1   R you ready to write a paper in R Markdown?

2   Rick O. Gilmore[1,2] & Michael Hallquist[1]

3   [1] The Pennsylvania State University

4   [2] Databrary.org

5   Author Note

10   Correspondence concerning this article should be addressed to Rick O. Gilmore,

11   Department of Psychology, The Pennsylvania State University, University Park, PA 16802

12   USA. E-mail: rogilmore@psu.edu

Abstract

Want to write a paper using R Markdown? Keep reading to see how.

Keywords: APA, R Markdown

Word count: Not that many.

¹⁷   R you ready to write a paper in R Markdown?

¹⁸   It is possible to write an entire APA-formatted article in R Markdown. This very brief

¹⁹ paper shows how it might be done. As illustration, we use the data from a short, informal

²⁰ survey of participants in the 2018 R Bootcamp at Penn State.

²¹   **Methods**

²²   Consistent with open and transparent science practices, we report how we determined

²³ our sample size, all data exclusions (if any), all manipulations, and all measures in the study

²⁴ (Simmons, Nelson, & Simonsohn, 2011).

²⁵ **Participants**

²⁶   We asked participants in an optional "R Bootcamp" held at the Pennsylvania State

²⁷ University Department of Psychology on August 16 and 17, 2018 to complete an anonymous

²⁸ survey using a Google Form. We asked participants to report how old they felt. A total of

²⁹ $n = 56$ respondents answered the survey with a reported felt age of M=50.75 and a range of

³⁰ [5-1000] years.

³¹ **Material**

³²   The survey can be found at this URL: https://docs.google.com/forms/d/e/

³³ 1FAIpQLSeGqic9Hrj-XvkESZmu_0t6H02R-U6yzYnRLuX6HDFDp4R39g/viewform.

³⁴ There were six questions asked:

³⁵   1. Your current level of experience/expertise with R

³⁶   2. Your enthusiasm for banjo music?

³⁷   3. How old do you feel (in years)?

³⁸   4. Preferred number of hours spent sleeping/day

³⁹   5. Favorite day of the week?

⁴⁰   6. Is there a reproducibility "crisis" in psychology?

Procedure

We emailed a link to the survey to the list of participants in advance. We also include a link to the survey on the web page containing the course schedule (https://psu-psychology.github.io/r-bootcamp-2018/schedule.html). We encouraged participants to complete the survey before the first day or during lunch.

Data analysis

We used R (Version 3.5.1; R Core Team, 2018) and the R-packages *afex* (Version 0.21.2; Singmann, Bolker, Westfall, & Aust, 2018), *bindrcpp* (Version 0.2.2; Müller, 2018), *dataMaid* (Version 1.1.2; Petersen & Ekstrøm, 2018), *dplyr* (Version 0.7.6; Wickham, François, Henry, & Müller, 2018), *emmeans* (Version 1.2.3; Lenth, 2018), *forcats* (Version 0.3.0; Wickham, 2018a), *Formula* (Version 1.2.3; Zeileis & Croissant, 2010), *ggplot2* (Version 3.0.0; Wickham, 2016), *gmodels* (Version 2.18.1; Warnes et al., 2018), *googlesheets* (Version 0.3.0; Bryan & Zhao, 2018), *haven* (Version 1.1.2; Wickham & Miller, 2018), *Hmisc* (Version 4.1.1; Harrell Jr, Charles Dupont, & others., 2018), *lattice* (Version 0.20.35; Sarkar, 2008), *lme4* (Version 1.1.17; Bates, Mächler, Bolker, & Walker, 2015), *Matrix* (Version 1.2.14; Bates & Maechler, 2018), *pander* (Version 0.6.2; Daróczi & Tsegelskyi, 2018), *papaja* (Version 0.1.0.9709; Aust & Barth, 2018), *purrr* (Version 0.2.5; Henry & Wickham, 2018), *readr* (Version 1.1.1; Wickham, Hester, & Francois, 2017), *stringr* (Version 1.3.1; Wickham, 2018b), *survival* (Version 2.42.6; Terry M. Therneau & Patricia M. Grambsch, 2000), *tibble* (Version 1.4.2; Müller & Wickham, 2018), *tidyr* (Version 0.8.1; Wickham & Henry, 2018), *tidyverse* (Version 1.2.1; Wickham, 2017), and *tufte* (Version 0.4; Xie & Allaire, 2018) for all our analyses. The code used to generate these analyses is embedded in this document. To view it, see the R Markdown file in the GitHub repository associated with this paper.

₆₄ **Results**

₆₅   Table 1 summarizes the banjo music enthusiasm ratings data by levels of R experience.

₆₆ As Gilmore predicted, the more participants know about R, the more they come to

₆₇ appreciate banjo music.

₆₈   Let's examine the correlations between our continuous variables. As indicated in Table

₆₉ 2, there is a non-significant negative correlation ($r = -.15$, 95% CI $[-.40, .12]$) between

₇₀ banjo music enthusiasm and age ($t(54) = -1.10$, $p = .275$), no correlation ($r = .06$, 95% CI

₇₁ $[-.21, .31]$) between banjo music enthusiasm and sleep ($t(54) = 0.41$, $p = .683$), but a

₇₂ positive correlation ($r = .59$, 95% CI $[.39, .74]$) between age and sleep ($t(54) = 5.44$,

₇₃ $p < .001$). Figures 1 and 2 depict these patterns.

₇₄   To test the hypothesis that banjo music enthusiasm varies as a function of R expertise,

₇₅ we carried out a one-way ANOVA. R experience ($F(3, 51) = 0.51$, $MSE = 7.84$, $p = .679$,

₇₆ $\hat{\eta}_p^2 = .029$) did not predict enthusiasm for banjo music, so Gilmore will have to continue

₇₇ searching for userRs who appreciate the banjo. Table 3 summarizes these results.

₇₈ **Discussion**

₇₉   These results aren't going to set the world on fire, but they do show how awesome it

₈₀ can be to use R, R Markdown, and literate programming principles to conduct and open,

₈₁ transparent, and reproducible psychological science. Yay, us!

₈₂   There are no limitations to what we can accomplish using these tools. So, let's get to it.

References

83

84   Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown.*

85   Retrieved from https://github.com/crsh/papaja

86   Bates, D., & Maechler, M. (2018). *Matrix: Sparse and dense matrix classes and methods.*

87   Retrieved from https://CRAN.R-project.org/package=Matrix

88   Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models

89   using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

90   Bryan, J., & Zhao, J. (2018). *Googlesheets: Manage google spreadsheets from r.* Retrieved

91   from https://CRAN.R-project.org/package=googlesheets

92   Daróczi, G., & Tsegelskyi, R. (2018). *Pander: An r 'pandoc' writer.* Retrieved from

93   https://CRAN.R-project.org/package=pander

94   Harrell Jr, F. E., Charles Dupont, & others. (2018). *Hmisc: Harrell miscellaneous.*

95   Retrieved from https://CRAN.R-project.org/package=Hmisc

96   Henry, L., & Wickham, H. (2018). *Purrr: Functional programming tools.* Retrieved from

97   https://CRAN.R-project.org/package=purrr

98   Lenth, R. (2018). *Emmeans: Estimated marginal means, aka least-squares means.* Retrieved

99   from https://CRAN.R-project.org/package=emmeans

100   Müller, K. (2018). *Bindrcpp: An 'rcpp' interface to active bindings.* Retrieved from

101   https://CRAN.R-project.org/package=bindrcpp

102   Müller, K., & Wickham, H. (2018). *Tibble: Simple data frames.* Retrieved from

103   https://CRAN.R-project.org/package=tibble

104   Petersen, A. H., & Ekstrøm, C. T. (2018). *DataMaid: A suite of checks for identification of*

105   *potential errors in a data frame as part of the data screening process.* Retrieved from

106   https://CRAN.R-project.org/package=dataMaid

107   R Core Team. (2018). *R: A language and environment for statistical computing.* Vienna,

108   Austria: R Foundation for Statistical Computing. Retrieved from

https://www.R-project.org/

Sarkar, D. (2008). *Lattice: Multivariate data visualization with r.* New York: Springer.
Retrieved from http://lmdvr.r-forge.r-project.org

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology:
Undisclosed flexibility in data collection and analysis allows presenting anything as
significant. *Psychol. Sci.*, *22*(11), 1359–1366. Retrieved from
http://journals.sagepub.com/doi/abs/10.1177/0956797611417632

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2018). *Afex: Analysis of factorial
experiments.* Retrieved from https://CRAN.R-project.org/package=afex

Terry M. Therneau, & Patricia M. Grambsch. (2000). *Modeling survival data: Extending the
Cox model.* New York: Springer.

Warnes, G. R., Bolker, B., Lumley, T., Randall C. Johnson are Copyright SAIC-Frederick, R.
C. J. C. from, Intramural Research Program, I. F. by the, NIH, . . . Cancer Research
under NCI Contract NO1-CO-12400., C. for. (2018). *Gmodels: Various r
programming tools for model fitting.* Retrieved from
https://CRAN.R-project.org/package=gmodels

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis.* Springer-Verlag New York.
Retrieved from http://ggplot2.org

Wickham, H. (2017). *Tidyverse: Easily install and load the 'tidyverse'.* Retrieved from
https://CRAN.R-project.org/package=tidyverse

Wickham, H. (2018a). *Forcats: Tools for working with categorical variables (factors).*
Retrieved from https://CRAN.R-project.org/package=forcats

Wickham, H. (2018b). *Stringr: Simple, consistent wrappers for common string operations.*
Retrieved from https://CRAN.R-project.org/package=stringr

Wickham, H., & Henry, L. (2018). *Tidyr: Easily tidy data with 'spread()' and 'gather()'
functions.* Retrieved from https://CRAN.R-project.org/package=tidyr

Wickham, H., & Miller, E. (2018). *Haven: Import and export 'spss', 'stata' and 'sas' files.*

136   Retrieved from https://CRAN.R-project.org/package=haven

137   Wickham, H., François, R., Henry, L., & Müller, K. (2018). *Dplyr: A grammar of data*

138   *manipulation.* Retrieved from https://CRAN.R-project.org/package=dplyr

139   Wickham, H., Hester, J., & Francois, R. (2017). *Readr: Read rectangular text data.*

140   Retrieved from https://CRAN.R-project.org/package=readr

141   Xie, Y., & Allaire, J. (2018). *Tufte: Tufte's styles for r markdown documents.* Retrieved

142   from https://CRAN.R-project.org/package=tufte

143   Zeileis, A., & Croissant, Y. (2010). Extended model formulas in R: Multiple parts and

144   multiple responses. *Journal of Statistical Software*, *34*(1), 1–13.

145   doi:10.18637/jss.v034.i01

Table 1

Descriptive statistics of banjo music enthusiasm
by R experience.

R_exp	Mean	Median	SD	Min	Max
none	3.90	4.00	2.08	1.00	7.00
limited	4.75	4.00	3.05	1.00	10.00
lots	4.09	3.00	2.02	1.00	7.00
pro	6.00	6.00	5.66	2.00	10.00
NA	1.00	1.00	NA	1.00	1.00

Note. This table was created with apa_table()

Table 2

Correlation table of the example data set.

	Banjo	Psych_age_yrs
Banjo		
Psych_age_yrs	-0.15	
Sleep_hrs	0.06	0.59***

Note. This is a correlation table created using apa_table().

Table 3

ANOVA table for the analyis of the example data set.

Effect	F	df_1	df_2	MSE	p	$\hat{\eta}_p^2$
R exp	0.51	3	51	7.84	.679	.029

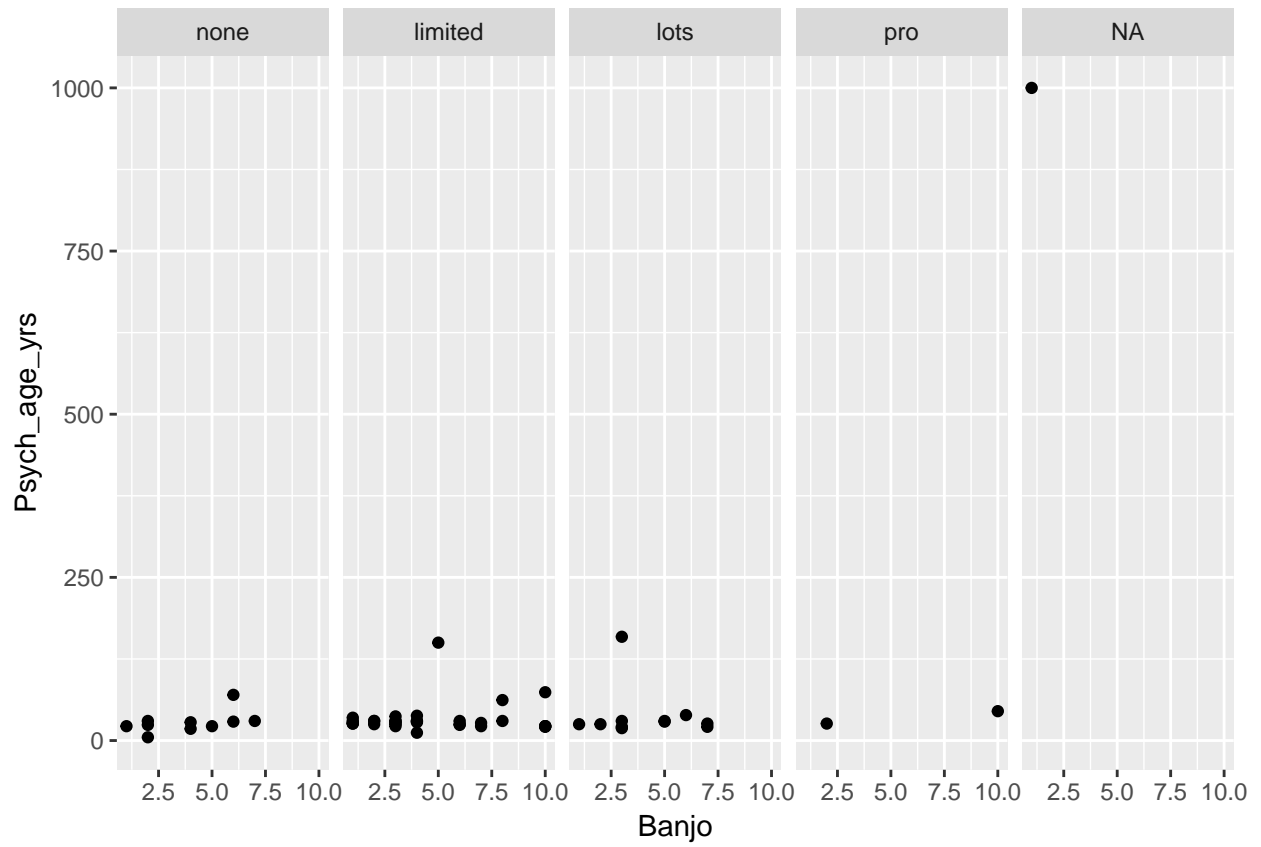*Note.* This is a table created using apa_print() and apa_table().
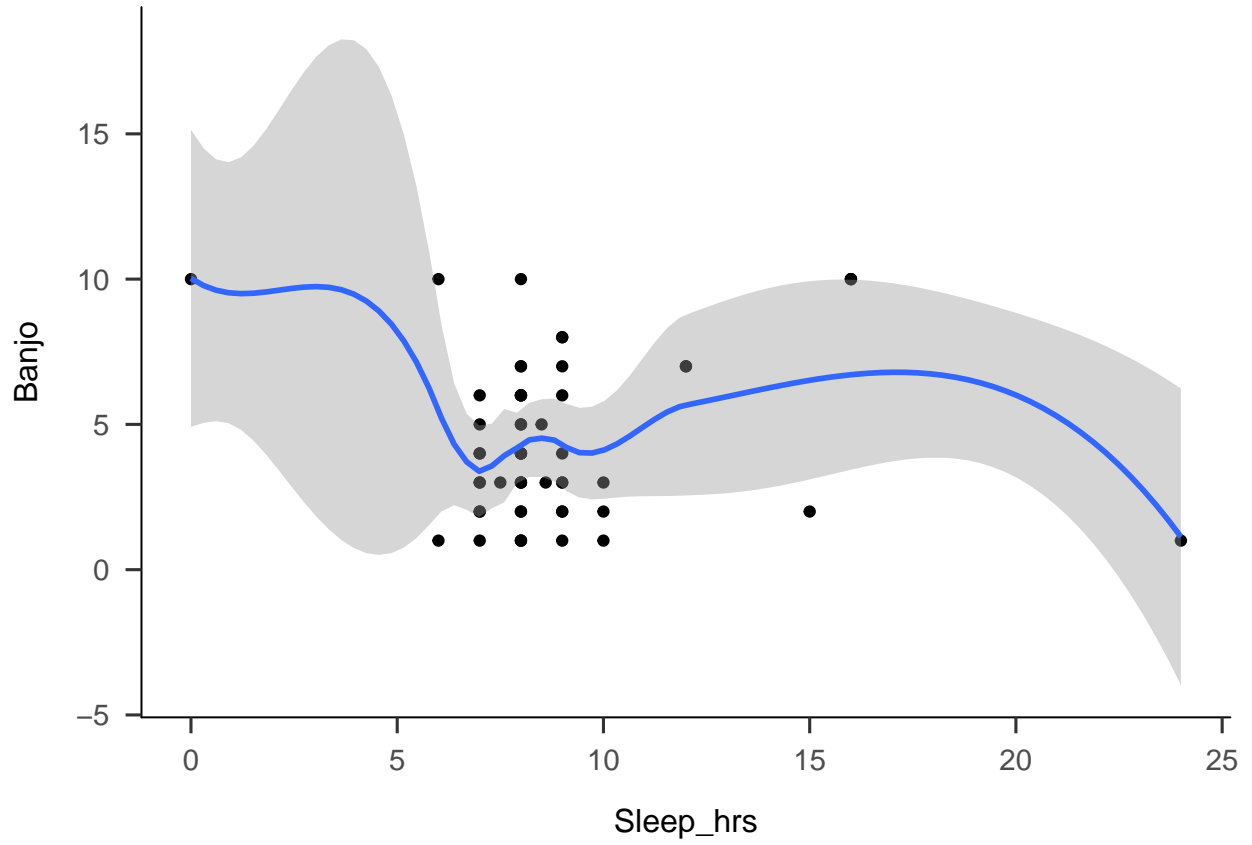
Figure 1. Banjo music enthusiasm by age and R experience

Figure 2. Banjo music enthusiasm by preferred hours of sleep