Stephanie Zimmer, Rebecca Powell, and Isabella Velásquez

Tidy Survey Book

To my son,

without whom I should have finished this book two years earlier

Contents

List of Figures

List of Tables

Preface

Hi there, this is my great book.

Why read this book

It is very important…

Structure of the book

Chapters **??** introduces a new topic, and …

Software information and conventions

I used the **knitr** package (Xie, 2015) and the **bookdown** package (Xie, 2022) to compile my book. My R session information is shown below:

```
xfun::session_info()
```

```
## R version 4.1.2 (2021-11-01)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur 10.16
##
## Locale: en_US.UTF-8 / en_US.UTF-8 / en_US.UTF-8 / C / en_US.UTF-8 / en_US.UTF-8
##
## Package version:
##   base64enc_0.1.3 bookdown_0.26   bslib_0.3.1
```

```
##    cli_3.3.0       compiler_4.1.2  digest_0.6.29
##    evaluate_0.15   fastmap_1.1.0   fs_1.5.2
##    glue_1.6.2      graphics_4.1.2  grDevices_4.1.2
##    highr_0.9       htmltools_0.5.2 jquerylib_0.1.4
##    jsonlite_1.8.0  knitr_1.39      magrittr_2.0.3
##    methods_4.1.2   R6_2.5.1        rappdirs_0.3.3
##    rlang_1.0.2     rmarkdown_2.14  rstudioapi_0.13
##    sass_0.4.1      stats_4.1.2     stringi_1.7.6
##    stringr_1.4.0   tinytex_0.39    tools_4.1.2
##    utils_4.1.2     xfun_0.31       yaml_2.3.5
```

Package names are in bold text (e.g., **rmarkdown**), and inline code and filenames are formatted in a typewriter font (e.g., `knitr::knit('foo.Rmd')`). Function names are followed by parentheses (e.g., `bookdown::render_book()`).

Acknowledgments

A lot of people helped me when I was writing the book.

<div align="right">

Frida Gomam
on the Mars

</div>

1

Introduction

2

Introducing survey data

3

Understanding survey data files

- Loading survey file formats
-

3.1 Loading survey files into R

Survey files come in different file types depending on the program used to output them. R gives us the flexibility to load datasets regardless of their file extension. We will cover the file types we are most likely to encounter in survey analysis: delimiter-separated text files, Excel files, `.dta` files from Stata, and `.sas` files from SAS.

3.1.1 Best practices for loading data into R

3.1.2 Loading delimiter-separated files into R

Delimiter-separated files use specific characters to separate values. For example, comma-separated values (CSV) files are separated by commas. These file types are widely used and we can use many applications to read and write them.

We can read delimiter-separated files using the tidyverse package {readr}. The {readr} package supports the following files with these `read_*()` functions:

- `read_csv()`: Comma-separated values (CSV) files
- `read_tsv()`: Tab-separated values (TSV) files
- `read_delim()`: Delimiter-separated files (CSV and TSV are important special cases)
- `read_fwf()`: Fixed-width files
- `read_table()`: Whitespace-separated files
- `read_log()`: Web log files

3.1.3   Loading Stata `.dta` files into R

Stata is a popular statistical software package for data manipulation, visualization, statistics, and automated reporting. Stata stores its data using the proprietary `.dta` format, a file type that can contain labels for the data's variables and values.

The {haven} package imports a variety of proprietary data formats, including Stata's .dta files, SAS's .sas files, and SPSS's .sav files. Once imported, we can analyze the resulting tibbles in R.

```
haven::read_dta("https://www.stata-press.com/data/r17/smho.dta")
```

```
## # A tibble: 120 x 10
##        id        hosptype exptotal  beds seencnt eoycnt
##     <dbl>        <dbl+lbl>    <dbl> <dbl>   <dbl>  <dbl>
## 1      14 1 [Psychiatric]       18   138    3083    308
## 2      19 1 [Psychiatric]        6    72    1557     15
## 3      30 1 [Psychiatric]       10    79    1822    191
## 4      32 1 [Psychiatric]        4    15    1201      0
## 5      44 1 [Psychiatric]        2    31    1346      0
## 6      47 1 [Psychiatric]       10   139    1739    126
## 7      52 1 [Psychiatric]       15   169    2614      0
## 8      54 1 [Psychiatric]       15    96    1137     82
## 9      58 1 [Psychiatric]       10    83    1783    164
## 10     62 1 [Psychiatric]       10   178    1956    305
## # ... with 110 more rows, and 4 more variables:
## #   findirct <dbl>, wt <dbl>, stratum <dbl>, N_h <dbl>
```

3.1.4   Loading SAS `.sas` files into R

Similar to the

3.2   Working with labelled data

We can use the `haven::labelled()` function to import labelled vectors into R. The goal is to create an intermediate object and then convert it into a regular R data frame to use in our analysis.

3.3 Missing data

For Stata and SAS, {haven} provides `tagged` missing values. This adds a single character label to R's regular `NA`.

Sources - https://www.stata-press.com/data/r17/svy.html

4

Introducing the srvyr package

5

Specifying sample designs in srvyr

6

Descriptive analyses in srvyr

7

Statistical testing

8

Modeling

9

Presenting results

A

More to Say

Yeah! I have finished my book, but I have more to say about some topics. Let me explain them in this appendix.

To know more about **bookdown**, see `https://bookdown.org`.

Bibliography

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.

Xie, Y. (2022). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.26.

Index

25