The Journal of Open Source Software

studentlife: Tidy Handling and Navigation of a Valuable Mobile-Health Dataset

Daniel Fryer[1, 2], **Hien Nguyen**[1], **and Pierre Orban**[3]

1 Department of Mathematics and Statistics, La Trobe University, Bundoora 3086, Victoria Australia **2** School of Mathematics and Physics, University of Queensland, St. Lucia 4072, Queensland Australia **3** Department of Psychiatry, University of Montreal, Montreal H3C 3J7, Quebec Canada

DOI: 10.21105/joss.01587

Software

- Review
- Repository
- Archive

Submitted: 06 July 2019
Published: 21 August 2019

Summary

Smartphones have become ubiquitous, and nearly constantly accompany their owners as they go about their daily lives. This sensor-rich and computationally-efficient technology allows quantifying continuous streams of human behaviors and activities unobtrusively and *in situ*, a process referred to as digital phenotyping (Onnela & Rauch, 2016). Two main types of data collected through this approach can be distinguished: on the one hand, active data that involve the voluntary participation of users, as required for surveys; and on the other hand, passive data that do not involve any dedicated action from users, as is the case for data generated by various sensors embedded in the smartphones.

The R package (R Core Team, 2019) `studentlife` is designed to assist in downloading, navigating and analysing the **StudentLife** dataset. The **StudentLife** dataset (Wang et al., 2014) is a one-of-its-kind digital phenotyping dataset made publicly available (https://studentlife.cs.dartmouth.edu) by Wang, Campbell and colleagues, at Dartmouth College (New Hampshire, United States). It consists of a rich array of data, collected through a dedicated mobile app, over a 10-week term in a sample of 48 college students. Active data were obtained through ecological momentary assessments, corresponding to self-reports of the students' states, prompted multiple times per day (3 to 13 daily assessments) at pseudorandom intervals. Likert scale-based questions (e.g., regarding stress, exercise, social activity) that changed across assessments, were systematically preceded by a photographic affect meter that captured the students' instantaneous mood in an original fashion, hence promoting the collection of a large amount of affect data. Passive data were also continuously collected through smartphone sensors (e.g., accelerometer, microphone, GPS). A key valuable aspect of the **StudentLife** dataset is that it does not merely share raw sensor data. Indeed, a number of behavioral inferences based on trained machine learners were automatically computed by the mobile app and are also included in the dataset, which includes, for example, type of physical activity, audio/conversation state, and sleep duration. Finally, the dataset contains data from pre- and post-term surveys aimed at assessing mental health with a series of well-validated questionnaires and scales (e.g., PHQ-9 to screen for the presence and severity of depression), as well as a number of measures related to academic context and performance.

While the primary focus of the **StudentLife** project is on mental well-being in educational settings, the publicly shared dataset could serve a more general purpose in addressing novel questions on the temporal dynamics of mental health at large. Digital phenotyping is promised to revolutionize clinical research and healthcare, yet research on mental health using smartphone sensing is only nascent and studies that have employed ecological momentary assessment in clinical samples typically follow subjects for much shorter durations than in the **StudentLife** project. Furthermore, this dataset is well-suited for developing methods tailored to analyzing the high-dimensional and intensively longitudinal data that arise from digital phenotyping.

The `studentlife` package provides functionality to assist users to:

- Download either the full dataset or a smaller sub-sample dataset.
- Organise the data intuitively into tables and schemas (i.e., collections of similar tables).
- Use a convenient interactive menu for browsing schemas and importing tables or subsets thereof.
- Classify tables based on their time information: timestamped, interval, date-only or dateless.
- Aggregate observations within blocks of time. For example, within each hour, day or week.
- Perform exploratory data analysis, such as visualisation of missing values and response frequencies.

For an introduction to using this package, view the README file on GitHub. The latest stable build can be found on CRAN. Please give your feedback and report bugs at the GitHub issues page.

References

Onnela, J.-P., & Rauch, S. L. (2016). Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health. *Neuropsychopharmacology*, *41*(7), 1691. doi:10.1038/npp.2016.7

R Core Team. (2019). R: a Language and Environment for Statistical Computing. Retrieved from https://www.R-project.org/

Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., et al. (2014). StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students Using Smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 3–14). ACM. doi:10.1145/2632048.2632054