking at the sex of the butterflies with all the numerical data: year, forewing length, mean temperature in June and rainfall in June. After creating my first graph (looking at the relationship between sex and year) I realised that there were 4 valuables in the sex column when I actually only wanted two, so this enabled me to then be able to fix this so it changed from: "Maes", "Males", "Female", "Females" to just "Male" and "Female". I then continued to create the other graphs to look for differences and uncertainty in the data. The other that I came across was the outlier data value in the forewing length column of 577.0, so I removed this row from my data in order to remove the anomaly. The packages that I used for this were: library (tidyverse), library(tidyr), library(readr), library(ggplot2), library(dplyr), library(knitr), library(here) and library(kableExtra).

After this I created a linear models script. I created lsmodel1 which is looking at two different data sets that directly relate to the butterflies: sex and forewing length. I used a variety of functions such as broom::tidy, performance::check_model, emmeans and summary to look more closely at this data and created graphs such as the Quantile-Qunatile plot and Cooks distance to check that outliers no longer existed after I had explored it. This once again, gave me a better insight into the data that i was working with. The packages that I used were: library(tidyverse), library(GGally), library(emmeans) and library(performance).

Using all three scripts and analysing the data I used these graphs to create a hypothesis for my Markdown file: Hypothesis: The temperature and rainfall has an effect on the forewing length of male and female butterflies. The null hypothesis goes against this and states that: the temperature and rainfall doesn't have an effect on the forewing length of male and female butterflies. My other two hypothesis I was also testing were: 1. The temperature in June affects the rainfall and therefore the number of male and female butterflies. 2. The rainfall in June across different years affects the number of male and female butterflies.

Firstly I created a box plot graph looking at the length of male and female butterfly wings during the rain in June. The point of creating this box plot graph is to show the mean of the forewing length for male and female butterflies to identify either their similarity or differences - it shows that the mean values are similar, however female butterfly's have slightly larger forewings with their mean value being around 55 mm and the mean of male butterflies being around 50 mm.

The next graph is a line graph and looks at the density of male and female butterflies in the month of June. It shows that there is a higher density of male butterfly's providing data of forewing length at around 12 - 14 degrees Celsius, however then from around 15.5 - 16 degrees Celsius when there are more males present. As the temperature decreases they become relatively similar.

The scatter plot shows how the temperature affects the rain in June for male and female butterflies, paying close attention to the p and the r values.

The next plot is a line graph produced shows the relationship between the temperature in June and the length of male and female butterfly forewings. It shows that females have a longer forewing length than

males across the years - and that as temperature increases so does forewing length.

I then created a bar graph to show the relationship between forewing length and the year of male and female butterflies.

I finally finished this off with three scatter plots, one to show the relationship between male and female butterflies and their year and rain in June and the next two to look into the main hypothesis, therefore looking at forewing length of male and female butterflies with temperature and rain in June.

I used R version 4.2.3