

rtweet: Collecting and analyzing Twitter data

Michael W. Kearney¹

¹ School of Journalism, Informatics Institute, University of Missouri

DOI: [10.21105/joss.01829](https://doi.org/10.21105/joss.01829)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Submitted: 21 October 2019

Published: 24 October 2019

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

Statement of need

Following the [announced \(2016\) deprecation of the twitterR package](#) (Gentry, 2013), R users seeking to interact with Twitter APIs have been encouraged to use the `rtweet` package. Use of the up-to-date and actively-maintained `rtweet` package is especially important in light of changes to Twitter's APIs since 2016. Most notably is the increased character limit (from 140 to 280) for Twitter statuses (Rosen & Ihara, 2017). In addition to providing an updated interface with similar functionality to `twitter`, allowing R users to communicate with various endpoints from Twitter's REST API, the `rtweet` package also provides support for communicating with Twitter's stream API.

Summary

Interest in Twitter data continues to grow, but for many the task of actually collecting and analyzing data via Twitter APIs remains daunting. For example, in order to interact with Twitter's APIs users must, in addition to identifying and digesting the relevant information from [Twitter's developer documentation](#), build/send/receive requests, manage rate limits, and wrangle nested and real-time response objects into analysis-friendly data structures. Fortunately, the `rtweet` R package (Kearney, 2018a) is designed to simplify these processes, making interacting with Twitter's APIs more accessible to a wider range of users.

The main goals of the `rtweet` package are two-fold. The first goal is to make interacting with Twitter's APIs more approachable and streamlined for less computationally-experienced users. The second goal is to assist in the analysis of Twitter data via converting information returned by Twitter's APIs into tabular data structures and providing several convenience functions for common analytical techniques such as examining Twitter networks or the frequency of tweets over time. In short, although it is certainly possible for users to write their own Twitter API wrapper functions, the heavy-lifting done by `rtweet` to (a) streamline the building, authorizing, and sending of API requests, (b) wrangle deeply nested JSON data into tabular structures, and (c) provide convenience functions for relevant and popular analytical techniques, make it a valuable contribution in the area of collecting and analyzing Twitter data.

Although `rtweet` provides some coverage to user context-behaviors (e.g., posting statuses, liking tweets, following users, etc.), the primary audience for the package to date has been researchers. Accordingly, `rtweet` has been featured in numerous popular press (e.g., Bajak & Wu, 2019; Machlis, 2019; Riley, 2019) and academic publications (e.g., Bossetta, 2018; Bradley & James, 2019; Buscema, Ferilli, Massini, & Zavarrone, 2018; Erlandsen, 2018; Gitto & Mancuso, 2019; Kearney, 2018b, 2019; Li, Chamrajnagar, Fong, Rizik, & Fu, 2019; Lutkenhaus et al., 2019a, 2019b; Molyneux, Lewis, & Holton, 2018; Tsoi et al., 2018; Unsihuay, 2018; Valls et al., 2017; Wu & Ying, 2018).

References

- Bajak, A., & Wu, F. (2019). Democrats 'went low' on Twitter leading up to 2018. *Roll Call*. Retrieved from <https://www.rollcall.com/news/lead-midterms-twitter-republicans-went-high-democrats>
- Bossetta, M. (2018). A simulated cyberattack on Twitter: Assessing partisan vulnerability to spear phishing and disinformation ahead of the 2018 u.s. Midterm elections. *First Monday*, 23(12). doi:10.5210/fm.v23i12.9540
- Bradley, A., & James, R. J. (2019). How are major gambling brands using Twitter? *International Gambling Studies*, 1–20. doi:10.1080/14459795.2019.1606927
- Buscema, M., Ferilli, G., Massini, G., & Zavarrone, E. (2018). Media content analysis on online hate speech. *Positive Messengers*. Retrieved from https://positivemessengers.net/images/library/pdfs/Media_content_analysis_form_eng.pdf
- Erlandsen, M. (2018). Twitter as a tool of para-diplomacy: An exploratory cohort study based on Catalonia (2013-2017). *Revista Chilena de Relaciones Internacionales*, 2(1), 211–231. doi:10.5281/zenodo.1188429
- Gentry, J. (2013). *twitterR: R based Twitter client*. Comprehensive R Archive Network. Retrieved from <https://cran.r-project.org/package=twitterR>
- Gitto, S., & Mancuso, P. (2019). Brand perceptions of airports using social networks. *Journal of Air Transport Management*, 75, 153–163. doi:10.1016/j.jairtraman.2019.01.010
- Kearney, M. W. (2018a). *rtweet: Collecting Twitter data*. The Comprehensive R Archive Network. doi:10.5281/zenodo.2528481
- Kearney, M. W. (2018b). *Analyzing tweets about the 2016 US presidential "blunder" election*. (B. R. Warner, D. G. Bystrom, M. S. McKinney, & M. C. Banwart, Eds.). ABC-CLIO.
- Kearney, M. W. (2019). Analyzing change in network polarization. *New Media & Society*. doi:10.1177/1461444818822813
- Li, T. R., Chamrajnagar, A., Fong, X., Rizik, N., & Fu, F. (2019). Sentiment-based prediction of alternative cryptocurrency price fluctuations using gradient boosting tree model. *Frontiers in Physics*, 7, 98. doi:10.3389/fphy.2019.00098
- Lutkenhaus, R. O., Jansz, J., & Bouman, M. P. (2019a). Tailoring in the digital era: Stimulating dialogues on health topics in collaboration with social media influencers. *Digital Health*, 5, 1–11. doi:10.1177/2055207618821521
- Lutkenhaus, R. O., Jansz, J., & Bouman, M. P. (2019b). Mapping the Dutch vaccination debate on Twitter: Identifying communities, narratives, and interactions. *Vaccine: X*, 100019. doi:10.1016/j.jvacx.2019.100019
- Machlis, S. (2019). R community blasts DataCamp response to exec's 'inappropriate behavior'. *ComputerWorld*. Retrieved from https://www.computerworld.com/article/3389684/r-community-blasts-datacamp-response-to-execs-inappropriate-behavior.html#tk.rss_news
- Molyneux, L., Lewis, S. C., & Holton, A. E. (2018). Media work, identity, and the motivations that shape branding practices among journalists: An explanatory framework. *New Media & Society*, 1–20. doi:10.1177/1461444818809392
- Riley, C. (2019). Does Twitter make political polarization seem worse? *Futurity*. Retrieved from <https://www.futurity.org/political-polarization-twitter-moderates-2025862/>
- Rosen, A., & Ihara, I. (2017). Giving you more characters to express yourself. Retrieved from https://blog.twitter.com/official/en_us/topics/product/2017/Giving-you-more-characters-to-express-y.html

Tsoi, K. K., Chan, N. B., Chan, F. C., Zhang, L., Lee, A. C., & Meng, H. M. (2018). How can we better use Twitter to find a person who got lost due to dementia? *npj Digital Medicine*, *1*(1), 14. doi:[10.1038/s41746-018-0017-5](https://doi.org/10.1038/s41746-018-0017-5)

Unsihuay, J. E. G. (2018). Topic modeling en datos de Twitter: Una aplicación en el contexto político peruano. *XXVIII Simposio Internacional de Estadística*.

Valls, F., Redondo, E., Fonseca, D., Torres-Kompen, R., Villagrasa, S., & Martí, N. (2017). Urban data and urban design: A data mining approach to architecture education. *Telematics and Informatics*. doi:[10.1016/j.tele.2017.09.015](https://doi.org/10.1016/j.tele.2017.09.015)

Wu, H., & Ying, S. (2018). Finding similar users over multiple attributes on the basis of intuitionistic fuzzy set. *Mobile Networks and Applications*, 1–9. doi:[10.1007/s11036-018-1055-6](https://doi.org/10.1007/s11036-018-1055-6)