R Tips and Tricks

Marielle Winarto

2024-09-09

Table of contents

Preface

These are my personal notes taken while learning R. There are many great and useful resources on R available and I do not intend to copy them. However, I thought that in order to create a bit of order in my chaos it would be nice to jot down some thoughts while working with R and share the resulting document with the rest of the world.

Marielle Winarto

1 Introduction

This is a book created from markdown and executable code.

See Knuth (1984) for additional discussion of literate programming.

```
1 + 1
```

```
[1] 2
```

2 Handling missing data

There exists a specialized package for just about everything. One package that is pretty handy is naniar (Tierney and Cook 2023). It has functionality for finding missing values, visualizing where the missings are in a dataset (per variable or per case), and for handling them.

They even wrote a whole book about it, the Missing Book (Tierney and Horst 2022), which is absolutely worth a look.

2.1 Invisible missing data in longitudinal datasets

A noteworthy case is how to handle "invisible" missing data, for example in experience sampling data where participants have missed a couple of beeps that are simply not listed in the dataset. A way to handle this, is with the function `complete` from `tidyr`.

An example:

```
library(tidyverse)

sample_data <- expand.grid(participant=1:5, beep=1:10)
sample_data$score <- runif(50, 0, 100)

# Remove 10% of the data randomly
sample_data <- slice_sample(sample_data, prop = .7) %>%
  arrange(participant, beep)

slice_head(sample_data, n = 15)
```

```
   participant beep    score
1            1    1 62.98754
2            1    2 92.29533
3            1    3 17.75540
4            1    7 27.29456
5            1    8 19.05686
6            1    9 62.00391
7            2    1 43.66097
```

```
8            2    2 91.22910
9            2    3 22.70806
10           2    4 41.80714
11           2    5 88.05338
12           2    6 33.85436
13           2    7 17.42570
14           2    8 21.37298
15           2    9 12.61362
```

In this dataset there are no *real* missings – every cell in the table is filled with data, but there are missing beeps.

```
sample_with_explicit_missings <- sample_data %>%
  tidyr::complete(participant, beep)

slice_head(sample_with_explicit_missings, n = 15)
```

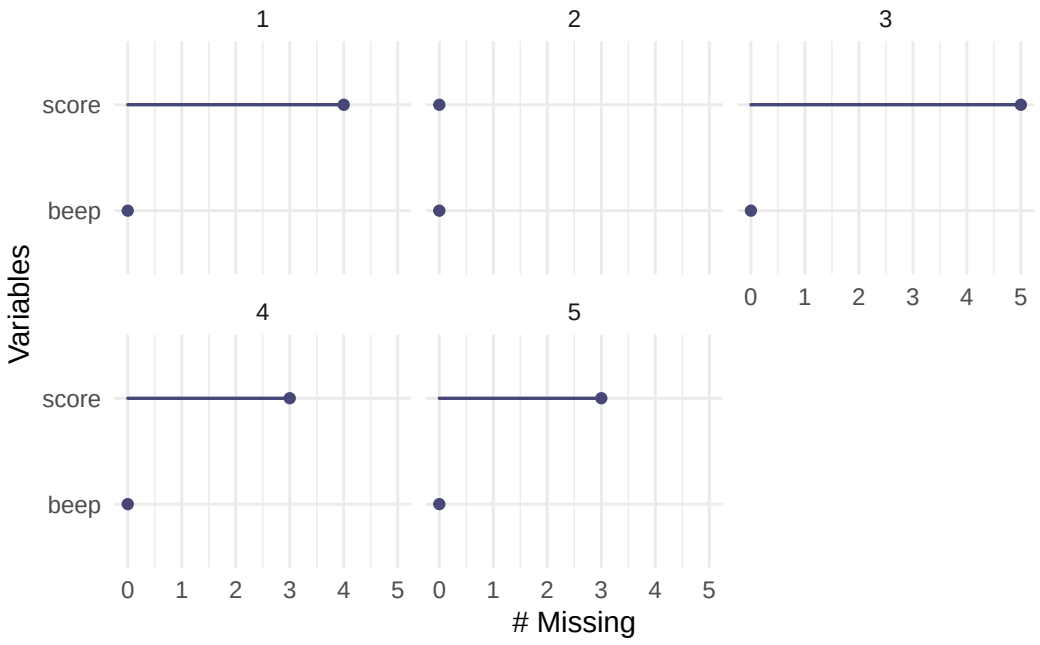
```
# A tibble: 15 x 3
   participant  beep score
         <int> <int> <dbl>
 1           1     1  63.0
 2           1     2  92.3
 3           1     3  17.8
 4           1     4  NA
 5           1     5  NA
 6           1     6  NA
 7           1     7  27.3
 8           1     8  19.1
 9           1     9  62.0
10           1    10  NA
11           2     1  43.7
12           2     2  91.2
13           2     3  22.7
14           2     4  41.8
15           2     5  88.1
```

The missing scores have now become explicit and are visible as `NA`.

```
library(naniar)
```

```
Warning: package 'naniar' was built under R version 4.3.3
```

```
gg_miss_var(sample_with_explicit_missings, facet = participant)
```



Tidyr also has a function `fill` to "fill down" missing data, i.e. fill the empty values in a column with the last non-empty value above it.

3 Summary

In summary, this book has no content whatsoever.

```
1 + 1
```

```
[1] 2
```

References

Knuth, Donald E. 1984. "Literate Programming." *Comput. J.* 27 (2): 97–111. https://doi.org/10.1093/comjnl/27.2.97.

Tierney, Nicholas, and Dianne Cook. 2023. "Expanding Tidy Data Principles to Facilitate Missing Data Exploration, Visualization and Assessment of Imputations." *Journal of Statistical Software* 105 (7): 1–31. https://doi.org/10.18637/jss.v105.i07.

Tierney, Nicholas, and Allison Horst. 2022. *The Missing Book*. https://tmb.njtierney.com/.