1   R you ready for some data?

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

3   [1] The Pennsylvania State University

4   [2] Databrary.org

5   Author Note

Abstract

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

Keywords: APA, R Markdown

Word count: Not that many.

16   R you ready for some data?

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

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

19 survey of participants in the inaugural R Bootcamp at Penn State. We predicted that higher

20 levels of enthusiasm for "Game of Thrones" would be reported by respondents with *lower*

21 reported hours/day of preferred sleep, at least among younger respondents.

22   **Methods**

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

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

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

26 **Participants**

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

28 University Department of Psychology to complete an anonymous survey using a Google

29 Form. We asked participants to report their age in years. A total of 50 respondents

30 answered the survey with a reported age of [23-55] years.

31 **Material**

32   The survey can be found at this URL: https:

33 //docs.google.com/forms/d/1l5OX8PcN_lfVn3ykr_PtHCzhRbWzMbxhqtgILD45zRg/edit.

34 There were five questions asked:

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

36   2. Your enthusiasm for Game of Thrones [1..10 scale]

37   3. Age in years

38   4. Preferred number of hours spent sleeping/day

39   5. Favorite day of the week?

40   6. Are your data tidy?

41 **Procedure**

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

43 the survey on the web page containing the course schedule

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

45 to complete the survey after the first day's material.

46 **Data analysis**

47   We used R (3.4.1, R Core Team, 2017) and the R-packages *bindrcpp* (0.2, Müller,

48 2016), *dplyr* (0.7.2, Wickham & Francois, 2016), *Formula* (1.2.2, Zeileis & Croissant, 2010),

49 *ggplot2* (2.2.1, Wickham, 2009), *googlesheets* (0.2.2, Bryan & Zhao, 2017), *Hmisc* (4.0.3,

50 Harrell Jr, Charles Dupont, & others., 2017), *lattice* (0.20.35, Sarkar, 2008), *papaja*

51 (0.1.0.9492, Aust & Barth, 2017), *purrr* (0.2.3, Henry & Wickham, 2017), *readr* (1.1.1,

52 Wickham, Hester, & Francois, 2017), *survival* (2.41.3, Terry M. Therneau & Patricia M.

53 Grambsch, 2000), *tibble* (1.3.3, Wickham, Francois, & Müller, 2017), *tidyr* (0.6.3, Wickham,

54 2017a), and *tidyverse* (1.1.1, Wickham, 2017b) for all our analyses. The code used to

55 generate these analyses is embedded in this document. To view it, see the R Markdown file

56 in the GitHub repository associated with this paper.

57   **Results**

58   Table 1 summarizes the Game of Thrones ratings data by levels of R experience.

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

60 2, there is a negative correlation ($r = -.92$, 95% CI $[-.95, -.86]$) between Game of Thrones

61 enthusiasm and age ($t(48) = -16.24$, $p < .001$), a negative correlation ($r = -.24$, 95% CI

62 $[-.48, .04]$) between Game of Thrones enthusiasm and sleep ($t(48) = -1.69$, $p = .098$), but

63 no correlation ($r = -.02$, 95% CI $[-.30, .26]$) between age and sleep ($t(48) = -0.13$,

64 $p = .896$). Figures 1 and 2 depict these patterns.

⁶⁵   To test the hypothesis that GoT enthusiasm varies as a function of R expertise and the

⁶⁶ extent to which respondents use tidy data, we carried out a one-way ANOVA. R experience

⁶⁷ $(F(4, 40) = 4.92, \text{MSE} = 2.87, p = .003, \eta_p^2 = .330)$ and the use of tidy data principles

⁶⁸ $(F(1, 40) = 0.15, \text{MSE} = 2.87, p = .699, \eta_p^2 = .004)$ did not predict enthusiasm for Game of

⁶⁹ Thrones. Table 3 summarizes these results.

⁷⁰   **Discussion**

⁷¹   These results show how awesome it can be to use R, R Markdown, and literate

⁷² programming principles to conduct and open, transparent, and reproducible psychological

⁷³ science. Yay, us!

⁷⁴   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

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

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/

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.

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

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	5.20	5.50	2.53	1.00	9.00
limited	3.40	3.00	1.71	1.00	7.00
some	3.80	4.00	1.55	1.00	6.00
lots	6.10	6.00	0.74	5.00	7.00
pro	5.50	5.50	2.32	1.00	9.00

Note. This table was created with apa_table()

Table 2

Correlation table of the example

data set.

	GoT	Age_yrs
GoT		
Age_yrs	-0.92***	
Sleep_hrs	-0.24	-0.02

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	η_p^2
R exp	4.92	4	40	2.87	.003	.330
Tidy data	0.15	1	40	2.87	.699	.004
R exp × Tidy data	3.78	4	40	2.87	.011	.274

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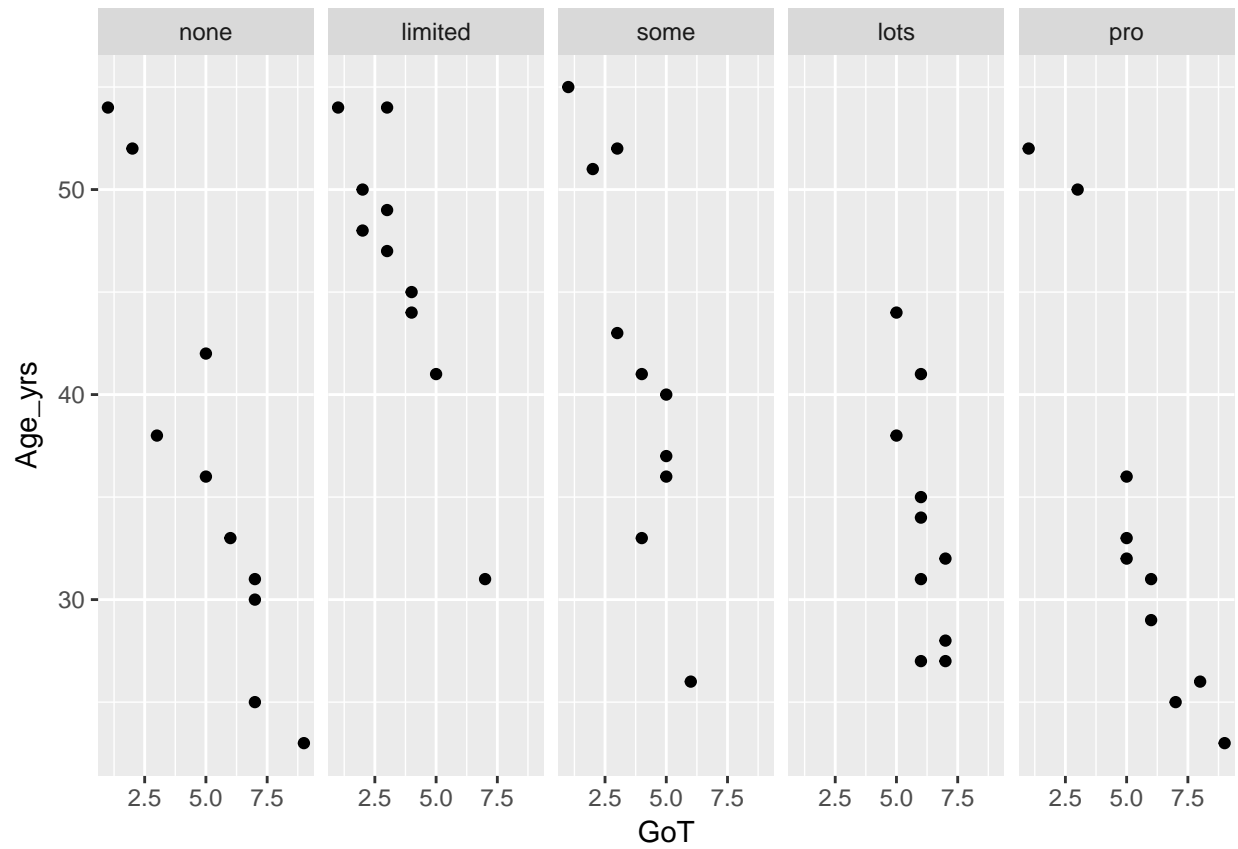
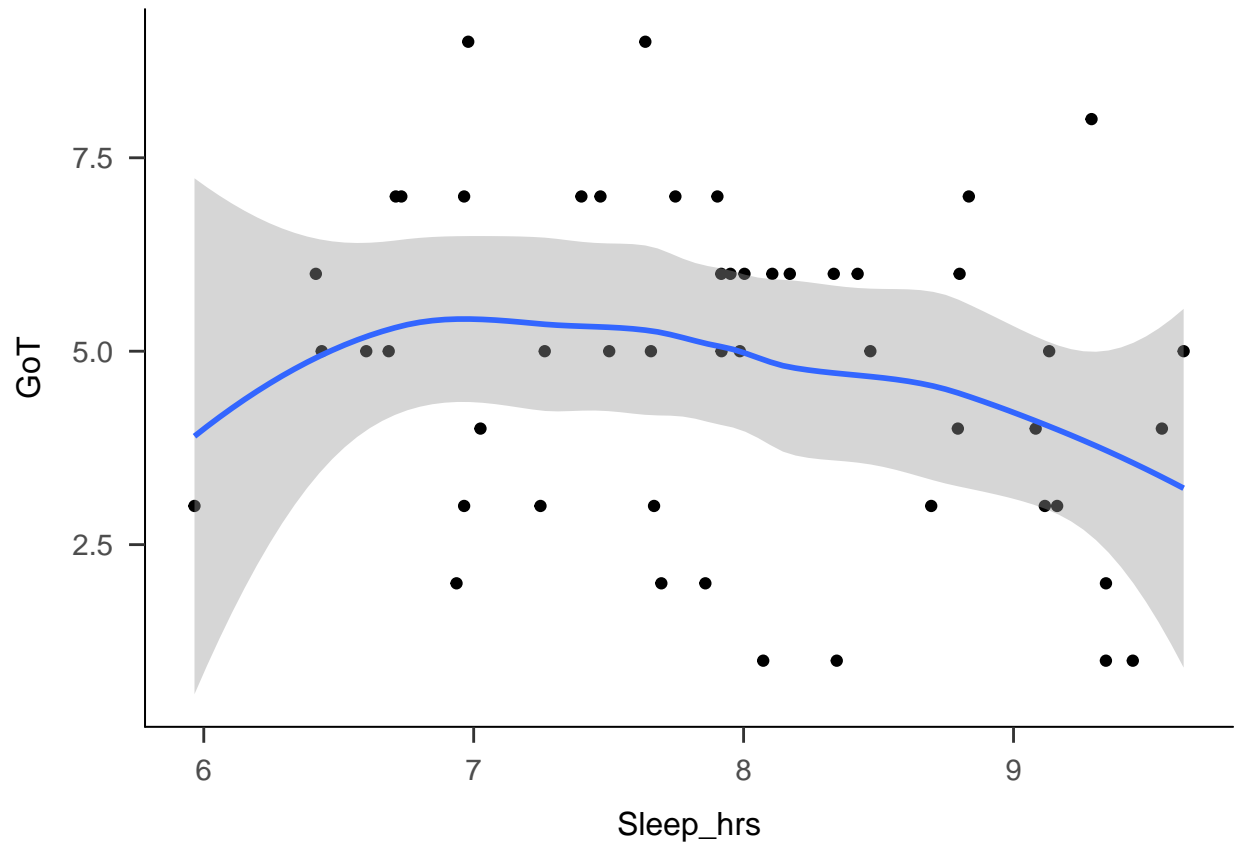*Figure 1*. Game of Thrones enthusiasm by age and R experience

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