**DATA DICTIONARY**

Please refer to the following manuscript for more details on the data collection:

Nguyen QC, Li D, Meng H-W, Kath S, Nsoesie EO, Li F, et al. Building a National Neighborhood Dataset From Geotagged Twitter Data for Indicators of Happiness, Diet, and Physical Activity. *JMIR Public Health Surveillance* 2016;2(2):e158.

 Twitter data collection period: April 2015-March 2016.

**Food variables:**

propfood: proportion of tweets about food

prophealthy: proportion of tweets about healthy food

propfastfood: proportion of tweets about fast food

senthealthy: proportion of tweets about healthy foods that are happy

sentfastfood: proportion of tweets about fast food that are happy

sentfood: proportion of food tweets that are happy

calories: average caloric density (per 100 grams) of food tweets

**Exercise variables:**

propexercise: proportion of tweets about exercise

sentex: proportion of exercise tweets that are happy

**Happiness variables:**

prophappy: proportion of tweets that are happy

**Alcohol variables:**

propalc: proportion of tweets about alcohol   
sentalc: proportion of tweets about alcohol that are happy