# 

## Hitchhikers' guide to analysing bird ringing data

Part 3: missing values

## Andrea Harnos<sup>1\*</sup>, Tibor Csörgő<sup>2</sup> & Péter Fehérvári<sup>1,3</sup>

Received: November 29, 2017 – Accepted: December 18, 2017



Harnos, A., Csörgő, T. & Fehérvári, P. 2017. Hitchhikers' guide to analysing bird ringing data – Part 3. – Ornis Hungarica 25(2): 141–156. DOI: 10.1515/orhu-2017-0020

**Abstract** This paper is the third part of our bird ringing data analysis series (Harnos *et al.* 2015a, 2016a) in which we continue to focus on exploring data using the R statistical

environment. Here, we give a short description of missing data patterns, then handling and visualization techniques that help simultaneously explore the data and the structure of missing values. We try to emphasize that missing data can seriously distort the results of the analyses, therefore the conclusions drawn.

Keywords: missing data pattern, exploration of incomplete data

Összefoglalás Sorozatunk harmadik részében folytatjuk a madárgyűrűzési adatok kezelését, elemzését az R statisztikai program használatával (Harnos *et al.* 2015a, 2016a). Röviden bemutatunk hiányzó adat mintázatokat, kezelésüket és azt, hogy hogyan lehet a meglévő adatokat és a hiányzó adatok struktúráját szimultán feltárni különböző ábrázolási módszerekkel. Egyúttal szeretnénk felhívni a figyelmet arra, hogy a hiányzó adatok komoly mértékben torzíthatják az elemzések eredményeit és a belőlük levont következtetéseket.

Kulcsszavak: hiányzó adatok mintázata, hiányzó adatok feltárása

<sup>1</sup>Department of Biomathematics and Informatics, University of Veterinary Medicine, 1078 Budapest, István utca 2., Hungary, e-mail: harnos.andrea@univet.hu

<sup>2</sup>Department of Anatomy, Cell- and Developmental Biology, Eötvös Loránd University, 1117 Budapest, Pázmány Péter sétány 1/C, Hungary

<sup>3</sup>Department of Zoology, Hungarian Natural History Museum, 1088 Budapest, Baross utca 13., Hungary <sup>\*</sup>corresponding author

## Introduction

This paper is the third part of our bird ringing data analyses series (Harnos *et al.* 2015a, 2016a) in which we focus on exploring missing data structures. We will give a short description of problems caused by missing data, and how they can bias basic descriptive statistics and explain how to generate and read useful figures to explore missing data patterns.

Our template dataset is of Pied Flycatchers trapped and ringed at the Ócsa Bird Ringing Station (Central Hungary) between 1984–2014 (for details see Harnos *et al.* 2015b, Csörgő *et al.* 2016, Harnos *et al.* 2016b). The dataset is available through our ringR (ringr. gryllosoft.hu) R package or from the online appendix of this paper along with the R codes (OrnisHungarica\_vol25(2)\_p141-156\_appendix.zip). We used the R 3.4.1 program (R

Core Team 2017), the VIM (Kowarik & Templ 2016), the RcmdrMisc (Fox 2017), the knitr (Xie 2014, 2015, 2016) and the xtable (Dahl 2016) R packages for the analyses and dynamic latex code generation, Tex Live, RStudio 1.0.136, Texmaker 4.4.1. for editing the R and LATEX scripts on the Ubuntu 14.04 platform.

Missing values are omnipresent in observational studies, including avian ringing datasets. In general, missing data are observations (e.g. morphometric measurements) that were intended to be made but for some reason were not. Missing values in a variable may arise independently of all the measured variables occurring in a purely haphazard fashion (MCAR: missing completely at random), or may be dependent on the observed or unobserved values of them. (For more detailed explanation see e.g. Carpenter & Kenward 2012, Little 2016.) For instance, consider a ringer who decides to neglect body mass measurements for individuals with brood patches, in order to reduce handling time for potentially incubating individuals. In this case missing values will depend on a binary variable 'brood patch presence'. The same "soft-hearted" ringer may also decide to drastically reduce handling time for individuals in apparently bad physical condition by skipping all morphometric measurements, including body mass. In this case, missing values in body mass will directly depend on the value of body mass. Using appropriate visualization techniques, the pattern of missingness and the relations of missing and non-missing data can be explored (Templ *et al.* 2012, Kowarik & Templ 2016).

#### Effect of missing values on basic descriptive statistics

To demonstrate the potential distorting effects of missing values we created several versions of the same variable (WING) excluding data from each with different models (WING.table.csv, the table and the codes generating the descriptive statistics are available in the online appendix). Initially, we randomly selected 200 individuals from the complete Pied Flycatcher dataset (All) with existing wing measurements and determined sex. We then assigned NAs to random 10% (Rand10), to random 30% (Rand30), to random 50% (Rand50) of the values, to random 50% of females (RandF50) and to random 30% males and 10% to females (RandF10M30) and to values below the 10% (Lower10) and above the 90% (Upper10) quantiles in the dataset. (Note that male wing length is significantly longer compared to females, see Harnos *et al.* 2015b.) We then calculated several descriptive statistics (mean, mean by sex (MeanF, MeanM), standard deviation (SD), median, minimum, maximum, number of non-missing values (*Table1*).

It is clear that randomly generated missing values hardly distort our sample estimates, except for Rand50, where the shear random reduction of sample size is causing the approximately 0.3 mm mean difference. If missingness depends on the variable itself (Lower10 and Upper10), the effect is apparent and substantial on all descriptives except the median and the opposite extreme value despite the relatively low number of introduced missing values. Sex dependent missingness (RandF50 and RandF10M50) also had a

Variable	Mean	MeanF	MeanM	SD	Median	Min.	Max.	N	NA
All	80.03	79.41	80.45	1.75	76	80	85	200	0
Rand10	80.03	79.43	80.45	1.73	76	80	85	180	20
Rand30	80.12	79.33	80.57	1.85	76	80	85	140	60
Rand50	80.30	79.76	80.68	1.65	77	80	85	100	100
RandF50	80.26	79.65	80.45	1.78	76	80	85	156	44
RandF10M50	79.78	79.34	80.28	1.67	76	80	83	137	63
Lower10	80.65	80.21	80.89	1.36	79	80	85	160	40
Upper10	79.40	79.09	79.64	1.30	76	80	81	159	41

Table 1.Numerical summaries of the WING.table (see text for explanations)1. táblázatA WING.table táblázat leíró statisztikái

sizeable impact on mean wing estimates, increasing it by 0.23 mm and decreasing it by 0.25 mm, respectively. In this case the mean estimates by sex groups are unbiased, the relatively large differences are caused by the high number of missing values. The mean estimate of the females, if only 10% of the values are missing (RandF10M50, MeanF = 79.34), is close to the mean of the females without missingness (All, MeanF = 79.41).

This little example (*Table1*) highlights the importance to understand the pattern of missing values to allow a more informed inference of results. However, this is not an easy task, and first of all needs a thorough knowledge on the precise details of data collection methodology. The nature of avian ringing datasets (i.e. spatial and temporal range, multiple participants) makes the evaluation of missing data patterns even more difficult, therefore we urge our readers to allocate extra time and energy in identifying the patterns and the underlying mechanisms.

#### Managing missing data

Three major approaches exist to deal with missing data: 1. to discard incomplete cases listwise (i.e. delete full rows from the data table; complete-case analysis) or pairwise (for each pair of variables for which data is available, the specific analysis will take that data into account) and analyse the remainder; 2. to impute (i.e. substitute using known data) the missing values and then analyse the filled-in data and 3. to analyse the incomplete data by a method that does not require a complete data table (Enders 2010, Little 2016).

In certain cases, deletion of incomplete cases listwise or pairwise can be a viable solution (i.e. producing unbiased estimates), but both requires MCAR data. Imputing requires deep statistical knowledge and may cause more harm than good in the wrong hands (Schafer & Graham 2002), thus we encourage our readers to consult a statistician. A good starting point may be the book of Enders (2010) which offers a good overview of imputing techniques, their usability and drawbacks. However, please note that once a data point is missing, the loss in information can not be substituted with statistical methods.

#### Importing data with missing values

In theory, missing values can be denoted arbitrarily, most commonly with character strings such as NA, na, null or with special characters like '\_', '-'. They may also be represented with values outside the range of the given variable like -99999 or with 0 in case of morphometrics. This potential variability of missing value representations is a major source of error and can have detrimental effects on our sample estimates (think of calculating mean wing length where missing values are marked with a 0). Furthermore, it is not uncommon to use multiple representations within the same dataset, depending on the variable. Prior to all analyses, one of our primary tasks is to understand and note how our missing values are denoted in each variable. We suggest to fix the missing data representation before data collection, if it is possible.

By default, only blanks and NAs are read as missing values in R. In case of other representations, the na.strings argument of the read.table() (and sibling functions e.g. read.csv()) should be specified appropriately. In our case, the lack of recorded sex of certain individuals was indicated with "N". If this sole letter is not used in other variables, we can import the data table giving "N" (besides "null") as a missing value string. First import the FICHYP.csv raw dataset, then summarise the SEX variable, then repeat it using the extended na.strings argument:

Note that the number of NAs in the second case equals the sum of NAs and Ns in the previous case.

#### **Recoding values as missing**

It is also possible to recode values to NAs after importing (see Harnos *et al.* 2015a) using the ifelse() function.

144

Another possibility is to directly substitute given values to NA. In the example below, we recode the values "FEJ" (developed bird with undetermined age) of the AGE variable to NA and then we check the result.

```
> mydata$AGE[mydata$AGE == "FEJ"] = NA
> summary(mydata$AGE)
1 1+ 1Y 2 2+ 2Y FEJ NA's
7 40 1614 31 3 251 0 20
```

Although we successfully recoded "FEJ" to NA, the AGE factor still retains the level "FEJ". The droplevels() and the factor() functions drop unused factor levels. The main difference between the two is while droplevels() can be applied to the whole dataset, the factor() function can be applied to only a single variable.

```
> mydata$AGE = droplevels(mydata$AGE)
> summary(mydata$AGE)
1 1+ 1Y 2 2+ 2Y NA's
7 40 1614 31 3 251 20
```

#### Locating and counting missing values

Once we recoded missing values to NA we have a decent selection of tools to locate them in the dataset. We illustrate how these work on a small sample vector called  $\times$  containing 10 Pied Flycatcher wing length measurements with 3 missing values.

```
> x = c(81, NA, 79, 82, 83, NA, 77, 81, NA, 80)
> anyNA(x)
[1] TRUE
> is.na(x)
[1] FALSE TRUE FALSE FALSE FALSE TRUE FALSE FALSE TRUE FALSE
> sum(is.na(x))
[1] 3
```

The anyNA() function returns TRUE if there are any missing value in the object, otherwise it returns FALSE. One of the most versatile functions is the is.na() function, which in our case returns a logical vector with the position of TRUE values indicating NAs. The sum of a logical vector returns the number of TRUE values, allowing for a quick count of missing values.

Combining is.na() with the which() function returns the index of missing values, or by using the NOT (negation) logical operator "!", we can get the index of non-missing values.

```
> which(is.na(x))
[1] 2 6 9
> which(!is.na(x))
[1] 1 3 4 5 7 8 10
```

The summary () function applied on a data frame also calculates the number of NAs in the last rows of the individual variables. First, we import the FICHYP1.csv dataset, than we make a subset data table called mydata2 containing only 5 variables. Then, using the summary () function, we can get a quick overview of missing data frequencies.

```
> mydata = read.table("FICHYP1.csv", sep=";",header=T)
> mydata2 = mydata[, c("SEX", "WING", "THIRD", "MASS", "AGE")]
> summary(mydata2)
               WING
  SEX
                            THIRD
                                           MASS
                                                       AGE
   :622 Min. :70.0 Min. :53.00 Min. : 9.20 1+ : 40
F
M :943 1st Qu.:79.0 1st Qu.:60.00 1st Qu.:11.90 1Y :1621
 NA's:401 Median :80.0 Median :61.00 Median :12.50
                                                    2+ : 3
          Mean :79.9 Mean :61.04 Mean :12.62 2Y : 282
          3rd Qu.:81.0 3rd Qu.:62.00 3rd Qu.:13.20 NA's: 20
          Max. :89.0 Max. :69.00 Max. :19.60
NA's :290 NA's :319 NA's :80
```

A more complex approach yielding similar results is to combine the sapply() function with the wrapped sum(is.na()) trick seen above.

```
> sapply(mydata2, function(x)(sum(is.na(x))))
SEX WING THIRD MASS AGE
401 290 319 80 20
```

The function complete.cases() returns a logical vector indicating which cases are complete. We apply it to the last 10 rows of the mydata2 dataset using the tail() function.

```
> mydata2.last10 = tail(mydata2,10)
> complete.cases(mydata2.last10)
[1] FALSE TRUE TRUE TRUE FALSE FALSE TRUE TRUE TRUE
```

Using the above function, we can get a dataset containing only the complete cases. (Note that the na.omit () function returns the same result.)

```
> mydata2.last10[complete.cases(mydata2.last10),]
    SEX WING THIRD MASS AGE
1958
    M 79 62 12.4 1Y
        79
    F
             61 11.6 1+
1959
1960 F 82
             63 12.5 1+
1964 M 82
             62 13.2 1+
1965
    М
         81
             61 13.9 1Y
1966 F 82 63 11.4 1Y
```

Or we can get the rows containing missing values by negating the expression with "!".

> mydata2.last10[!complete.cases(mydata2.last10),]
 SEX WING THIRD MASS AGE
1957 M NA NA 12.9 1Y
1961 <NA> 79 59 11.2 1Y
1962 <NA> 85 64 13.0 1+
1963 <NA> 82 62 13.3 1+

Certain arithmetic and descriptive functions (e.g. sum(), sd()) return NA when the focal object contains missing values. Using the generic na.rm = TRUE parameter may remove missing values in case of the specific function, however, we recommend to always refer to the help page of the function for details.

```
> mean(x)
[1] NA
> mean(x, na.rm = TRUE)
[1] 80.42857
```

#### Visualization of missing values

In this section, we introduce the reader to how the data and the structure of missing values can be explored simultaneously with useful plots of the VIM package (Kowarik & Templ 2016). We will use the FICHYP1.csv dataset (mydata), and its subset mydata2, and the previously used WING.table containing differently structured missing values (*Table1*).

#### **Aggregation Plot**

The highly customizable aggr() (see the help page) function produces a barplot and an aggregation plot. The barplot on the left of *Figure 1* shows a bar for each considered variable describing the frequencies of missing values (alternatively, the proportion can be shown using the prop = T (default) parameter). The aggregation plot on the right hand side shows all different missing value combinations horizontally that are present in the observations for the same variables. The grey colour indicates missingness, the white colour represents available data. (Default colours are red and light blue). The horizontal barplot to the right describes the frequencies of observations to the corresponding combinations. The proportion or frequencies of the different combinations can be represented by numbers by using the number = T parameter setting.

The first row from the bottom shows the most frequent combination (complete cases) that has the highest frequency (1281 observations). The second row shows the second most common combination (only SEX is missing). Adult Pied Flycatchers in general can be sexed with high certainty in all seasons, while juveniles in autumn can only be sexed after they initiate their postjuvenile moult. This relatively high number of missing values for SEX (287 cases) may be a result of trapped unmoulted juveniles. The third row shows that 174 observations are missing in both WING and THIRD. A potential explanation may be that ringers decided not to measure wing and third primary lengths due to high level of abrasion or active moult. We will further explore these patterns below.

```
> library(VIM)
> aggr(mydata2, col = c("white", "grey"), prop = F, numbers = T)
```



- *Figure 1.* Barplot and aggregation plot. Grey colour indicates missing, white colour represents available data
- 1. ábra Oszlopdiagram és kombinációs ábra. A szürke a hiányzó, a fehér a meglévő adatokat jelöli

With summary (aggr (mydata2)), summaries of the above graphs can be printed out.

```
> summary(aggr(mydata2))
 Missings per variable:
 Variable Count
      SEX
            401
            290
     WING
    THIRD
            319
     MASS
             80
      AGE
              0
 Missings in combinations of variables:
 Combinations Count
                       Percent
    0:0:0:0:0
              1284 65.3102747
    0:0:0:1:0
                 45
                     2.2889115
    0:0:1:0:0
                 29
                     1.4750763
    0:1:0:0:0
                 10
                     0.5086470
    0:1:1:0:0
              178
                     9.0539166
    0:1:1:1:0
                19
                     0.9664293
                294 14.9542218
    1:0:0:0:0
    1:0:0:1:0
                11
                     0.5595117
    1:0:1:0:0
                 13
                     0.6612411
    1:1:0:0:0
                  3
                     0.1525941
    1:1:1:0:0
                 75
                     3.8148525
    1:1:1:1:0 5 0.2543235
```

A. Harnos, T. Csörgő, & P. Fehérvári



*Figure 2.* Barplot and spineplot for the year variable. Grey colour indicates missing, white colour represents available data in SEX

 2. ábra Oszlopdiagram és spineplot a year változóra. A szürke a hiányzó, a fehér a meglévő adatokat jelöli a SEX változóban

#### Barplot/histogram and spineplot/spinogram with information on missing values

We can also make barplots (*Figure 2 a, 3 a*) and spineplots in case of categorical (*Figure 2 b, 3 b*) or histograms (*Figure 4 a, 4 c*) and spinograms (*Figure 4 b, 4 d*) in case of numerical variables with highlighting of missing values in other variables by splitting each bar into two parts. For spineplots (*Figure 2 b, 3 b*) and spinograms (*Figure 4 b, 4 d*) the width of the horizontal bars is scaled according to relative frequencies of the categories and the proportion of missing to the observed values are shown on the vertical axes (grey rectangles). Using the only.miss = F parameter setting, two additional barplots at the right appear on the plot. The first shows the number/ratio of missing values of the variable of interest (*Figure 2, 3*), the second one has meaning only in case of having imputed values. In case of only.miss = T (default) setting, only the missing values in the variable of interest are visualized.

Let us further explore the patterns of missing values in the SEX variable, by evaluating inter-annual proportions of NAs relative to the total number of ringed birds.

```
> x = mydata[, c("year", "SEX")]
> barMiss(x, col=c("white","grey"), only.miss = F)
> spineMiss(x, col=c("white","grey"), only.miss = F)
```





3. ábra Oszlopdiagram és spineplot az age változóra. A szürke a hiányzó, a fehér a meglévő adatokat jelöli a SEX változóban

*Figure 2 a* shows that the relative frequency of missing values is high until 1988 and in the last two years of the dataset. Potentially, the lack of available literature on sexing and ageing of passerines in the 80's is one of the causes of this pattern. A new principal ringer started working at the station in 2013, and the pattern suggests that he might have been less experienced in sexing Pied Flycatchers, therefore the slight increase in non-sexed individuals. However, the spineplot shows that these fluctuations in missing data relative frequency do not effect the bulk of the data, and excluding the aforementioned years, the temporal distribution of NAs seems fairly even.

```
> x <- mydata[, c("age", "SEX")]
> barMiss(x, col=c("white","grey"), only.miss = F)
> spineMiss(x, col=c("white","grey"), only.miss = F)
```

*Figure 3* shows the barplot (a) and spineplot (b) for missing values in case of age and SEX. It is apparent from the latter that missing SEX values are more frequent in case of juveniles, somewhat corroborating our previous assumptions, however there is a substantial number of unsexed adults, despite the fact that in general all adults can be readily sexed in all seasons.

```
> x <- mydata[, c("MASS", "WING")]
> barMiss(x, col=c("white","grey"), only.miss = F)
> spineMiss(x, col=c("white","grey"), only.miss = F)
```

To demonstrate how the barMiss() and spineMiss() functions depict continuous variables we used the MASS and WING variables (*Figure 4 a, b*). An MCAR pattern can be seen on the plot, since the proportion of missing values in WING seemingly does not dependent on MASS.





4. ábra Hisztogramok (a, b) és spinogramok (b, d) a MASS változóra. A szürke a hiányzó, a fehér a meglévő adatokat jelöli a WING változóban

Compare *Figure 4 a, b* to an artificial example (*Figure 4 c, d*) that resembles the behaviour of the "soft-hearted ringer" mentioned previously. Here, we deleted the WING values, that were measured on birds with MASS values below the 10% quantile (unmeasured birds in bad condition). The figures show a clear indication of a not MCAR pattern.

```
> barMiss(MASS.WING, col=c("white", "grey"), only.miss = F)
```

```
> spineMiss(MASS.WING, col=c("white","grey"), only.miss = F)
```



Figure 5. Mosaic plot for age and year. Grey colour indicates missing, white colour represents available data in SEX

5. ábra Mozaik ábra a age és year változókra. A szürke a hiányzó, a fehér a meglévő adatokat jelöli a SEX változóban

#### Mosaic plot with highlighted missing values

Mosaic plots are graphical representations of multi-way contingency tables (Harnos *et al.* 2015a). The frequencies of the different category combinations (cells) are visualized by area-proportional rectangles. Additionally, missing values in selected variables can be highlighted using the mosaicMiss() function. The first parameter is the data table, plotvars is a vector with the index of the variables to be plotted (age and year) and finally highlight is the index of the variable (SEX) to be used for highlighting. In our case the second and third variables in the subset table x are used for the mosaic plot, and the missing values in the first variable (SEX) are highlighted. The miss.labels = "F" deletes the labels for observed and missing values from the right side of the plot.

The figure shows that missing values in SEX nearly exclusively occur for juveniles in the first couple of years, and are seemingly more frequent in most of the following years of the study period. The figure also corroborates our previous assumptions regarding the inexperience of the ringer in the last two years; juveniles have disproportionately more missing values compared to adults and to previous years. Overall, there is also a sizeable proportion of adults with missing SEX values (see also *Figure 3*) despite that all adults can be sexed in hand. The only explanation here is human error and/or bad data recoding practice. However, we may assume (as we have no reason not to assume) that both males and females have equally been affected by missing values, therefore the an MCAR model can be assumed.

#### Scatterplots with Missing Data Information in the Margins

The marginplot () function (*Figure 6*) creates a scatterplot and also depicts information on missing values in the context of the variables of the x and y axes. Two box plots for missing and observed data are shown in the outer margins of the plots. The two parallel box plots along the vertical axis both represent the variable WING, but the light and dark grey boxplots were made from the WING values, where the corresponding MASS values are missing or available, respectively. (Note that the grey () function creates colours with different grey levels. 0 indicates black and 1 indicates white.) The plot additionally shows a univariate scatterplot (stripchart) of the values that are missing in the second variable (light grey points), and the frequency of these values by a number (290 and 80). The number in the lower left corner (here 24) is the number of observations that are missing in both variables.

```
> x <- mydata[, c("MASS","WING")]
> marginplot(x, col=c(grey(.5), grey(.8)))
```

At first glance this may be a bit confusing but (*Figure 6 b*) will help disentangle how the marginal boxplots relate to each other.

```
> marginplot(MASS.WING, col=c(grey(.5), grey(.8)))
```

*Figure 6 b* depicts the marginplot of the MASS variable and the WING variable, however we deleted the values of WING where the MASS was below the 10% quantile (i.e. the "softhearted ringer" effect). Comparing *Figure 6 a* and *Figure 6 b* the first apparent difference is that the scatterplot of the latter has a clear cutoff point. The marginal boxplots along the y axes are identical in both figures, however the horizontal boxplots of *Figure 6 b* are alarmingly different. The median of the light grey boxplot (showing MASS for individuals with missing wing measurements) is approximately equal to the lower bound of the dark grey box (showing MASS for individuals with existing WING measurements). This pattern indicates a not MCAR model, and should sound the alarm bell of any analyst.



- *Figure 6.* Scatterplots for the MASS and WING variables. Light grey colour indicates missing, dark grey colour represents available data in the other variable
- 6. ábra Szórásdiagram a MASS és WING változókra. A világosszürke jelöli a másik változóban hiányzó, a sötétszürke a meglévő adatokat

#### **Concluding remarks**

The methods described above are hopefully a good starting point to reveal the nature of missing values in our dataset, however we need to emphasize that by no means did we present the full set of available tools. For further information see e.g. the books of Little (2016) and Enders (2010). We also encourage our readers to refer to other functions of the VIM package (Kowarik & Templ 2016). Consult a statistician if unsure of the potential mechanisms and effects of the missing values in the dataset. Finally, we recommend that when publishing results, discuss in detail how missing values were treated and why they may or may not affect the inference.

#### Acknowledgements

The authors express their gratitude for the work of all the volunteers who collected data at the Ócsa Bird Ringing Station. We are grateful to János Kis and Imre Sándor Piross whom helped improve earlier versions of the manuscript. This work was supported by the National Scientific Research Fund of Hungary (OTKA under Grant No. 108571).

#### References

- Carpenter, J. & Kenward, M. 2012. Multiple imputation and its application. John Wiley & Sons
- Csörgő, T., Harnos, A., Rózsa, L., Karcza, Zs. & Fehérvári, P. 2016. Detailed description of the Ócsa Bird Ringing Station, Hungary. Ornis Hungarica 24(2): 91–108. DOI: 10.1515/orhu-2016-0018
- Dahl, D. B. 2016. xtable: Export tables to LaTeX or HTML. URL: https://CRAN.R-project.org/ package=xtable, R package version 1.8-2
- Enders, C. K. 2010. Applied missing data analysis. Guilford Press
- Fox, J. 2017. RcmdrMisc: R Commander Miscellaneous Functions. URL: https://CRAN.R-project.org/ package=RcmdrMisc, R package version 1.0-6
- Harnos, A., Csörgő, T. & Fehérvári, P. 2016a Hitchhikers' guide to analysing bird ringing data. Part 2. Ornis Hungarica 24(1): 172–181. DOI: 10.1515/orhu-2016-0010
- Harnos, A., Fehérvári, P. & Csörgő, T. 2015a Hitchhikers' guide to analysing bird ringing data. Part 1. Ornis Hungarica 23(2): 163–188. DOI: 10.1515/orhu-2015-0018
- Harnos, A., Fehérvári, P., Piross, I. S., Karcza, Zs., Ágh, N., Kovács, Sz. & Csörgő, T. 2016b Exploratory analyses of migration timing and morphometrics of the Pied Flycatcher (*Ficedula hypoleuca*). – Ornis Hungarica 24(2): 109–126. DOI: 10.1515/orhu-2016-0019
- Harnos, A., Lang, Zs., Fehérvári, P. & Csörgő, T. 2015b Sex and age dependent migration phenology of the Pied Flycatcher in a stopover site in the Carpathian Basin. – Ornis Hungarica 23(2): 10–19. DOI: 10.1515/orhu-2015-0010
- Kowarik, A. & Templ, M. 2016. Imputation with the R package VIM. Journal of Statistical Software 74(7): 1–16. DOI: 10.18637/jss.v074.i07
- Little, R. J. 2016. Missing Data/Imputation. The Encyclopedia of Adulthood and Aging 1–5 DOI: 10.1002/9781118528921.wbeaa147
- R Core Team 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria URL: https://www.R-project.org/
- Schafer, J. L. & Graham, J. W. 2002. Missing data: our view of the state of the art. Psychological Methods 7(2): 147–177.
- Templ, M., Alfons, A. & Filzmoser, P. 2012. Exploring incomplete data using visualization techniques. Advances in Data Analysis and Classification 6(1): 29–47. DOI: 10.1007/s11634-011-0102-y

#### ORNIS HUNGARICA 2017. 25(2)

- Templ, M., Alfons, A., Kowarik, A. & Prantner, B. 2015. VIM: Visualization and imputation of missing values. CRAN
- Xie, Y. 2014. knitr: A comprehensive tool for reproducible research in R. In: Stodden, V., Leisch, F. & Peng, R. D. (eds.) Implementing reproducible computational research. Chapman and Hall/CRC

Xie, Y. 2015. Dynamic documents with R and knitr. – Chapman and Hall/CRC, Boca Raton 2<sup>nd</sup> ed.

Xie, Y. 2016. knitr: A general-purpose package for dynamic report generation in R. – URL: http://yihui.name/knitr/, R package version 1.15.1



156