Reproduction Materials for "Collective Emotions and Social Resilience in the Digital Traces After a Terrorist Attack"

David Garcia and Bernard Rimé, 10 October 2022

Following the recommendations of Crüwell et al., 2022, this document explains an updated version of the replication materials for the article "Collective Emotions and Social Resilience in the Digital Traces After a Terrorist Attack" (Garcia & Rimé, 2018). The data and codes have been prepared to reproduce the results on the main text of the article. The new code and datasets with more accessible file formats can be found on the following github repository: https://github.com/dgarcia-eu/ParisAttacks

Please address any questions or comments to david.garcia@uni-konstanz.de

New Data files

Main tweets file: Tweets.csv.gz and Tweets.RData

Size: 17,899,591 rows, 12 columns
This file contains the essential content data for each tweet in the study.
Columns:

- userid: Integer. Anonymized user id of the author of the tweet.
- date: String. Date when the tweet was produced in YYYY-MM-DD format.
- i: Integer. Count of words of the first person singular class in LIWC.
- soc: Integer. Count of words of the social processes class in LIWC.
- posemo: Integer. Count of words of the PA class in LIWC.
- negemo: Integer. Count of words of the NA class in LIWC.
- anx: Integer. Count of words of the anxiety class in LIWC.
- ang: Integer. Count of words of the anger class in LIWC.
- sad: Integer. Count of words of the sadness class in LIWC.
- prosoc: Integer. Count of words of the prosocial terms dictionary.
- frenchValues: Integer. Count of words matching the French Values terms.
- n: Integer. Total amount of words in the tweet.

This file contains tweets produced by 61,677 unique users. The datasets in the article were produced after several steps of data filtering, starting with a sample of tweets by hashtag, which was used to identify users, which then were filtered according to location and other criteria, and from which twitter timelines were gathered. The tweets of these timelines were again filtered to remove retweets and tweets not in French and to focus on the periods of our analysis. This process took several months and only tweets available at the retrieval date were considered in the analysis. The total user count reported in the article (62,114) corresponds to the users for which a timeline was requested. However, after filtering tweets in the timeline and updating datasets to comply with the Twitter Developer Agreement, some of those users might contribute zero tweets to the final dataset, thus the small discrepancy (less than 0.1%) between user counts in the reporting of the paper and what can be seen in the datasets for replication. In complex data analyses like this one, there is not one single sample size that matters but a large and interdependent set of steps and filters with different resulting sizes. We documented this in the Supplemental Materials of the article but this should be reported in more detail, including flow charts and descriptive statistics in various steps.

Attacks references: TweetsAttacksData.csv

Size: 1,086,089 rows, 2 columns This file contains basic information on original tweets that included one of the terms in reference to the attacks, as explained in the Supplementary Materials.
Columns:

- userid: Integer. Anonymized user id of the author of the tweet.
- date: String. Date when the tweet was produced in YYYY-MM-DD format.

Tweet pairs: TweetPairs.csv.gz and TweetPairsDF.RData

Size: 890,994 rows, 11 columns
This file contains time-ordered tweet pairs by the same user in a period after the attacks.
Columns:

- userid: Integer. Anonymized user id of the author of the tweet.
- soc: Integer. Count of words of the social processes class in LIWC.
- prosoc: Integer. Count of words of the prosocial dictionary.
- frenchValues: Integer. Count of words of the French values dictionary.
- posemo: Integer. Count of words of the PA class in LIWC.
- negemo: Integer. Count of words of the NA class in LIWC.
- presoc: Integer. Count of social process words in previous tweet.
- preprosoc: Integer. Count of words of prosocial class in previous tweet.
- prefrenchValues: Integer. Count of words from the French values dictionary in prev tweet.
- preposemo: Integer. Count of PA words in previous tweet.
- prenegemo: Integer. Count of NA words in previous tweet.

Running the code

The main file to reproduce results is Paris_Attacks.Rmd. It is a markdown file with R code that generates a pdf report with the results of the main text of the article and additional statistics and descriptive information. Functions used several times can be found under Scripts/AuxFunctions.R. It loads data contained in the Data/ folder.

Some chunks of Paris_Attacks.Rmd are very computationally intensive and require large memory. These chunks are set not to be run automatically "eval=FALSE" and the results of our execution of those chunks are saved in the temp/ folder. This way you can inspect the results and statistical analyses without having to run time-demanding code for bootstrapping and other computational tasks.

The current version and the pdf output was produced with R version 3.6.3 (the original paper is a few years old). Attached package versions in sessionInfo() are the following: texreg_1.36.23 arm_1.12-2 lme4_1.1-23 Matrix_1.2-18 MASS_7.3-51.5 magrittr_2.0.1 dplyr_1.0.10 zoo_1.8-8 ggplot2_3.3.2 sfsmisc_1.1-13

References

Crüwell, S., Apthorp, D., Baker, B. J., Colling, L. J., Elson, M., Geiger, S. J., . . . Brown, N. J. L. (2022). What's in a Badge? A Computational Reproducibility Investigation of the Open Data Badge Policy in one Issue of Psychological Science. https://doi.org/10.31234/osf.io/729qt

Garcia, D., & Rimé, B. (2019). Collective emotions and social resilience in the digital traces after a terrorist attack. Psychological science, 30(4), 617-628.