

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

2 Rick O. Gilmore^{1,2} & Michael Hallquist¹

3 ¹ The Pennsylvania State University

4 ² Databrary.org

5 Author Note

6 The authors are with the Department of Psychology at The Pennsylvania State
7 University. The authors acknowledge support from the Department of Psychology, the Social,
8 Life, & Engineering Sciences Imaging Center (SLEIC), and the Child Study Center's Open
9 Data in Developmental Science (ODDS) initiative.

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

13

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

15 *Keywords:* APA, R Markdown

16 Word count: Not that many.

R you ready to write a paper in R Markdown?

It is possible to write an entire APA-formatted article in R Markdown. This very brief paper shows how it might be done. As illustration, we use the data from a short, informal survey of participants in the 2018 R Bootcamp at Penn State.

Methods

Consistent with open and transparent science practices, we report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study (Simmons, Nelson, & Simonsohn, 2011).

Participants

We asked participants in an optional “R Bootcamp” held at the Pennsylvania State University Department of Psychology on August 16 and 17, 2018 to complete an anonymous survey using a Google Form. We asked participants to report how old they felt. A total of $n = 46$ respondents answered the survey with a reported felt age of $M=55.33$ and a range of [5-1000] years.

Material

The survey can be found at this URL: https://docs.google.com/forms/d/e/1FAIpQLSeGqic9Hrj-XvkESZmu_0t6H02R-U6yzYnRLuX6HDFDp4R39g/viewform.

There were six questions asked:

1. Your current level of experience/expertise with R
2. Your enthusiasm for banjo music?
3. How old do you feel (in years)?
4. Preferred number of hours spent sleeping/day
5. Favorite day of the week?
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, 2017b) and the R-packages *afex* (Version 0.21.2; Singmann, Bolker, Westfall, & Aust, 2018), *bindrcpp* (Version 0.2.2; Müller, 2016), *dplyr* (Version 0.7.6; Wickham & Francois, 2016), *emmeans* (Version 1.2.3; Lenth, 2018), *forcats* (Version 0.3.0; Wickham, 2018a), *foreign* (Version 0.8.71; R Core Team, 2017a), *Formula* (Version 1.2.3; Zeileis & Croissant, 2010), *ggplot2* (Version 3.0.0; Wickham, 2009), *gmodels* (Version 2.18.1; Warnes et al., 2015), *googlesheets* (Version 0.3.0; Bryan & Zhao, 2017), *Hmisc* (Version 4.1.1; Harrell Jr, Charles Dupont, & others., 2017), *lattice* (Version 0.20.35; Sarkar, 2008), *lme4* (Version 1.1.17; Bates, Mächler, Bolker, & Walker, 2015), *MASS* (Version 7.3.50; Venables & Ripley, 2002), *Matrix* (Version 1.2.14; Bates & Maechler, 2018), *multilevel* (Version 2.6; Bliese, 2016), *nlme* (Version 3.1.137; Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2017), *papaja* (Version 0.1.0.9709; Aust & Barth, 2017), *plyr* (Wickham, 2011; Version 1.8.4; Wickham & Francois, 2016), *psych* (Version 1.8.4; Revelle, 2017), *purrr* (Version 0.2.5; Henry & Wickham, 2017), *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; Wickham, Francois, & Müller, 2017), *tidyr* (Version 0.8.1; Wickham, 2017a), and *tidyverse* (Version 1.2.1; Wickham, 2017b) 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.

Results

Table 1 summarizes the banjo music enthusiasm ratings data by levels of R experience. As Gilmore predicted, the more participants know about R, the more they come to appreciate banjo music.

Let's examine the correlations between our continuous variables. As indicated in Table 2, there is a non-significant negative correlation ($r = -.18$, 95% CI $[-.44, .12]$) between banjo music enthusiasm and age ($t(44) = -1.18$, $p = .244$), no correlation ($r = .08$, 95% CI $[-.22, .36]$) between banjo music enthusiasm and sleep ($t(44) = 0.50$, $p = .620$), but a positive correlation ($r = .60$, 95% CI $[.38, .76]$) between age and sleep ($t(44) = 4.99$, $p < .001$). Figures 1 and 2 depict these patterns.

To test the hypothesis that banjo music enthusiasm varies as a function of R expertise, we carried out a one-way ANOVA. R experience ($F(3, 41) = 0.19$, $MSE = 8.35$, $p = .902$, $\hat{\eta}_p^2 = .014$) did not predict enthusiasm for banjo music, so Gilmore will have to continue searching for userRs who appreciate the banjo. Table 3 summarizes these results.

Discussion

These results aren't going to set the world on fire, but they do show how awesome it can be to use R, R Markdown, and literate programming principles to conduct and open, transparent, and reproducible psychological science. Yay, us!

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

References

- Aust, F., & Barth, M. (2017). *papaja: Create APA manuscripts with R Markdown*. Retrieved from <https://github.com/crsh/papaja>
- Bates, D., & Maechler, M. (2018). *Matrix: Sparse and dense matrix classes and methods*. Retrieved from <https://CRAN.R-project.org/package=Matrix>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi:10.18637/jss.v067.i01
- Bliese, P. (2016). *Multilevel: Multilevel functions*. Retrieved from <https://CRAN.R-project.org/package=multilevel>
- Bryan, J., & Zhao, J. (2017). *Googlesheets: Manage google spreadsheets from r*. Retrieved from <https://CRAN.R-project.org/package=googlesheets>
- Harrell Jr, F. E., Charles Dupont, & others. (2017). *Hmisc: Harrell miscellaneous*.
- Henry, L., & Wickham, H. (2017). *Purrr: Functional programming tools*. Retrieved from <https://CRAN.R-project.org/package=purrr>
- Lenth, R. (2018). *Emmeans: Estimated marginal means, aka least-squares means*. Retrieved from <https://CRAN.R-project.org/package=emmeans>
- Müller, K. (2016). *Bindrcpp: An 'rcpp' interface to active bindings*. Retrieved from <https://CRAN.R-project.org/package=bindrcpp>
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2017). *nlme: Linear and nonlinear mixed effects models*. Retrieved from <https://CRAN.R-project.org/package=nlme>
- R Core Team. (2017a). *Foreign: Read data stored by 'minitab', 's', 'sas', 'spss', 'stata', 'sysstat', 'weka', 'dBase', ...*. Retrieved from <https://CRAN.R-project.org/package=foreign>
- R Core Team. (2017b). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from

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

Revelle, W. (2017). *Psych: Procedures for psychological, psychometric, and personality research*. Evanston, Illinois: Northwestern University. Retrieved from

<https://CRAN.R-project.org/package=psych>

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.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (Fourth.). New York: Springer. Retrieved from <http://www.stats.ox.ac.uk/pub/MASS4>

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. (2015). *Gmodels: Various r programming tools for model fitting*. Retrieved from

<https://CRAN.R-project.org/package=gmodels>

Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from <http://ggplot2.org>

Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, 40(1), 1–29. Retrieved from <http://www.jstatsoft.org/v40/i01/>

Wickham, H. (2017a). *Tidyr: Easily tidy data with 'spread()' and 'gather()' functions*.

Retrieved from <https://CRAN.R-project.org/package=tidyr>

Wickham, H. (2017b). *Tidyverse: Easily install and load 'tidyverse' packages*. 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., & Francois, R. (2016). *Dplyr: A grammar of data manipulation*. Retrieved from <https://CRAN.R-project.org/package=dplyr>

Wickham, H., Francois, R., & Müller, K. (2017). *Tibble: Simple data frames*. Retrieved from <https://CRAN.R-project.org/package=tibble>

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

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

Zeileis, A., & Croissant, Y. (2010). Extended model formulas in R: Multiple parts and multiple responses. *Journal of Statistical Software*, 34(1), 1–13. Retrieved from

<http://www.jstatsoft.org/v34/i01/>

Table 1

*Descriptive statistics of banjo music enthusiasm
by R experience.*

R_exp	Mean	Median	SD	Min	Max
none	4.33	5.00	2.42	1.00	7.00
limited	4.48	3.00	3.11	1.00	10.00
lots	4.40	4.00	1.84	2.00	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	-0.18	
Sleep_hrs	0.08	0.60***

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

Table 3

ANOVA table for the analysis of the example data set.

Effect	F	df_1	df_2	MSE	p	$\hat{\eta}_p^2$
R exp	0.19	3	41	8.35	.902	.014

Note. This is a table created using `apa_print()` and `apa_table()`.

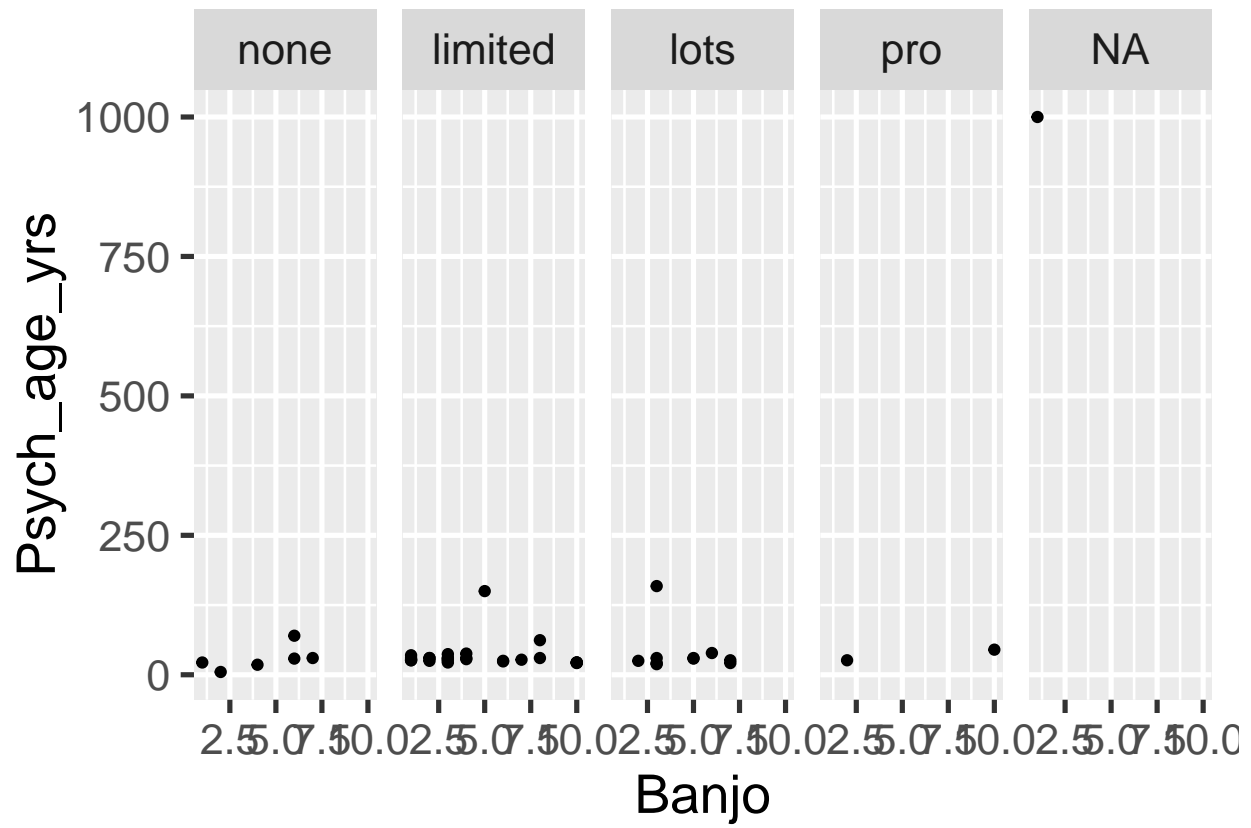


Figure 1. Banjo music enthusiasm by age and R experience

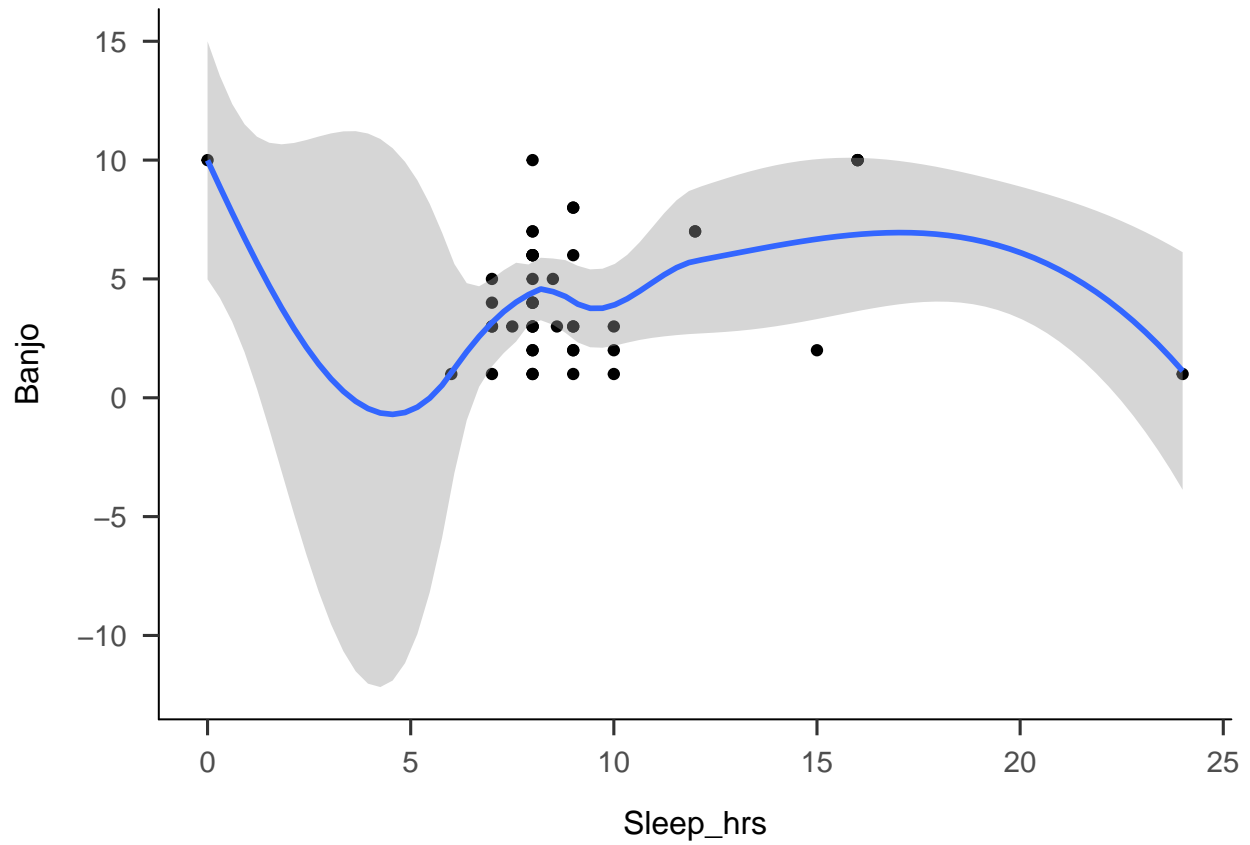


Figure 2. Banjo music enthusiasm by preferred hours of sleep