High sparrow body length decreases survival

A demonstration of Rmarkdown using Herman Bumpus' data

Brad Duthie

Biological and Environmental Sciences, University of Stirling, Stirling, UK, FK9 4LA

Abstract

Writing documents in Rmarkdown using Rstudio can make scientific workflow more efficient, and here I demonstrate how a scientific manuscript can be written using a classical data set first published by Herman Bumpus. I integrate Bumpus' data with Rmarkdown to produce a sample manuscript, testing whether or not sparrow body length decreases survival following a storm in southern New England. Using a t-test, I show that surviving birds have lower body length than birds that do not survive. All analyses of data are incorporated into the underlying Rmarkdown document, including figures and a table. References are incorporated using BibTeX. The underlying code for this manuscript is publicly available on GitHub as part of the Stirling Coding Club organisation.

Introduction

In the late 1800s, there was a particularly severe snowstorm in Providence, Rhode Island. At the time, Herman Bumpus was a professor of comparative zoology at Brown University. Bumpus noticed that the storm had a particularly negative effect on the local sparrow population (*Passer domesticus*) and decided to use the event to test Charle's Darwin's theory of natural selection (Darwin 1859). Bumpus collected 136 sparrows; some of these sparrows survived the storm, while others perished. Bumpus (1898) published a paper and all of the data that he had collected. These data are now a classic data set in biology, and have been analysed multiple times (e.g., Johnston et al. 1972). Here I will use Bumpus' data to demonstrate how to write a scientific manuscript in Rmarkdown.

The focus of this manuscript is therefore not on Bumpus' data or survival of sparrows *per se*, but the process of scientific writing using Rmarkdown. I have chosen the Bumpus data set because it provides a useful tool for working through most key features of Rmarkdown that scientists might want to use when writing a manuscript. The example question that I will address through this data set and R analysis in Rmarkdown is whether or not increasing sparrow body length is associated with decreased survival following a storm.

Methods

Bumpus focused his study on the House Sparrow (*Passer domesticus*; see Figure 1), which has a very wide global distribution. It is native to Europe and Asia, but not the Americas where Bumpus collected his original study (Bumpus 1898). In addition to measuring total length and survival for 136 sparrows, Bumpus measured sparrow sex, wingspan, and mass, and also the length of each sparrow's head, humerus, tibiotarsus, skull, and sternum. While modern ornithologists believe that the total body length measurement that I will use today is subject to high observational error (Johnston et al. 1972), it will be more than sufficient for demonstrating Rmarkdown.

Figure 1: *Passer domesticus*

I performed an independent two-sample student's t-test on sparrow total body length to test whether or not sparrows that died in the 1898 storm were larger than sparrows that survived. I assume that both groups of sparrows (dead and living) have equal variances, so the test statistic t is calculated as follows,

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

In the above, $\bar{X}_1$ and $\bar{X}_2$ are the mean of the samples of sparrows that died and lived, respectively. Similarly,

40 n_1 and n_2 are the sample sizes of sparrows that died and lived, and s_p is the pooled sample mean, which is

41 calculated as follows,

$$s_p = \sqrt{\frac{s_{X_1}^2 + s_{X_2}^2}{2}}.$$

42 In the above, the $s_{X_1}^2$ and $s_{X_2}^2$ are the sample standard deviations for sparrows that died and lived, respectively.

43 I conducted the two sample t-test using the `t.test` function in R (R Core Team 2018).

Results

45 Bumpus' data included 72 sparrows that lived and 64 sparrows that died. The mean total length of living

46 sparrows was 158.71 mm, and the mean total length of dead sparrows was 160.48 mm. The two sample t-test

47 revealed a t-statistic of -2.99, which corresponds to a p-value of $P = 0.00167$.

48 Figure 2 shows the difference between total length in sparrows that survived versus sparrows that died.

49 Overall, dead sparrows were 1.78 mm longer than living sparrows, and ranged between 152 and 163 mm.

50 Living sparrows ranged between 153 and 160.25 mm (Figure 2).

Discussion

52 I have analysed data collected by Herman Bumpus (Bumpus 1898) on the relationship between sparrow

53 (*Passer domesticus*) total length and surival following an unusually severe storm. I found that sparrows that

54 died in the storm were longer than sparrows that survived, which suggests that higher sparrow body length

55 decreased survival. Of course, it is not possible to definitively conclude a causal relationship between any

56 aspect of body size and sparrow survival , and even the available data collected by Bumpus would permit a

57 more thoughtful analysis than that conducted in this study (see Appendix Table 1).

58 Overall, this document demonstrates how high quality, professional looking documents can be written using

59 Rmarkdown. The underlying code for this manuscript is publicly available, along with accompanying notes

60 to understand how it was written. By using Rmarkdown to write manuscripts, authors can more easily use
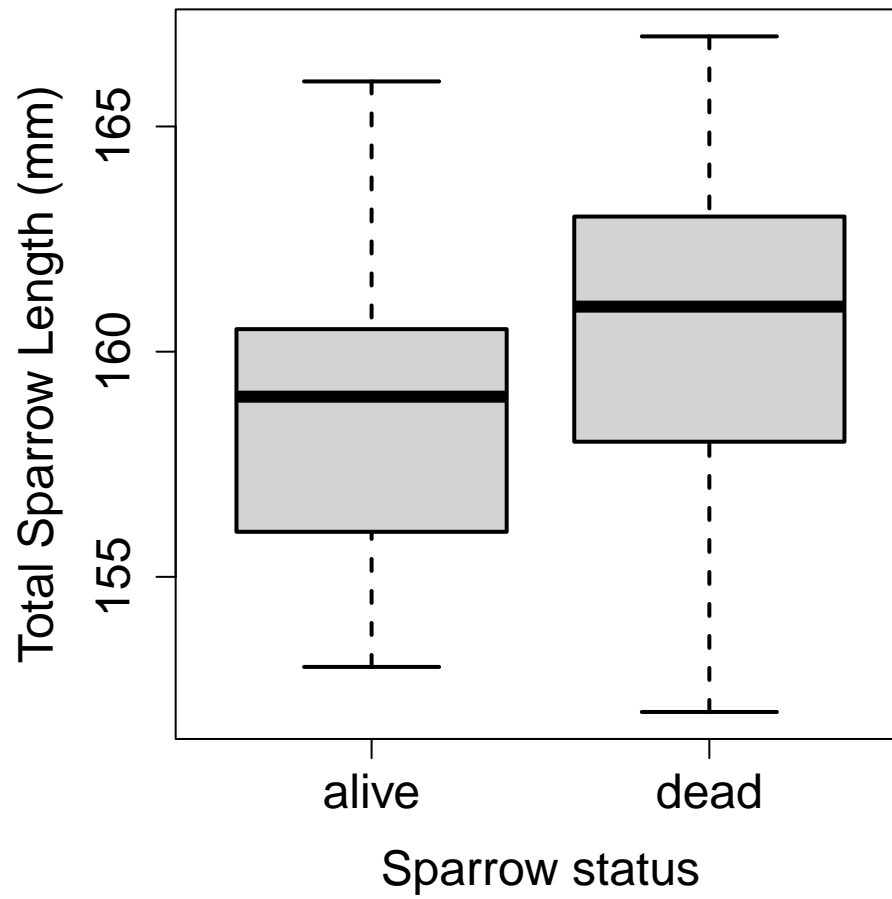
Figure 2: Box plot of the total lengths of live and dead sparrows following a snowstorm in Providence, RI, as originally collected by Hermon Bumpus. The central horizontal line shows median values. Boxes and whiskers show inter-quartile ranges and extreme values, respectively.

version control (e.g., git) throughout the writing process. The ability to easily integrate citations though BibTeX, LaTeX tools, and dynamic R code can also make writing much more efficient and more enjoyable. Further, obtaining the benefits of using Rmarkdown does not need to come with the cost of isolating colleagues who prefer to work with Word or LaTeX because Rmarkdown can easily be converted to these formats (in the case of Word, with the push of a button). By learning all of the tools used in this manuscript, readers should have all of the necessary knowledge to get started writing and collaborating in Rmarkdown.

References

Bumpus, H. C. 1898. Eleventh lecture. The elimination of the unfit as illustrated by the introduced sparrow, *Passer domesticus*. (A fourth contribution to the study of variation.). Biological Lectures: Woods Hole Marine Biological Laboratory 209–225.

Darwin, C. 1859. The Origin of Species. Penguin, New York.

Johnston, R. F., D. M. Niles, and S. A. Rohwer. 1972. Hermon Bumpus and natural selection in the House Sparrow *Passer domesticus* . Evolution 26:20–31.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Appendix Table 1

An example table is shown below, which includes all of the variables collected by Bumpus (1898) for the first 10 measured sparrows. The full data set can be found online in GitHub.

Table 1: First ten rows of the original data set collected by Hermon Bumpus

sex	surv	totlen	wingext	wgt	head	humer	femur	tibio	skull	stern
male	alive	154	241	24.5	31.2	0.687	0.668	1.022	0.587	0.830
male	alive	160	252	26.9	30.8	0.736	0.709	1.180	0.602	0.841

sex	surv	totlen	wingext	wgt	head	humer	femur	tibio	skull	stern
male	alive	155	243	26.9	30.6	0.733	0.704	1.151	0.602	0.846
male	alive	154	245	24.3	31.7	0.741	0.688	1.146	0.584	0.839
male	alive	156	247	24.1	31.5	0.715	0.706	1.129	0.575	0.821
male	alive	161	253	26.5	31.8	0.780	0.743	1.144	0.607	0.893
male	alive	157	251	24.6	31.1	0.741	0.736	1.153	0.610	0.862
male	alive	159	247	24.2	31.4	0.728	0.718	1.126	0.609	0.793
male	alive	158	247	23.6	29.8	0.703	0.673	1.079	0.602	0.820
male	alive	158	252	26.2	32.0	0.749	0.739	1.153	0.614	0.857