

R you ready for some data?

Rick O. Gilmore^{1,2}, James LeBreton¹, & Michael Hallquist¹

¹ The Pennsylvania State University

² 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

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 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 enthusiams 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.

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 [23-54] years.

Material

The survey can be found at this URL: 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, 2017) and the R-packages *dplyr* (0.5.0, Wickham & Francois, 2016), *ggplot2* (2.2.1, Wickham, 2009), *googlesheets* (0.2.2, Bryan & Zhao, 2017), *papaja* (0.1.0.9492, Aust & Barth, 2017), *purrr* (0.2.2.2, Henry & Wickham, 2017), *readr* (1.1.1, Wickham, Hester, & Francois, 2017), *tibble* (1.3.0, 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. There is a negative correlation ($r = -.92$, 95% CI $[-.95, -.86]$) between Game of Thrones enthusiasm and age ($t(48) = -16.19$, $p < .001$), a negative correlation ($r = -.48$, 95% CI $[-.67, -.23]$) between Game of Thrones enthusiasm and sleep ($t(48) = -3.79$, $p < .001$), but no correlation ($r = .22$, 95% CI $[-.06, .47]$) between age and sleep ($t(48) = 1.55$, $p = .128$). Figures 1 and 2 depict these patterns.

To test the hypothesis that GoT enthusiasm varies as a function of R expertise and the extent to which respondents use tidy data, we carried out a one-way ANOVA. R experience ($F(4, 40) = 0.12$, $MSE = 5.29$, $p = .974$, $\eta_p^2 = .012$) and the use of tidy data principles ($F(1, 40) = 0.17$, $MSE = 5.29$, $p = .686$, $\eta_p^2 = .004$) did not predict enthusiasm for Game of

66 Thrones. Table 2 summarizes these results.

67 **Discussion**

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

71 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>
- Bryan, J., & Zhao, J. (2017). *Googlesheets: Manage google spreadsheets from r*. Retrieved from <https://CRAN.R-project.org/package=googlesheets>
- Henry, L., & Wickham, H. (2017). *Purrr: Functional programming tools*. Retrieved from <https://CRAN.R-project.org/package=purrr>
- R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from <http://ggplot2.org>
- 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>

Table 1

*Descriptive statistics of Game of Thrones
enthusiasm by R experience.*

R_exp	Mean	Median	SD	Min	Max
none	4.60	5.00	1.51	2.00	6.00
limited	4.80	4.50	2.53	1.00	8.00
some	5.00	4.50	2.00	2.00	8.00
lots	5.30	5.00	2.87	1.00	10.00
pro	4.90	4.50	1.85	2.00	8.00

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

Table 2

ANOVA table for the analysis of the example data set.

Effect	F	df_1	df_2	MSE	p	η_p^2
R exp	0.12	4	40	5.29	.974	.012
Tidy data	0.17	1	40	5.29	.686	.004
R exp \times Tidy data	0.32	4	40	5.29	.865	.031

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

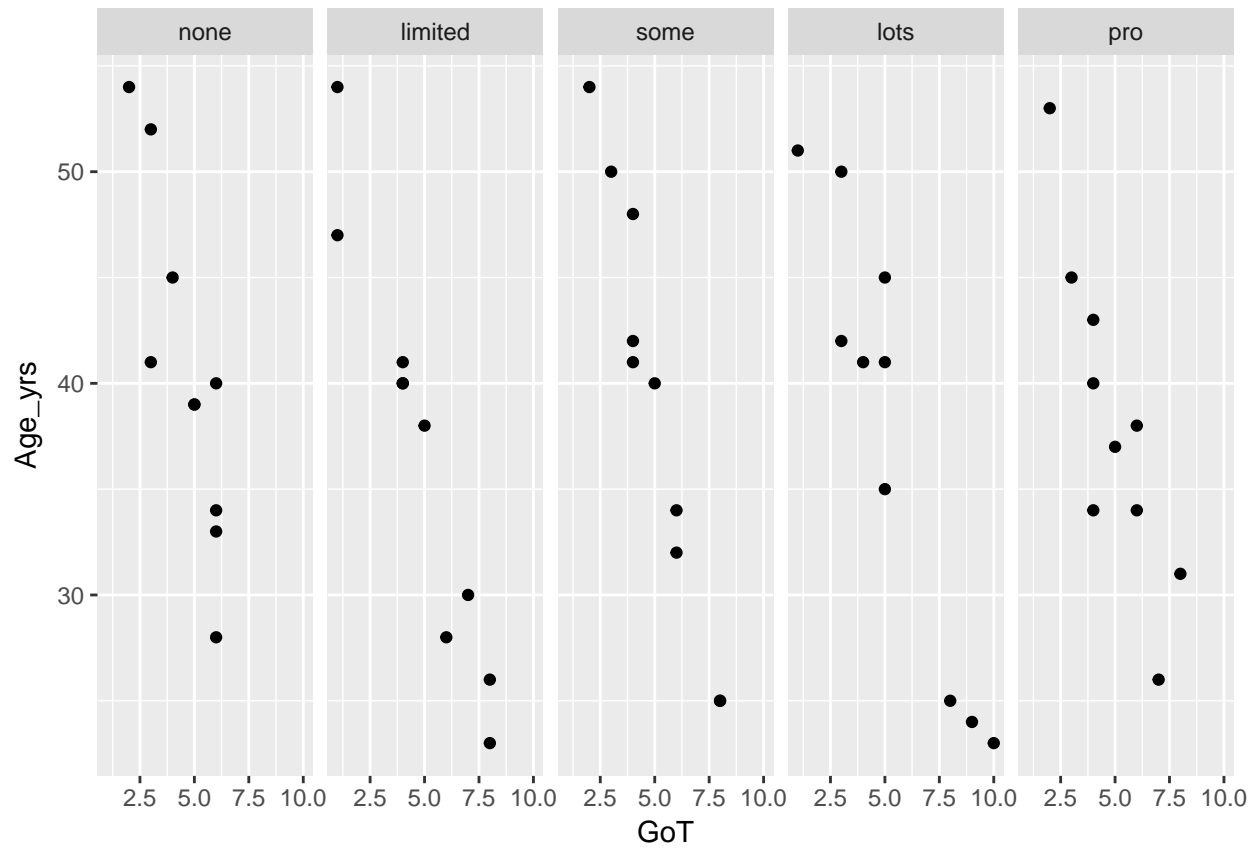


Figure 1. Game of Thrones enthusiasm by age and R experience

