Hoops Longwing Data Analysis

Zane Billings

15 November 2019

In order to start analyzing the Hoops' Longwing sample data, we will first load the `tidyverse` package suite. After loading the packages we need, we can use `readr::read_csv()` to load in the data. But, notice the imported data frame has a useless column at the beginning, which we can easily remove manually.

```
library(tidyverse)

butterfly <- read_csv("hoops_longwing_study.csv")
butterfly <- butterfly[ , -1]
```

Now that we have the data imported, we can go ahead and take a quick look at the summary and structure.

```
summary(butterfly)
```

```
##   wing_length      wing_width          age        num_offspring
##  Min.   :11.06   Min.   : 4.640   Min.   : 8.00   Min.   :20.00
##  1st Qu.:14.05   1st Qu.: 6.202   1st Qu.:12.00   1st Qu.:24.00
##  Median :15.72   Median : 7.610   Median :21.50   Median :28.00
##  Mean   :19.70   Mean   : 8.397   Mean   :22.18   Mean   :27.78
##  3rd Qu.:27.29   3rd Qu.:10.070   3rd Qu.:28.00   3rd Qu.:31.00
##  Max.   :33.93   Max.   :14.990   Max.   :50.00   Max.   :36.00
##  feeding_range      color_peak      num_mates     avg_scale_size
##  Min.   : 1.160   Min.   :375.3   Min.   : 1.00   Min.   :23.09
##  1st Qu.: 2.645   1st Qu.:388.3   1st Qu.: 2.00   1st Qu.:27.65
##  Median : 3.475   Median :394.1   Median : 3.50   Median :31.58
##  Mean   : 5.545   Mean   :392.7   Mean   : 5.62   Mean   :38.21
##  3rd Qu.: 5.067   3rd Qu.:397.6   3rd Qu.: 9.00   3rd Qu.:50.37
##  Max.   :25.950   Max.   :409.4   Max.   :15.00   Max.   :66.13
##  antenna_length    num_spots      population        dispersal_distance
##  Min.   :0.040   Min.   : 3.00   Length:50          Min.   :23.15
##  1st Qu.:1.123   1st Qu.: 4.25   Class :character   1st Qu.:24.48
##  Median :1.605   Median : 6.00   Mode  :character   Median :24.72
##  Mean   :2.233   Mean   : 5.98                      Mean   :24.70
##  3rd Qu.:3.857   3rd Qu.: 8.00                      3rd Qu.:25.00
##  Max.   :4.880   Max.   :10.00                      Max.   :26.15
##   body_length      sample_id
##  Min.   : 2.440   Length:50
##  1st Qu.: 5.433   Class :character
##  Median : 6.365   Mode  :character
##  Mean   : 6.855
##  3rd Qu.: 8.110
##  Max.   :11.830
```

```
str(butterfly)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    50 obs. of  14 variables:
##  $ wing_length       : num  14.1 12.2 21.3 32.4 15.5 ...
##  $ wing_width        : num  8.38 6.19 9.78 12.75 9.02 ...
##  $ age               : num  13 36 13 24 11 19 10 8 12 30 ...
##  $ num_offspring     : num  25 33 23 31 24 28 23 21 25 30 ...
```

```
##  $ feeding_range     : num  2.24 9.79 2.53 4.39 2.46 4.23 1.96 2.74 2.32 4.13 ...
##  $ color_peak        : num  401 387 387 404 393 ...
##  $ num_mates         : num  4 2 8 13 4 14 9 1 3 2 ...
##  $ avg_scale_size    : num  29.6 25.5 39 66.1 26.3 ...
##  $ antenna_length    : num  1.2 0.59 2.9 4.55 1.6 4.55 4.28 0.04 0.43 1.52 ...
##  $ num_spots         : num  5 8 5 3 5 3 3 8 7 6 ...
##  $ population        : chr  "Ternate" "Ternate" "Kayoa" "Tidore" ...
##  $ dispersal_distance: num  25.8 24.4 24 26.1 25.1 ...
##  $ body_length       : num  5.76 3.57 5.62 8.05 7.23 ...
##  $ sample_id         : chr  "Ter_01_ZB" "Ter_02_MP" "Kay_03_MP" "Tid_04_ZB" ...
```

The only real change we need to make is to convert the `population` variable into a factor, since the functions provided in `readr` do not coerce strings to factors by default.

```
butterfly$population <- as.factor(butterfly$population)
summary(butterfly$population)
```

```
##   Kayoa Ternate  Tidore
##       4      31      15
```

Now that all of our variables are imported correctly, let's go ahead and document what each variable means.

Variable	Description
Wing length	
Wing width	
Age	
Number of offspring	
Feeding range	
Color peak	
Number of mates	
Avg. scale size	
Antenna length	
Number of spots	
Population	
Dispersal distance	
Body length	
Sample ID	