

R you ready for some data?

Rick O. Gilmore<sup>1,2</sup>, James LeBreton<sup>1</sup>, & Michael Hallquist<sup>1</sup>

<sup>1</sup> The Pennsylvania State University

<sup>2</sup> Databrary.org

#### Author Note

The authors are with the Department of Psychology at The Pennsylvania State University. The authors acknowledge support from the Department of Psychology and the Social, Life, & Engineering Sciences Imaging Center (SLEIC).

Correspondence concerning this article should be addressed to Rick O. Gilmore, Department of Psychology, The Pennsylvania State University, University Park, PA 16802 USA. E-mail: [rogilmore@psu.edu](mailto:rogilmore@psu.edu)

## Abstract

12

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

14 *Keywords:* APA, R Markdown

15 Word count: Not that many.

R you ready for some data?

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 brief, informal survey of participants in the inaugural R Bootcamp at Penn State. We predicted that higher levels of enthusiasm for “Game of Thrones” would be reported by respondents with *lower* reported hours/day of preferred sleep, at least among younger respondents.

## 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 to complete an anonymous survey using a Google Form. We asked participants to report their age in years. A total of 50 respondents answered the survey with a reported age of [22-55] years.

## Material

The survey can be found at this URL: [https://docs.google.com/forms/d/1l5OX8PcN\\_lfVn3ykr\\_PtHCzhRbWzMbxhqtgILD45zRg/edit](https://docs.google.com/forms/d/1l5OX8PcN_lfVn3ykr_PtHCzhRbWzMbxhqtgILD45zRg/edit).

There were five questions asked:

1. Your current level of experience/expertise with R
2. Your enthusiasm for Game of Thrones [1..10 scale]
3. Age in years
4. Preferred number of hours spent sleeping/day
5. Favorite day of the week?

## 6. Are your data tidy?

### Procedure

We emailed a link to the survey to the list of participants. We also include a link to the survey on the web page containing the course schedule (<https://psu-psychology.github.io/r-bootcamp/schedule.html>). We encouraged participants to complete the survey after the first day's material.

### Data analysis

We used R (3.4.1, R Core Team, 2017b) and the R-packages *bindrcpp* (0.2, Müller, 2016), *dplyr* (0.7.2, Wickham & Francois, 2016), *foreign* (0.8.69, R Core Team, 2017a), *Formula* (1.2.2, Zeileis & Croissant, 2010), *ggplot2* (2.2.1, Wickham, 2009), *gmodels* (2.16.2, Warnes et al., 2015), *googlesheets* (0.2.2, Bryan & Zhao, 2017), *Hmisc* (4.0.3, Harrell Jr, Charles Dupont, & others., 2017), *lattice* (0.20.35, Sarkar, 2008), *MASS* (7.3.47, Venables & Ripley, 2002), *multilevel* (2.6, Bliese, 2016), *nlme* (3.1.131, Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2017), *papaja* (0.1.0.9492, Aust & Barth, 2017), *plyr* (Wickham, 2011; 1.8.4, Wickham & Francois, 2016), *psych* (1.7.5, Revelle, 2017), *purrr* (0.2.3, Henry & Wickham, 2017), *readr* (1.1.1, Wickham, Hester, & Francois, 2017), *survival* (2.41.3, Terry M. Therneau & Patricia M. Grambsch, 2000), *tibble* (1.3.3, Wickham, Francois, & Müller, 2017), *tidyr* (0.6.3, Wickham, 2017a), and *tidyverse* (1.1.1, 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 Game of Thrones ratings data by levels of R experience.

Let's examine the correlations between our continuous variables. As indicated in Table 2, there is a negative correlation ( $r = -.93$ , 95% CI  $[-.96, -.88]$ ) between Game of Thrones enthusiasm and age ( $t(48) = -17.78$ ,  $p < .001$ ), a negative correlation ( $r = -.25$ , 95% CI

65  $[-.49, .03]$ ) between Game of Thrones enthusiasm and sleep ( $t(48) = -1.78, p = .082$ ), but  
66 no correlation ( $r = -.01$ , 95% CI  $[-.29, .27]$ ) between age and sleep ( $t(48) = -0.07$ ,  
67  $p = .944$ ). Figures 1 and 2 depict these patterns.

68 To test the hypothesis that GoT enthusiasm varies as a function of R expertise and the  
69 extent to which respondents use tidy data, we carried out a one-way ANOVA. R experience  
70 ( $F(4, 40) = 2.72, \text{MSE} = 4.19, p = .043, \eta_p^2 = .214$ ) and the use of tidy data principles  
71 ( $F(1, 40) = 0.00, \text{MSE} = 4.19, p = .985, \eta_p^2 = .000$ ) did not predict enthusiasm for Game of  
72 Thrones. Table 3 summarizes these results.

## 73 Discussion

74 These results show how awesome it can be to use R, R Markdown, and literate  
75 programming principles to conduct and open, transparent, and reproducible psychological  
76 science. Yay, us!

77 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>
- 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>
- 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>
- 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., & 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 Game of Thrones  
enthusiasm by R experience.*

R_exp	Mean	Median	SD	Min	Max
none	4.80	4.50	2.66	1.00	9.00
limited	4.90	4.50	1.91	2.00	8.00
some	4.30	4.00	2.54	1.00	8.00
lots	2.70	3.00	1.42	1.00	5.00
pro	5.00	5.00	1.15	3.00	7.00

*Note.* This table was created with `apa_table()`

Table 2

*Correlation table of the example  
data set.*

	GoT	Age_yrs
GoT		
Age_yrs	-0.93***	
Sleep_hrs	-0.25	-0.01

*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	$p$	$\eta_p^2$
R exp	2.72	4	40	4.19	.043	.214
Tidy data	0.00	1	40	4.19	.985	.000
R exp $\times$ Tidy data	1.02	4	40	4.19	.410	.092

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

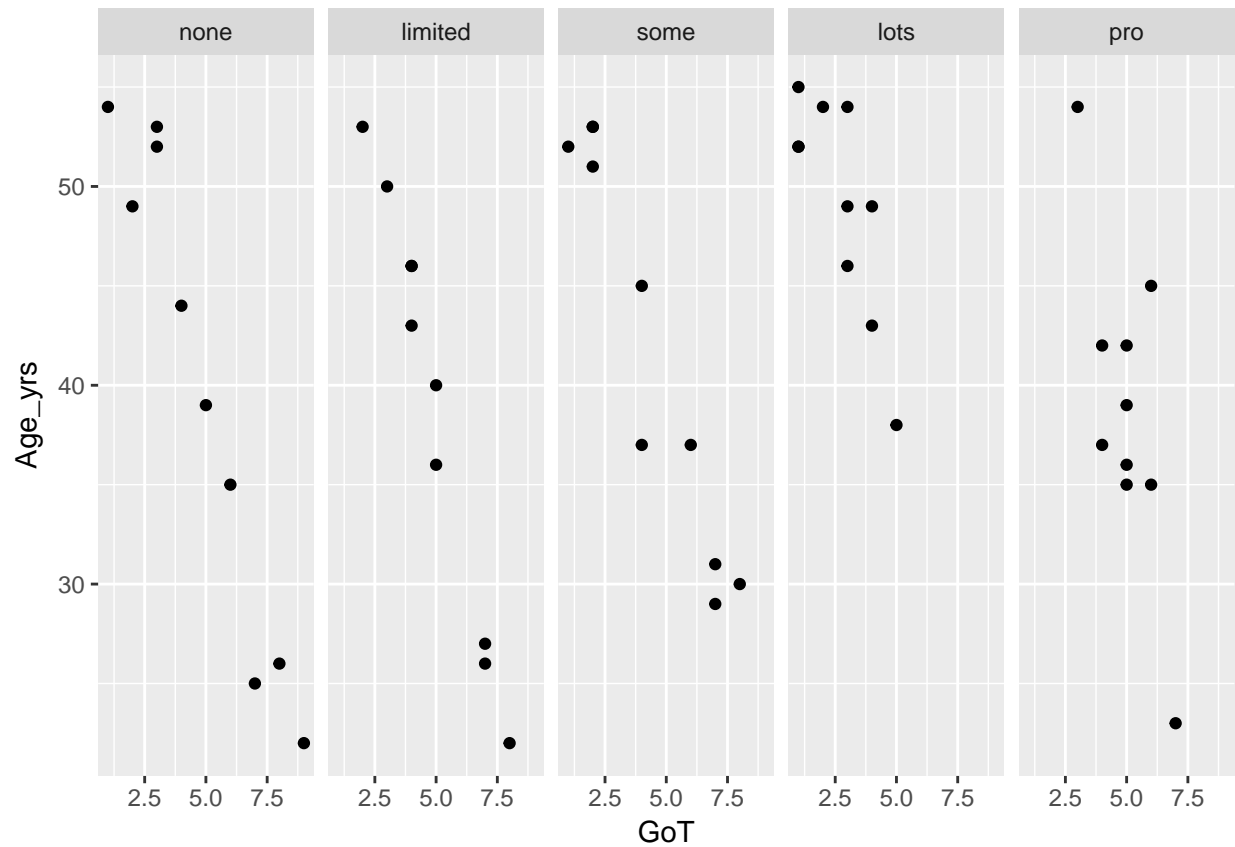
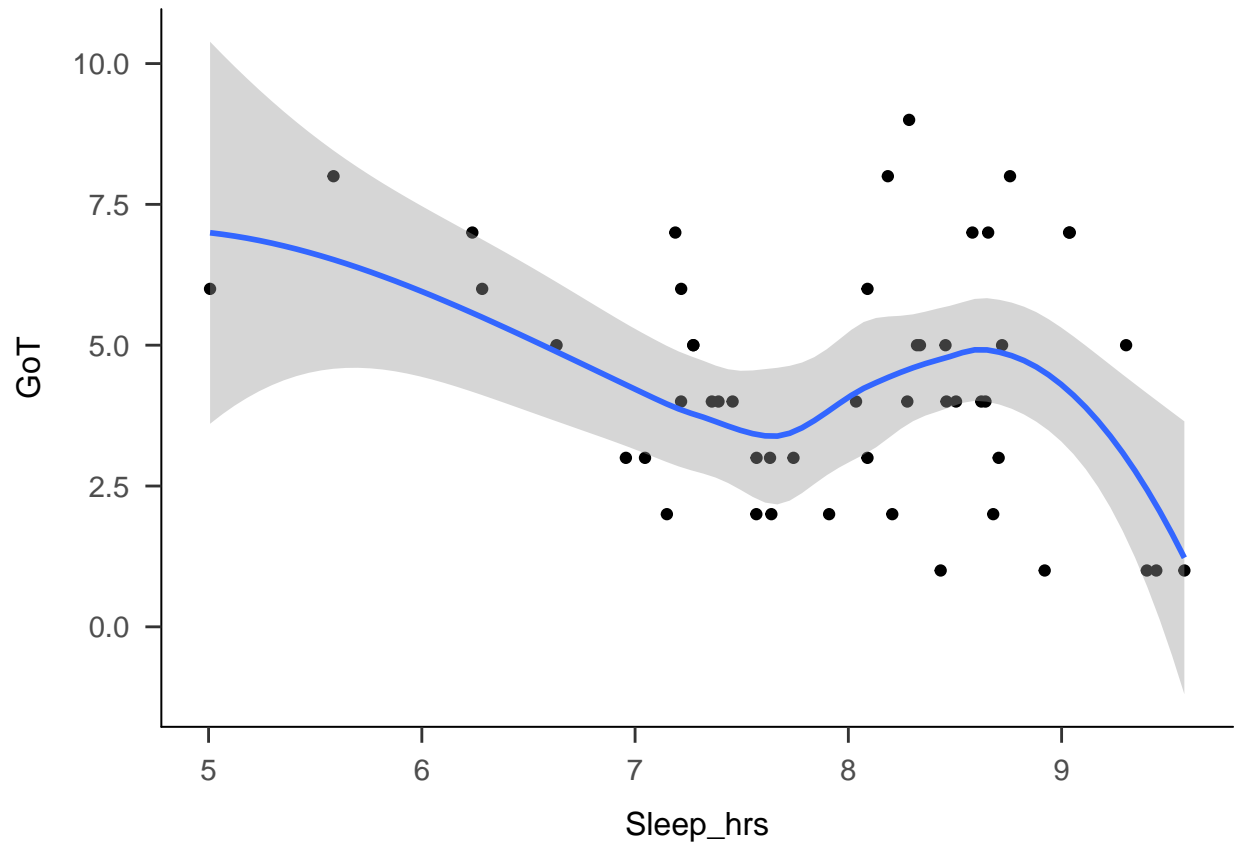


Figure 1. Game of Thrones enthusiasm by age and R experience



*Figure 2.* Game of Thrones enthusiasm by preferred hours of sleep