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

Abstract

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.

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 $n = 37$ respondents answered the survey with a reported felt age of $M{=}62.54$ and a range of

30 [19-1000] 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

⁴¹

⁴²   We emailed a link to the survey to the list of participants in advance. We also include

⁴³ a link to the survey on the web page containing the course schedule

⁴⁴ (https://psu-psychology.github.io/r-bootcamp-2018/schedule.html). We encouraged

⁴⁵ participants to complete the survey before the first day or during lunch.

Data analysis

⁴⁶

⁴⁷   We used R (Version 3.5.1; R Core Team, 2017b) and the R-packages *afex* (Version

⁴⁸ 0.21.2; Singmann, Bolker, Westfall, & Aust, 2018), *bindrcpp* (Version 0.2.2; Müller, 2016),

⁴⁹ *dplyr* (Version 0.7.6; Wickham & Francois, 2016), *emmeans* (Version 1.2.3; Lenth, 2018),

⁵⁰ *forcats* (Version 0.3.0; Wickham, 2018a), *foreign* (Version 0.8.71; R Core Team, 2017a),

⁵¹ *Formula* (Version 1.2.3; Zeileis & Croissant, 2010), *ggplot2* (Version 3.0.0; Wickham, 2009),

⁵² *gmodels* (Version 2.18.1; Warnes et al., 2015), *googlesheets* (Version 0.3.0; Bryan & Zhao,

⁵³ 2017), *Hmisc* (Version 4.1.1; Harrell Jr, Charles Dupont, & others., 2017), *lattice* (Version

⁵⁴ 0.20.35; Sarkar, 2008), *lme4* (Version 1.1.17; Bates, Mächler, Bolker, & Walker, 2015), *MASS*

⁵⁵ (Version 7.3.50; Venables & Ripley, 2002), *Matrix* (Version 1.2.14; Bates & Maechler, 2018),

⁵⁶ *multilevel* (Version 2.6; Bliese, 2016), *nlme* (Version 3.1.137; Pinheiro, Bates, DebRoy, Sarkar,

⁵⁷ & R Core Team, 2017), *papaja* (Version 0.1.0.9709; Aust & Barth, 2017), *plyr* (Wickham,

⁵⁸ 2011; Version 1.8.4; Wickham & Francois, 2016), *psych* (Version 1.8.4; Revelle, 2017), *purrr*

⁵⁹ (Version 0.2.5; Henry & Wickham, 2017), *readr* (Version 1.1.1; Wickham, Hester, & Francois,

⁶⁰ 2017), *stringr* (Version 1.3.1; Wickham, 2018b), *survival* (Version 2.42.6; Terry M. Therneau

⁶¹ & Patricia M. Grambsch, 2000), *tibble* (Version 1.4.2; Wickham, Francois, & Müller, 2017),

⁶² *tidyr* (Version 0.8.1; Wickham, 2017a), and *tidyverse* (Version 1.2.1; Wickham, 2017b) for all

⁶³ our analyses. The code used to generate these analyses is embedded in this document. To

⁶⁴ view it, see the R Markdown file in the GitHub repository associated with this paper.

⁶⁵ **Results**

⁶⁶   Table 1 summarizes the banjo music enthusiasm ratings data by levels of R experience.

⁶⁷ As Gilmore predicted, the more participants know about R, the more they come to

⁶⁸ appreciate banjo music.

⁶⁹   Let's examine the correlations between our continuous variables. As indicated in Table

⁷⁰ 2, there is a non-significant negative correlation ($r = -.21$, 95% CI $[-.50, .12]$) between

⁷¹ banjo music enthusiasm and age ($t(35) = -1.27$, $p = .211$), no correlation ($r = .04$, 95% CI

⁷² $[-.29, .36]$) between banjo music enthusiasm and sleep ($t(35) = 0.24$, $p = .813$), but a

⁷³ positive correlation ($r = .60$, 95% CI $[.34, .77]$) between age and sleep ($t(35) = 4.42$,

⁷⁴ $p < .001$). Figures 1 and 2 depict these patterns.

⁷⁵   To test the hypothesis that banjo music enthusiasm varies as a function of R expertise,

⁷⁶ we carried out a one-way ANOVA. R experience ($F(3, 32) = 0.36$, $MSE = 9.40$, $p = .780$,

⁷⁷ $\hat{\eta}_p^2 = .033$) did not predict enthusiasm for banjo music, so Gilmore will have to continue

⁷⁸ searching for userRs who appreciate the banjo. Table 3 summarizes these results.

⁷⁹ **Discussion**

⁸⁰   These results aren't going to set the world on fire, but they do show how awesome it

⁸¹ can be to use R, R Markdown, and literate programming principles to conduct and open,

⁸² transparent, and reproducible psychological science. Yay, us!

⁸³   There are no limitations to what we can accomplish using these tools. So, let's get to it.

References

84

85 Aust, F., & Barth, M. (2017). *papaja: Create APA manuscripts with R Markdown.*

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

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

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

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

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

91 Bliese, P. (2016). *Multilevel: Multilevel functions.* Retrieved from

92   https://CRAN.R-project.org/package=multilevel

93 Bryan, J., & Zhao, J. (2017). *Googlesheets: Manage google spreadsheets from r.* Retrieved

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

95 Harrell Jr, F. E., Charles Dupont, & others. (2017). *Hmisc: Harrell miscellaneous.*

96 Henry, L., & Wickham, H. (2017). *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. (2016). *Bindrcpp: An 'rcpp' interface to active bindings.* Retrieved from

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

102 Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2017). *nlme: Linear and*

103   *nonlinear mixed effects models.* Retrieved from

104   https://CRAN.R-project.org/package=nlme

105 R Core Team. (2017a). *Foreign: Read data stored by 'minitab', 's', 'sas', 'spss', 'stata',*

106   *'systat', 'weka', 'dBase', …* Retrieved from

107   https://CRAN.R-project.org/package=foreign

108 R Core Team. (2017b). *R: A language and environment for statistical computing.* Vienna,

109   Austria: R Foundation for Statistical Computing. Retrieved from

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

111   Revelle, W. (2017). *Psych: Procedures for psychological, psychometric, and personality*

112   *research.* Evanston, Illinois: Northwestern University. Retrieved from

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

114   Sarkar, D. (2008). *Lattice: Multivariate data visualization with r.* New York: Springer.

115   Retrieved from http://lmdvr.r-forge.r-project.org

116   Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology:

117   Undisclosed flexibility in data collection and analysis allows presenting anything as

118   significant. *Psychol. Sci.*, *22*(11), 1359–1366. Retrieved from

119   http://journals.sagepub.com/doi/abs/10.1177/0956797611417632

120   Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2018). *Afex: Analysis of factorial*

121   *experiments.* Retrieved from https://CRAN.R-project.org/package=afex

122   Terry M. Therneau, & Patricia M. Grambsch. (2000). *Modeling survival data: Extending the*

123   *Cox model.* New York: Springer.

124   Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (Fourth.). New

125   York: Springer. Retrieved from http://www.stats.ox.ac.uk/pub/MASS4

126   Warnes, G. R., Bolker, B., Lumley, T., Randall C. Johnson are Copyright SAIC-Frederick, R.

127   C. J. C. from, Intramural Research Program, I. F. by the, NIH, . . . Cancer Research

128   under NCI Contract NO1-CO-12400., C. for. (2015). *Gmodels: Various r*

129   *programming tools for model fitting.* Retrieved from

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

131   Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis.* Springer-Verlag New York.

132   Retrieved from http://ggplot2.org

133   Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of*

134   *Statistical Software*, *40*(1), 1–29. Retrieved from http://www.jstatsoft.org/v40/i01/

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

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

137   Wickham, H. (2017b). *Tidyverse: Easily install and load 'tidyverse' packages.* Retrieved

138   from https://CRAN.R-project.org/package=tidyverse

139   Wickham, H. (2018a). *Forcats: Tools for working with categorical variables (factors).*

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

141   Wickham, H. (2018b). *Stringr: Simple, consistent wrappers for common string operations.*

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

143   Wickham, H., & Francois, R. (2016). *Dplyr: A grammar of data manipulation.* Retrieved

144   from https://CRAN.R-project.org/package=dplyr

145   Wickham, H., Francois, R., & Müller, K. (2017). *Tibble: Simple data frames.* Retrieved from

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

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

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

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

150   multiple responses. *Journal of Statistical Software, 34* (1), 1–13. Retrieved from

151   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	6.50	6.50	0.71	6.00	7.00
limited	4.77	3.50	3.37	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	-0.21	
Sleep_hrs	0.04	0.60***

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	0.36	3	32	9.40	.780	.033

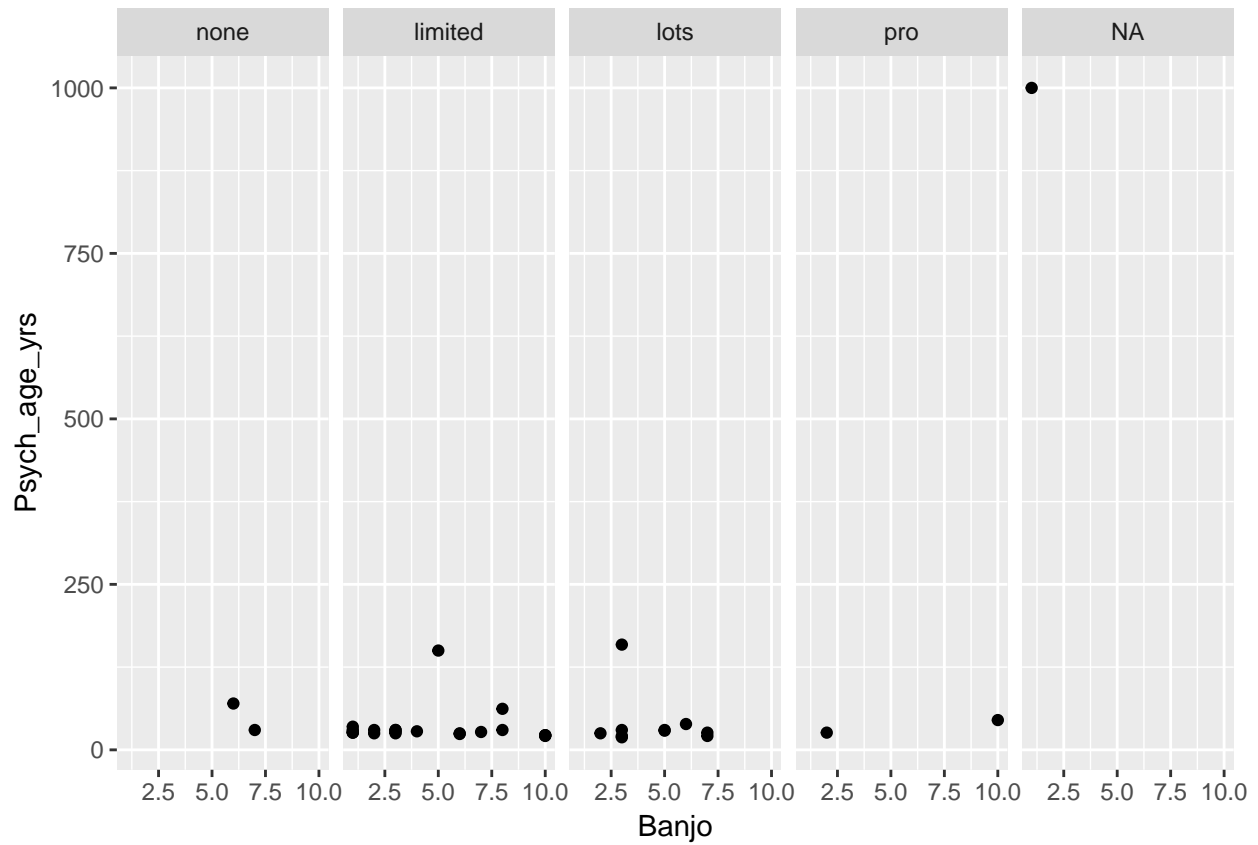*Note.* This is a table created using apa_print()

and apa_table().

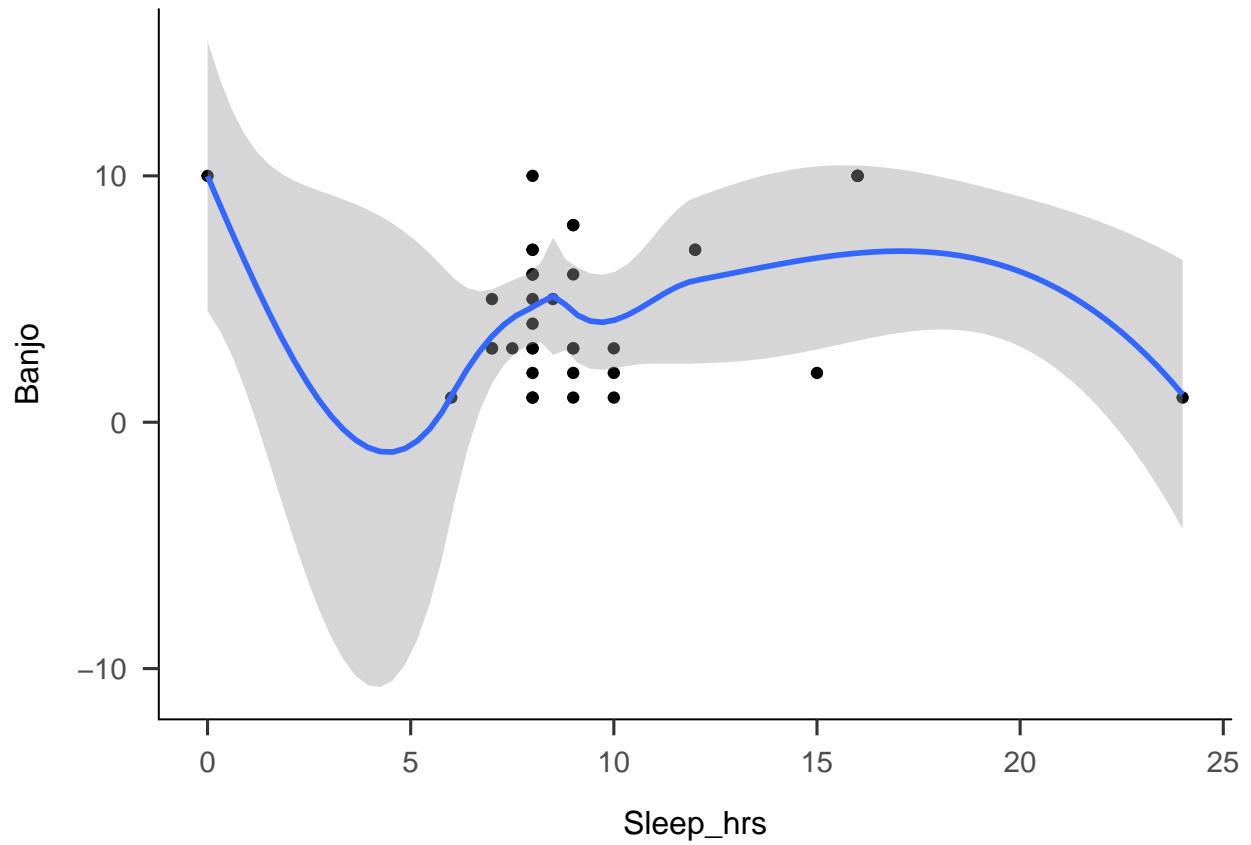*Figure 1*. Banjo music enthusiasm by age and R experience

Figure 2. Banjo music enthusiasm by preferred hours of sleep