Each row in the data is a tweet. The fields are described below.

Field	**Description**	**How Many Tweeters Have It?**
tweet_text	The text of the tweet	Everyone (barring weird stuff)
lat	Tweeter geolocated latitude. Available for only a small fraction of the tweeters	1-10%
lon	Tweeter geolocated longitude	1-10%
tweeter_location	Tweeter's self-described location	70%
tweet_time_gmt	Time at which tweet was tweeted (GMT)	Everyone (barring weird stuff)
tweet_time_local	Time at which tweet was tweeted (tweeter's local timezone)	70%
name	Tweeter's name (not their screen name)	Everyone (barring weird stuff)
screen_name	Tweeter's screen name	Everyone (barring weird stuff)
description	Tweeter's self-description	Everyone (barring weird stuff)
n_friends	Number of friends Tweeter has	Everyone (barring weird stuff)
n_followers	Number of followers Tweeter has	Everyone (barring weird stuff)
n_statuses	Number of statuses Tweeter has posted	Everyone (barring weird stuff)
tweeter_gender	The tweeter's gender (inferred from their name)	50-60%
hashtags	Hashtags in the tweet	Everyone (barring weird stuff)
tweet_is_retweet	True if the tweet is a retweet; false otherwise	Everyone (barring weird stuff)

original_screen_name, original_name, original_gender, original_followers	If the tweet is a retweet, these fields describe the original tweeter	Everyone (barring weird stuff) for whom it's a retweet
tweeted_at_screennames, tweeted_at_names, tweeted_at_genders	If the tweeter tweeted at someone (ie, @emma) these fields describe who they tweeted at	Everyone (barring weird stuff)
favorites	How many times tweet has been favorited (keep in mind this field is computed *as the tweet is streaming*). On	Everyone (barring weird stuff)
retweets	How many times tweet has been retweeted	Everyone (barring weird stuff)
tweeter_location_from_gps	Tweeter's address, inferred from latitude and longitude	No one, at present, since it takes too long -- the best way I've found to do reverse geocoding is the Google API, which gives great data but is rate-limited at 2,500 requests a day. If someone really needs this data I can probably figure out a way to implement it.
All other fields	How many words fit the LIWC categories used for sentiment analysis. For category descriptions, see here.	Everyone (barring weird stuff)

Warnings:
1. Files are large. At present, about half a megabyte per thousand tweets.
2. Sentiment analysis is sketchy. Examine results to make sure they're actually intuitive.
3. Some data is available only for a subset of the tweeters and this may introduce biases.
4. All fields are computed *as the tweet is streaming*. Eg, if I tweet something, and then later tons of people favorite it, its value of favorites will still be zero because when it was originally tweeted, no one had favorited it.