TITLE: Social Media Captures Demographic and Regional Physical Activity

Nina Cesare1,2, Quynh C. Nguyen3, Christan Grant4, Elaine O. Nsoesie1,2

**AFFILIATIONS:**

1Department of Global Health, School of Public Health, Boston University, Boston, Massachusetts, 02119.

2 Institute for Health Metrics and Evaluation, University of Washington, Seattle, Washington, 98121.

3Department of Epidemiology and Biostatistics, School of Public Health, University of Maryland, College Park, Maryland, 20742.

4Department of Computer Science, University of Oklahoma, Norman, Oklahoma, USA, 73019.

CORRESPONDING AUTHOR:

Nina Cesare, Department of Global Health, Boston University, 801 Massachusetts Avenue, Suite 300, Boston, MA, 02118, United States

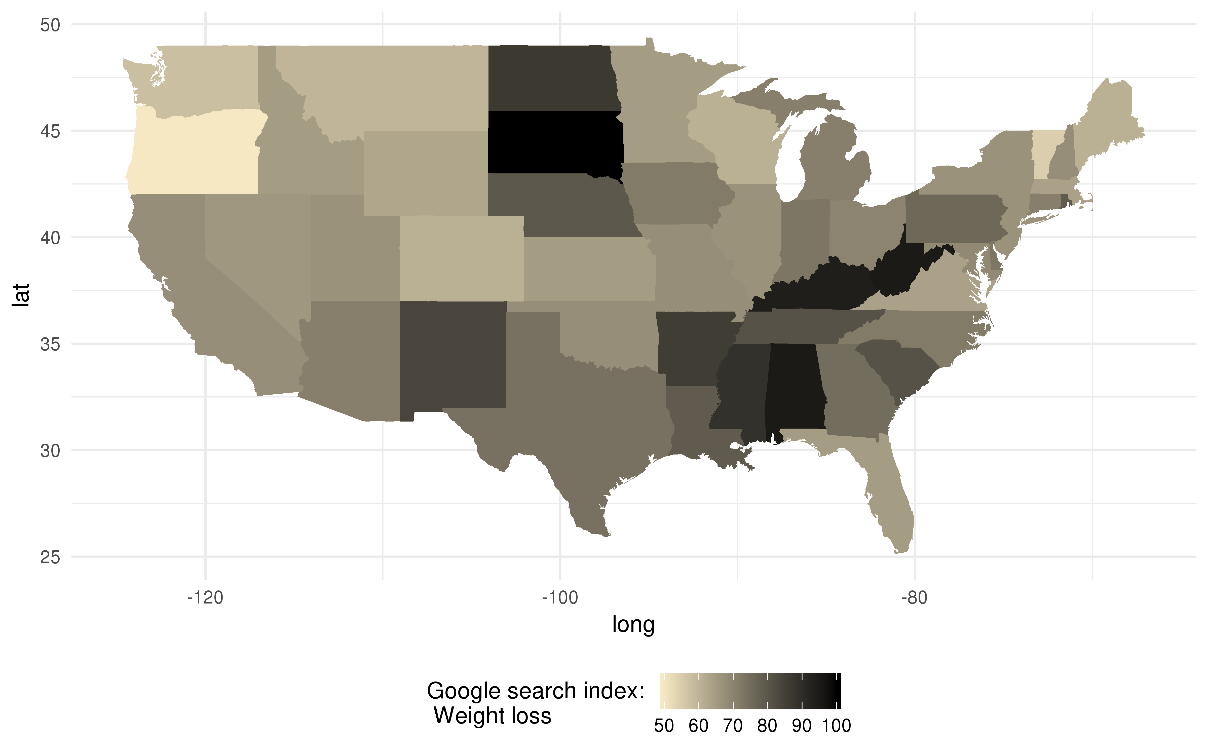
Tel: 617-358-2415, Email: ncesare@bu.edu



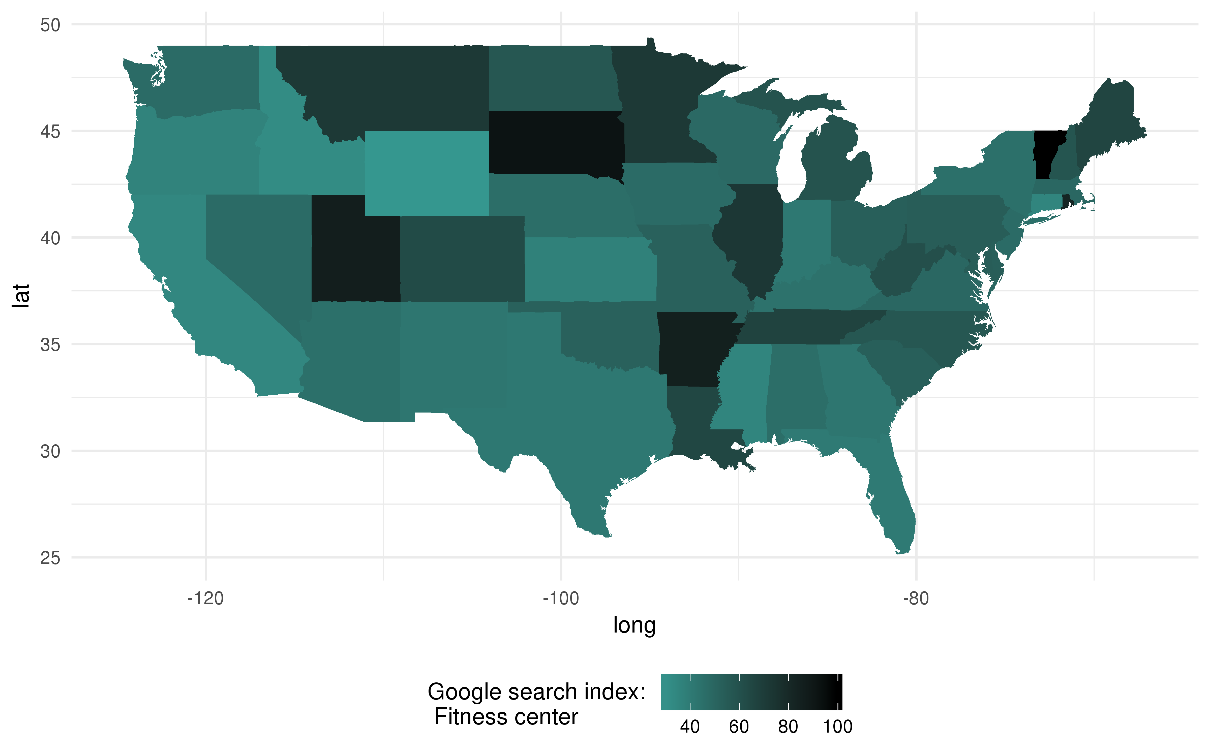
**Table S1.** Fitness keywords used to collect data from Twitter’s streaming API. Note that some keywords (e.g. ‘run late’) permitted keyword-based filtering of irrelevant content, as described in the *Methods: Data processing* section of the manuscript.

|  |
| --- |
| "I love Spectrum Fitness more every time I go. Indoor pool: winning. Planet Fitness: lacking" |
| "Out with my cousin kayaking today @ smugglers cove https://t.co/3erzvy3sum" |
| "I love swimming in the ocean!!" |
| "Just so everyone knows Dan bowled a 38 last night??? @greekkinder @vanmoutsos" |
| "Dancing was her favorite part of today’s party. @ wdiv https://t.co/dol0b5coc5" |
| "#whoareyou #namaste @ the Family Yoga Center Inc. https://t.co/1ft6rwvmya" |
| "Disfrutando el fin muy relax... #feelgood #resting @ Jazzercise Fitness in West Jordan https://t.co/wbdmzzzylk" |
| "Mixing it up with an early morning workout = photo safari #denver livingsocial @ Denver Capitol Building https://t.co/yvtaqqsuiv" |
| "#spring #springbreak sailing camp on #lakeaustin #austin #atx for #kids just days away! https://t.co/oaed967cd6" |
| "#arcadefun #grandparentslove #margaret @ putt putt golf & games https://t.co/j3rl7g6g5g" |

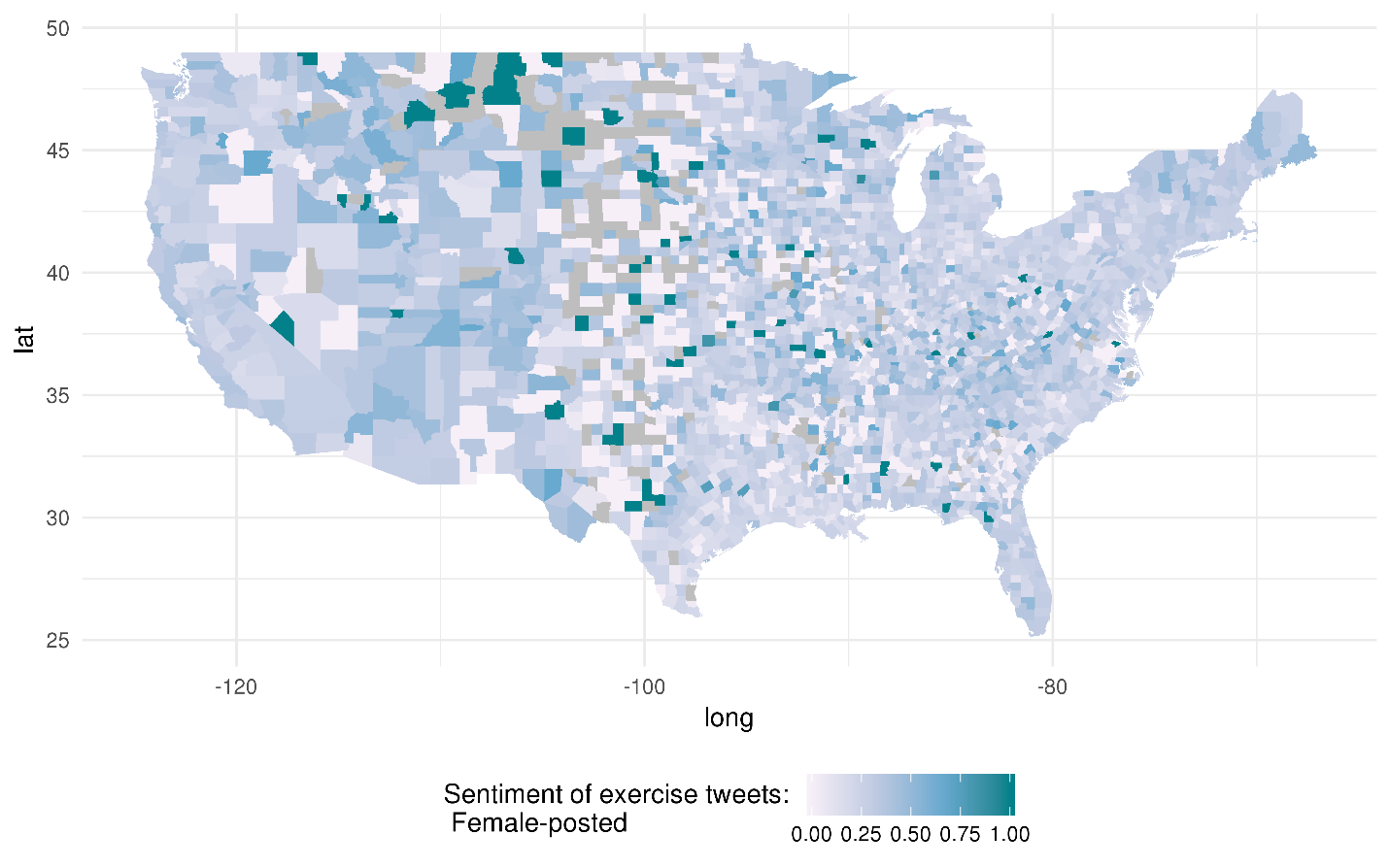
**Table S2.** A sample of ten tweets from the dataset. As indicated in these tweets, users may post updates in the midst of their activity, or discuss recent or planned activity.



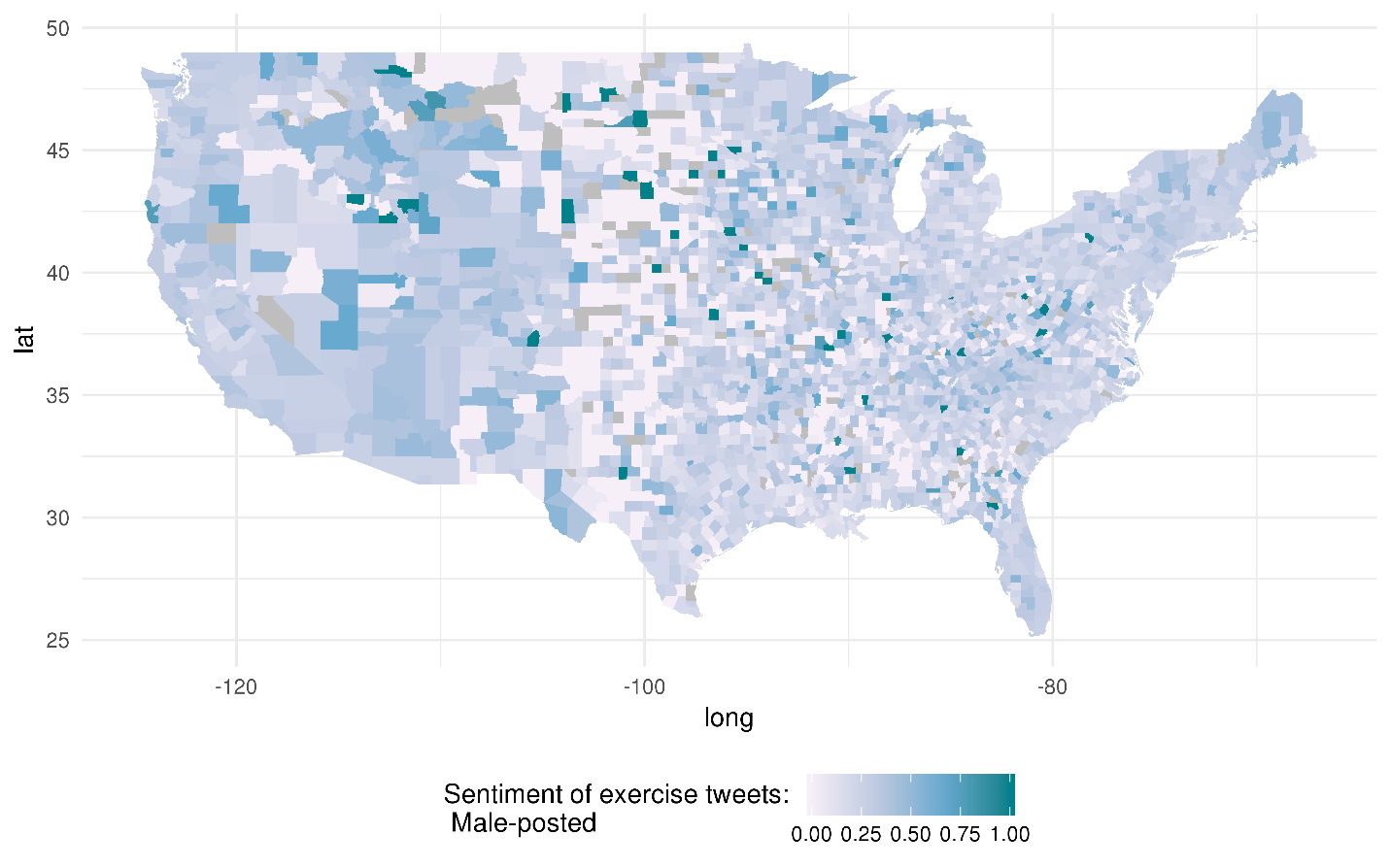
**Fig. S1.** Google Trends search index distribution for topic “fitness center”



**Fig. S2.** Google Trends search index distribution for topic “weight loss”

****

**Fig. S3.** Sentiment toward exercise measured as the proportion of tweets classified as ‘happy’ sentiment score of 0.80 or higher within each county for women.

****

**Fig. S4.** Sentiment toward exercise measured as the proportion of tweets classified as ‘happy’ sentiment score of 0.80 or higher within each county for and men.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table S3.** Pearson correlations between Twitter variables and physical inactivity. Negative correlations are higher between overall tweet volume and inactivity prevalence for both males and females. | | | | | |
|  |  | | Percent inactive | | |
|  |  | | Female | Male | |
| Percent tweets exercise-related (logged) | Midwest | | -0.132 | -0.047 | |
|  | Northeast | | -0.373 | -0.234 | |
|  | South | | -0.174 | -0.203 | |
|  | West | | -0.267 | -0.217 | |
|  | National | | -0.263 | -0.251 | |
| Tweet sentiment | Midwest | | -0.077 | -0.063 | |
|  | Northeast | | -0.271 | -0.063 | |
|  | South | | -0.051 | -0.027 | |
|  | West | | -0.076 | -0.194 | |
|  | National | | -0.116 | -0.115 | |
| Exercise intensity | Midwest | | 0.001 | -0.123 | |
|  | Northeast | | -0.048 | 0.060 | |
|  | South | | 0.063 | 0.006 | |
|  | West | | -0.004 | -0.203 | |
|  | National | | -0.001 | -0.061 | |
| **Table S4.** Mixed-effects regression for 2015 county-level, sex-specific inactivity prevalence across the United States | | | | | | |
|  | | Female | | | Male | |
| Percent of tweets exercise-related logged | | -0.425\*\*\* | | | -0.476\*\*\* | |
|  | | 0.078 | | | 0.075 | |
| Tweet sentiment toward exercise | | -0.008\*\*\* | | | -0.004 | |
|  | | 0.003 | | | 0.003 | |
| Avg. exercise intensity, via Tweets | | 0.001 | | | -0.002 | |
|  | | 0.001 | | | 0.001 | |
| 'Fitness center' Google search index | | -0.023 | | | -0.003 | |
|  | | 0.028 | | | 0.020 | |
| 'Weight loss' Google search index | | 0.210\*\*\* | | | 0.138\*\*\* | |
|  | | 0.042 | | | 0.030 | |
| Median age | | -0.054 | | | -0.027\*\* | |
|  | | 0.011 | | | 0.011 | |
| Median household income (in 1000s) | | -0.090\*\*\* | | | -0.085\*\*\* | |
|  | | 0.005 | | | 0.005 | |
| Percent non-Hispanic black | | 0.031\*\*\* | | | -0.007 | |
|  | | 0.006 | | | 0.006 | |
| Percent Hispanic | | -0.006 | | | -0.003 | |
|  | | 0.005 | | | 0.005 | |
| Percent with access to exercise space | | -0.027\*\*\* | | | -0.030\*\*\* | |
|  | | 0.003 | | | 0.003 | |
| Driving death rate | | 0.848\*\*\* | | | 0.904\*\*\* | |
|  | | 0.099 | | | 0.099 | |
| Violent crime rate | | 0.001\*\* | | | 0.00002 | |
|  | | 0.0003 | | | 0.0003 | |
| Constant | | 24.393\*\*\* | | | 26.878\*\*\* | |
|  | | 3.447 | | | 2.562 | |
| Observations | | 2,768 | | | 2,805 | |
| Adjusted R-Squared | | 0.793 | | | 0.674 | |
| Std. deviation, random intercept | | 3.143 | | | 2.216 | |
| Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01 | |  | | |  | |
|  | |  | | |  | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table S5.** Mixed-effects regression for county-level, female-specific inactivity by region - Twitter variables only | | | | |
|  | Midwest | Northeast | South | West |
| Percent of tweets exercise-related (logged) | -0.223\* | -1.345\*\*\* | -0.905\*\*\* | -1.185\*\*\* |
|  | -0.134 | -0.445 | -0.132 | -0.223 |
| Sentiment toward exercise | -0.006 | -0.02 | -0.009\* | -0.01 |
|  | -0.004 | -0.023 | -0.005 | -0.008 |
| Avg. calories expended | 0.0003 | 0.014\* | 0.002 | -0.0002 |
|  | -0.002 | -0.008 | -0.002 | -0.003 |
| Constant | 23.952\*\*\* | 19.314\*\*\* | 28.523\*\*\* | 19.949\*\*\* |
|  | -0.816 | -1.871 | -0.978 | -0.97 |
| Observations | 961 | 217 | 1,367 | 385 |
| Adjusted R-Squared | 0.487 | 0.390 | 0.600 | 0.427 |
| Std. deviation, random intercept | 2.546 | 1.895 | 3.606 | 2.341 |
| Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01 |  |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table S6:** Mixed-effects regression for county-level, male-specific inactivity by region - Twitter variables only | | | | |
|  | Midwest | Northeast | South | West |
| Percent of tweets exercise-related (logged) | -0.177 | -1.149\*\*\* | -0.954\*\*\* | -1.135\*\*\* |
|  | 0.132 | 0.426 | 0.114 | 0.226 |
| Sentiment toward exercise | -0.001 | -0.019 | -0.005 | -0.030\*\*\* |
|  | 0.005 | 0.018 | 0.005 | 0.009 |
| Avg. calories expended | -0.006\*\*\* | 0.01 | -0.001 | -0.006\* |
|  | 0.002 | 0.007 | 0.002 | 0.004 |
| Constant | 25.385\*\*\* | 18.492\*\*\* | 24.757\*\*\* | 21.711\*\*\* |
|  | 0.746 | 1.51 | 0.829 | 1.056 |
| Observations | 994 | 216 | 1,384 | 396 |
| Adjusted R-Squared | 0.387 | 0.152 | 0.545 | 0.439 |
| Std. deviation, random intercept | 2.228 | 0.9078 | 3.013 | 2.452 |
| Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01 |  |  |  |  |