Happiness in the U.S.: Measuring the Sentiment of Geolocated Tweets

João Ferreira Loff
joao.loff@tecnico.ulisboa.pt

Universidade de Lisboa - Instituto Superior Técnico

Abstract. Individual happiness is a fundamental societal metric. Normally measured through self-report or survey-like schemes, happiness has often been indirectly characterized and overshadowed by more readily quantifiable economic indicators. In this paper we analyze data from the micro-blogging site Twitter and generate a happiness map and ranking for all Continental U.S. states and counties. We do so by combining (1) a massive, geotagged data set comprising over 233 million tweets in 2012 and (2) comparing those results with the annually happiness index made by Gallup-Healthway. In measuring happiness we use the aproched introduced by Dodds and Danforth in 2009 [6], and greatly expanded upon by [7, 14]. I found out that there is a good enough correlation between my approach and Gallup-Healthway survey. Most important (and never hear of) conclusion I reached was that there is a very strong correlation between the number of tweets from a state and its happiness ranking, unveiling that Twitter usage is a very strong variable to take into account regarding state happiness.

1 Introduction

Numerous studies on well-being are published every year. The UN's 2012 World Happiness Report attempts to quantify happiness on a global scale using a 'Gross National Happiness' index which uses data on rural-urban residence and other factors [12]. In the U.S., Gallup and Healthways produce a yearly report on the well-being of different states, cities, communities and congressional districts, with a well-being index based on continual polling and survey data [9]. Another known initiative is the Happy Planet Index, introduced by the New Economics Foundation and that measures human well-being and environmental impact [20].

While these and other approaches to quantifying the sentiment of a population (be either a state or a city) as a whole rely almost exclusively on survey data, there are now a range of complementary, remote-sensing methods available to researchers. For the social sciences, the explosion in the amount and availability of data relating to social network use in the past 10 years has led to a collective, open recording of an enormous number of transactions, interactions, and expressions, marking a clear transition in our ability to quantitatively characterize, and thereby potentially understand, previously hidden as well as novel micro-scale mechanisms underlying socio-technical systems [13]. This has

driven a rapid increase in the application of data-driven techniques to the social sciences and sentiment analysis of large-scale populations.

My overall aim in this paper is to investigate how geographic place correlates with societal levels of happiness. In particular, I will be examining happiness dynamics at the level of U.S. states and counties, asking if it is possible to accurately measure the overall average happiness of people located in states and counties. My methodology for answering the first question uses word frequency distributions collected from a large corpus of messages posted on Twitter, with individual words scored for their happiness independently by users of Amazon's Mechanical Turk service. This technique was introduced by Dodds and Danforth in 2009 [6], and greatly expanded upon in [7], as well as used and tweaked by others such as [14].

Among many results, I generated three choropleths: (1) one with the distribution of happiness score by U.S. state, (2) another that uses the same technique that Gallup-Healthway uses fo their reports, and I use this choropleth to compare my results with their index (3) and finally one with the same set of happiness score, but distributed by U.S. counties. In fact, many of the other results have a focus on the comparison between my results and Gallup-Healthway report. Finally, a striking correlation value appeared between the number of tweets and the position in the happiness state ranking, the more tweets a state has the more sad it appears to be.

My paper is structured as follows. In Section 2 we describe the previous related work related to my sutdy. In Section 3, I describe the data sets we used. In Section 4 I describe my methodology for measuring happiness. In Section 5 I measure the happiness of different states and counties and determine the happiest and saddest states in the U.S., and at the same time I compare my results with the Gallup-Healthways Well-being Index [9]. I conclude with a discussion and a reference for future work in Section 6.

2   Related Work

In [2, 7, 8, 14] the overall aim is to investigate how geographic place correlates with and potentially influences societal levels of happiness. In particular they ask if it is possible to (a) measure the overall average happiness of people located in U.S. states and cities, and (b) explain the variation in happiness across different states and cities. The first question is answered using word frequency distributions collected from a large corpus of *tweets* posted on Twitter, with individual words scored for their happiness using the technique by Dodds et al. [6], which is an improvement of the original ANEW approach [4]. To answer the second question of happiness variability, they examine how individual word usage correlates with happiness and various social and economic factors, using the 'word shift graph' technique developed in [6, 7], and socioeconomic census data to attempt to explain the usage of certain words. To obtain a score for a given tweet T containing N unique words, they calculate the average happiness h_{avg} using an weighted average of the number of times a word occurs in T, multiplied by

the the h_{avg} of that specific word. They then divide the weighted average of all the words, with the total number of detected LabMT words. Finally they used the average of the scores, to achieve the happiness scoring of a given state.
RESULTADOS DESTES PAPERS

In [1] they use Twitter to study the fine-grained geography and dynamics of sentiment in the greater New York City area, identifying areas and times of positive and negative sentiment. They develop a classifier based upon supervised learning specifically tuned for 140-character Twitter messages, using key words, phrases and emoticons to determine the mood of each tweet. They use labeled tweets regarding the presence of emoticons, to build two classifiers: one to classify whether or not a tweet is positive, and another to classify whether or not a tweet is negative. This method, combined with geotagging provided by users, enabled them to gauge public sentiment on extremely fine-grained spatial and temporal scales. They find that public mood is generally highest in public parks and lowest at transportation hubs, and locate other areas of strong sentiment such as cemeteries, medical centers, a jail, and a sewage facility. Sentiment progressively improves with proximity to Times Square. Periodic patterns of sentiment fluctuate on both a daily and a weekly scale: more positive tweets are posted on weekends than on weekdays, with a daily peak in sentiment around midnight and a nadir between 9:00 a.m. and noon.

In [5] they connect measures of public opinion derived from polls with sentiment measured from analysis of text from Twitter. They explicitly link measurement of textual sentiment in microblog messages through time, comparing to contemporaneous polling data. In this preliminary work, summary statistics derived from extremely simple text analysis techniques are demonstrated to correlate with polling data on consumer confidence and political opinion, and can also predict future movements in the polls. They derive day-to-day sentiment scores by counting positive and negative messages. Positive and negative words are defined by the subjectivity lexicon from OpinionFinder. They do not use the lexicon's distinctions between weak and strong words. They define the sentiment score on day as the ratio of positive versus negative messages on the topic, counting from that day's messages.

In [18] they address subjective well-being, as measured by life satisfaction. They correlated the words within Twitter messages (in the form of LDA-generated word topics) mapped to U.S. counties, with the results for life satisfaction, as measured by questionnaires answered in those counties. Their goal is not to count positive sentiment words, like the studies above, but to study the language of well-being, to better understand the multiple components that contribute to it, i.e. they find the lexica and topics that correlate with life satisfaction as measured in questionnaires. They also add several control variables such as age, sex, minority, income, education. Finally, they run all the variable - the lexica, the topics and the control variables - through a LASSO linear regression with life satisfaction over counties. The surprising fundamental result of this paper is that they can predict (on average) the happiness of one set of people (those who

answered the LS questionnaires) from the tweets of other people (people in the same county).

3 Datasets

I examine a corpus of over 232 million tweets gathered from all over the world during the calendar year of 2012. After an initial parsing effort, I reduced the data set to have only geotagged tweets, which is near 0.78% of the whole data set (less than 1.8 million tweets). For the present study, I focus only on the geotagged tweets located in the Continental U.S., excluding Alaska and Hawaii, which is near 27% of the geotagged tweets (less than 0.5 million tweets).

Polygonal areas corresponding to the different U.S. were obtained from the state data set that is available from the *maps* package for the R system for statistical computing [1].

To measure sentiment (hereafter happiness) in these areas from the corpus of words collected, I use the Language Assessment by Mechanical Turk (LabMT) word list (available online in the supplementary material of [7]), assembled by combining the 5'000 most frequently occurring words in each of four text sources: Google Books (English), music lyrics, the New York Times and Twitter. A total of 10'233 of these individual words have been scored by users of Amazon's Mechanical Turk service on a scale of 1 (sad) to 9 (happy), resulting in a measure of average happiness for each given word [11]. For example, 'rainbow' is one of the happiest words in the list with a score of $h_{avg} = 8.1$, while 'earthquake' is one of the saddest, with $h_{avg} = 1.9$. Neutral words like 'the' or 'thereof' tend to score in the middle of the scale, with $h_{avg} = 4.98$ and 5 respectively.

Fig. 1 presents (i) the total number of tweets per state, (ii) the number of tweets containing LabMT words, and (iii) the percentage of tweets containing LabMT words. In median 75% of the total tweets have LabMT words.

Finally I will be using the 2012 Gallup-Healthways Well-being Index results [9] [2]. This study has an *Overall Well-Being Index* that I will use as a happiness index to correlate with our results.

4 Methodology

My section is structured as follows. In Section 4.1 I detail how I processed the tweets with the LabMT approach from [7]. In Section 4.2 I present how I processed the placement of each geotagged tweets in its appropriate state. In Section 4.3 I present some processing I made regarding the Gallup-Healthway Well-Being Index. In Section 4.4 I explain how I reached a single score for each state. In Section 4.5 I present some visualization techniques I adopted to present the results from my work.

---

[1] available at http://cran.r-project.org/web/packages/maps/
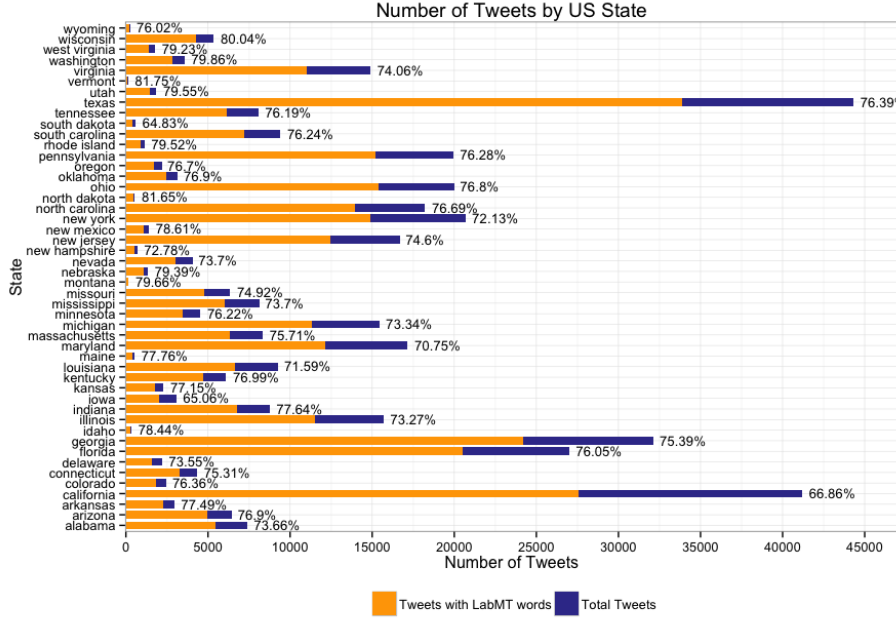[2] available at http://info.healthways.com/wbi2013

Fig. 1: Stacked bar chart for the number of tweets by U.S. state. In blue the total number of tweets, and in orange the number of tweets with words from the LabMT corpus of words. In front of each bar I have the percentage of tweets that have LabMT words.

4.1 LabMT Tweet Processing

As stated in Section 3, to measure sentiment (hereafter happiness) in the areas from the corpus of words collected, I use the Language Assessment by Mechanical Turk (LabMT) word list [7]. This word list has a total of roughly 10'000 of individual words, which have been scored on a scale of 1 (sad) to 9 (happy), resulting in a measure of average happiness for each given word [11]. For a given text T containing N unique words, I calculate the average happiness h_{avg} by:

$$h_{avg}(T) = \frac{\sum_{i=1}^{N} h_{avg}(w_i) f_i}{\sum_{i=1}^{N} f_i} \tag{1}$$

where f_i is the frequency of the ith word w_i in T for which I have a happiness value $h_{avg}(w)$, and $p_i = f_i / \sum_{i=1}^{N} f_i$ is the normalized frequency of word w_i. This is a similar approach of the original ANEW research [4], that's explained in [6]. This method makes no attempt to take the context of words or the meaning of a text into account. While this might lead to difficulties in accurately determining the emotional content of small texts, it was proved that for sufficiently large texts this approach nonetheless gives reliable (if eventually improvable) results [6, 7,

11,14]. Following Dodds et al. [7], I remove all words w_i for which the happiness score falls in the range $4 < h_{avg}(w_i) < 6$ when calculating $h_avg(T)$. Removal of these neutral words has been demonstrated to provide a suitable balance between sensitivity and robustness in [7]. Following [14], to avoid the problem that some states have happier names than others I removed (a) each state name from the calculation for h_{avg}; and (b) all variants of the racial pejorative or 'N-word', since variants of this word have very low happiness values, and consequently were found to be highly influential in determining the state happiness score, however upon examining individual tweets [14] found that this word appeared to be being used in conversation as a more colloquial stand in for the word 'friend', and not in fact in any particularly negative sense, therefore the word was unfairly biasing [14] the results towards the negative and removed it.

To increase the amount of tweets considered in the study, I also processed the same data set but using a Spanish list of words resulting from work by [17] [3]. This process allowed me to increase the number of tweets in near 40'000. To avoid tweets with score with both lists, leading to duplicated tweets, we only choose the score that from the language were it were detected more words with. Example, if we have this tweet: "You are a very good friend, compadre!", even if we have the one word from the Spanish list ('compadre'), there are 2 words form the English list ('good', 'friend'). Therefore, we choose to score this tweet with the English word list.

4.2 Geolocating Tweets

Each geotagged tweet from the data set was processed in order to extract its latitude and longitude coordinates, searching for those fields in the JSON object corresponding to each Twitter message [4]. After stripping those two fields, and to be able to match each tweet to a U.S. state, I used the *sp* R package [5] more specifically the *over* function, which does consistent spatial overlay for points, grids and polygons, i.e., if a latitude/longitude point x and a polygon P representing a state is given, the over function returns *true* or *false*, if that point is inside the polygon - if a certain tweets belongs to a certain state.

4.3 Gallup-Healthway Score

I will correlate our happiness results with with the 2012 Gallup-Healthways Well-being Index [9] [6]. Since LabMT score goes from 1 to 9, I had to rescale the Gallup-Healthways Overall Well-Being Index from a percentage scale (0-100) to the 1-9 point scale used in the LabMT lexicon. For this I used the known 'typical' function to rescale values from one range to another:

$$rescale_(x) = \frac{(b-a)(x-min)}{max-min} + a \tag{2}$$

---

[3] word list available at http://link.springer.com/article/10.3758%2FBF03193031
[4] available at https://dev.twitter.com/docs/platform-objects/tweets
[5] available at http://cran.r-project.org/web/packages/sp/
[6] available at http://info.healthways.com/wbi2013

where, for our case: $min = 0$, $max = 100$, $a = 1$ and $b = 9$.

4.4 Score by U.S. State

After having all tweet associated with a state, I had to choose one of the three central tendency statistics that allowed us to get a score for a whole state state, using either the *mean*, *median* or the *mode*. To choose between them I plotted the histograms for the score distributions of some states (see Fig. 2), to observe the patterns for tweet distribution, and I calculated the correlation factor between those three statistics and the Gallup-Healthway Index.
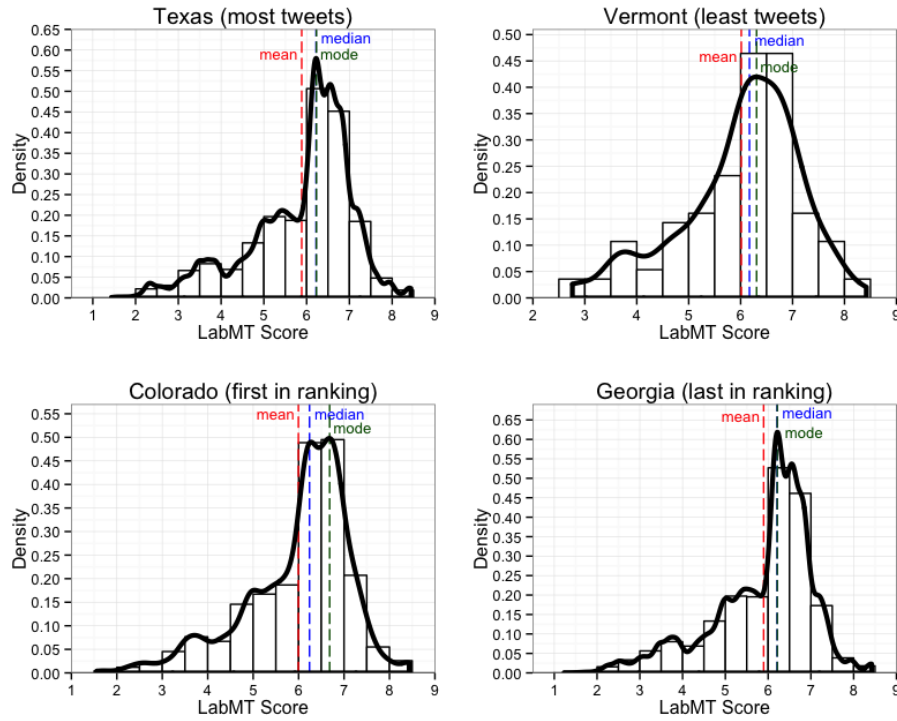
Fig. 2: Histograms and density lines for Texas (state with most tweets), Vermont (state with least tweets), Colorado (first in state happiness ranking) and Georgia (last in happiness state ranking). Vertical red line is the mean of the scores, vertical blue line is the median of the scores, and vertical green line is the mode for the scores.

An initial analysis lead us to choose the *mode* because the correlation was the highest. The mode method I used, the most common one, was the Pearson

correlation coefficient. The Pearson correlation is a measure of the linear correlation (dependence) between two variables X and Y, giving a value between +1 and -1 inclusive, where 1 is total positive correlation, 0 is no correlation, and -1 is total negative correlation. It is widely used as a measure of the degree of linear dependence between two variables. It is defined as the covariance of the two variables divided by the product of their standard deviations. [16]

I limited the mode density function interval to $[1, 9]$, the range of values of the LabMT word scores, and used a adjust (smooth) factor of 0.80. This allowed us to achieve a correlation of approximately 0.45. Taking in consideration what has been shown in other studies [2,7,8], that a correlation value between studies near similar to this one are usually around the 0.50 mark, we can conclude that our approach is yielding very good results. The correlation results are shown below:

$$X = \text{Gallup-Healthway Well-Being Index score by state} \qquad (3)$$
$$cor(X, Y) = 0.187, \text{ with Y} = \text{mean of LabMT tweet scores by state}$$
$$cor(X, W) = 0.029, \text{ with W} = \text{median of LabMT tweet scores by state}$$
$$cor(X, Z) = 0.448, \text{ with Z} = \text{mode of LabMT tweet scores by state}$$

4.5 Visualization

Finally I present our results using one of the visualization methods discussed in [15], which is also used in [7,14], the Choropleth map. Th Choropleth map uses the graphic variables describing properties of color or texture to show properties of non-overlapping areas, such as provinces, districts, or other units of territory division. A number of categories is mapped to distinct colors or textures which are used to fill the areas accordingly. For the colors used on our choropleths, I used the *RdYlGn* palette from the *RColorBrewer* R package [7], which has a set of default color palettes based on the work done by [10] [8]. I presented our maps using the *globular projection* detailed in [3].

5   Results

First I plot the average happiness of all Continental U.S., excluding Alaska and Hawaii in Fig. 3. In this chart I plot our results redefining the scale so I get a greater amount of different colors near the mean value of our state scores. This way I avoid using an uniform scale, that causes most of the states to be with the same color. Our scale is an approximation of a Normal Distribution, but in a ad-hoc manner: $scale = \langle min, mean - (sd/2), mean - (sd/3), mean, mean + (sd/3), mean + (sd/2), max \rangle$. The result is a scale that increases the amount of different colors near the mean, and uses the same ones in the extremities.

---

[7] available at http://cran.r-project.org/web/packages/RColorBrewer/
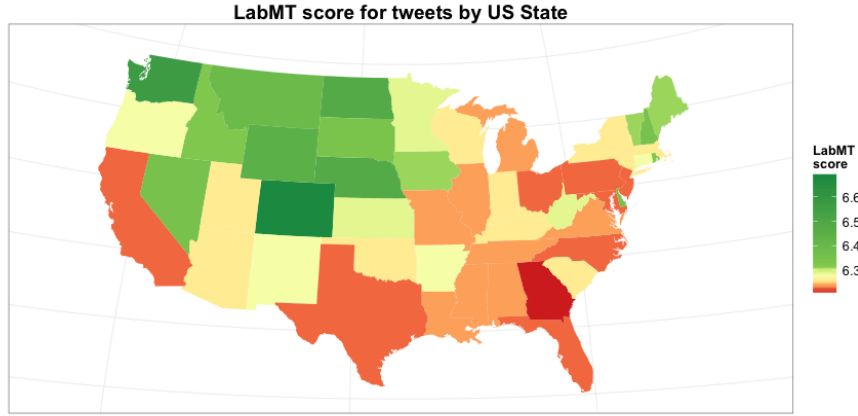[8] available at http://colorbrewer.org

Fig. 3: Choropleth showing *mode* word happiness for geotagged tweets in all Continental U.S. states collected during the calendar year 2012. The happiest 5 states, in order, are: Colorado, Washington, Nebraska, North Dakota, Wyoming. The saddest 5 states, in order, are: Georgia, Texas, Pennsylvania, Ohio, North Carolina. The full state ranking is shown in Appendix section.

At such a coarse resolution there is little variation between states, which lie between 0.09 and 0.38. The h_{avg} of the mean value for the entire U.S. is 6.30 and the standard deviation has a value of 0.09. The top 5 states are the ones that are far apart from the mean value, are in average 0.23 above the mean. The happiest state is Colorado with a score of $h_{avg} = 6.68$ and the saddest state is Georgia with a score of $h_{avg} = 6.21$. The striking observation that I take from the ranking is the high correlation between the state's position in the ranking and the number of tweets of that state, cor(number of tweets, ranking) = 0.77, and also a high correlation between the score and the number of tweets, cor(number of tweets, LabMT score) = -0.48. This means that the more tweets a state has the lower it will be in the ranking (and the sadder it will appear to be). In fact, if I split our states into equal parts (6 states each group), I can clearly see the average number of tweets per group increases as I lower in the ranking (see Table 1).

Another interesting way to look at the data is to emulate the style that Gallup-Healthway uses for the choropleth in their report. They split up the scores per state in five quintiles, coloring them in the map. For the sake of easier comparison, I present both the original and our own choroplet using the same scheme Gallup-Healthway uses.

Even tough there is no striking resemblance between the two charts, if you take into account the little variation of the scores that I discussed before, I can assume that a state could easily be on another quintile (except the top 5, as I also discussed before). In the Southeast I have similar results, with some of the

Ranking	Average number of tweets
1-12th	1376
13-24th	4391
25-36th	5344
37-48th	14479

Table 1: Average of tweets per position in state ranking

states that should be in the 4th Quintile are in the 5th. Northeast section is very far apart with many of the values on the 4th/5th Quintile when they should be in the 3rd/2nd, still in the north part there's some improvement. Midwest section has yet again some striking resemblance between the 1st and 2nd Quintile states. Southwest state also have some resemblance with some states switching between the 3rd and 5th Quintile. Finally, the West appears similar with some states switching between the 1st and 2nd Quintile, but with a staggering difference in California. In fact, there are very few states that jump more than 3 Quintiles in their classification.

You can attest that analyzing the ranking differences in the bump chart in Fig. 5 that compares the changes in ranking between the LabMT approach and Gallup-Healthway score. What I aim with this graph is a high number of straight (or almost straight) lines. Those indicate that the rank stayed the same (or almost the same) in both approaches. As I expected (since I achieved a correlation of approximately 0.45), a reasonable amount of lines are straight. Regarding the lines that aren't straight, I can observe that they connect extremes of the ranking (with rare exceptions).

Regarding score difference in spite of the little variation there is our scores, the average difference between Gallup-Healthway score and is just 0.02. What happens is that, and I can verify that analyzing Fig. 6, I either have very small differences (in the range of few hundredths) or I have very large differences (in the range of several tenths). This leads to situations I can see in Fig. 5, where many times I see a state changing very few positions in ranking and other times changing many positions.

Approaches similar to Gallup-Healthway report, i.e. approaches that quantifying the sentiment of a population that rely almost exclusively on survey data, are limited in the scale they are made. Using this approach it would be very difficult, if not impossible, to use the exact same data set to analyze a different scale, e.g. at the level of counties. Currently Gallup-Healthway focus their research on State, Community and (some) cities. Meanwhile using my approach and dataset, I could be able to reach a level of detail based on street if necessary. I demonstrate this with analysis with a chropleth at the level of U.S. counties (Fig. 7), clearly showing the huge potential of these types of approaches when compares with survey-like approaches.

The scale is the same as in the original choropleth (Fig. 3) so the comparison between the two maps is easier. This forced us to cap the values used in the county map to the minimum and maximum score values that I found in the state map. There is a great amount of dispersion regarding the number of tweets by county, with a mean of 230 tweets per county but with a sd of 675. That's why I observe weird patterns like the state of California, where the counties are mainly green but the red counties are the ones with the most tweets. When I move to a coarser scale, those counties with the highest number of tweets will contribute more towards the mode score, leading to a shift in the overall state score.

6   Discussion and Future Work

In this paper I have examined word use in urban areas in the U.S., using a simple mathematical method which has been shown to have great flexibility, sensitivity and robustness. I have used this tool to map areas of high and low happiness and score individual states (and also counties) for average word happiness.

My state ranking is backed up with a strong correlation between my results and the results of Gallup-Healthways report but also with the results in [14]. In fact, if you consider that the correlation between [14] and the Gallup-Healthway report is 0.50, my correlation value of 0.48 for the same report, is an excellent value for the scale of my research in comparison with [14] study. Meanwhile, the correlation between my study and [14] is also of 0.45. This indicates what have been shown in other studies [7], that a correlation value between studies similar to these are usually around the 0.50 mark, and if we reach such a value we can consider that our approach is yielding good results and that's robust enough.

The one conclusion that we take from our study, that I didn't find in any of the other related work is the very high correlation between the amount of tweets and the ranking (or score) of a state. The value of 0.77 value that we found clearly indicates that Twitter usage might be correlated with happiness in that state. A different interpretation we can take form this value, is that the more individuals use Twitter, the more they write messages with negative connotation. And who uses less, uses Twitter to broadcast messages with positive connotation.

There are a number of legitimate concerns to be raised about how well the Twitter data set can be said to represent the happiness of the greater population. Roughly 15% of online adults regularly use Twitter, and 18-29 year-olds and minorities tend to be more highly represented on Twitter than in the general population [19]. Furthermore, the fact that the collected sample is small compared with the overall Twitter usages, and that the geotagged tweets are even less means that our data set is a non-uniform subsample of statements made by a non-representative portion of the population. One could improve the amount of geotagged tweets, considering for the non-geotagged tweets that they were made from the city the user filled in its Twitter registration form.

In this work I have only scratched the surface of what is possible using this particular dataset. I have not examined whether or not these methods have

any predictive power—future research could look at how observed changes in the Twitter data set, predict changes in the underlying social and economic characteristics measured using traditional survey methods. A possible approach would implement a predictive regression model similar with [18], but taking into consideration the new conclusions I have withdrawn from this work, specially that there is a strong correlation between the amount of twitter messages (i.e. Twitter usage) and the happiness score, introducing in their LASS model another variable indication Twitter usage.

References

1. Bertrand, K.Z., Bialik, M., Virdee, K., Gros, A., Bar-yam, Y.: Sentiment in new york city : A high resolution spatial and temporal view pp. 1–12 (2013)
2. Bliss, C.a., Kloumann, I.M., Harris, K.D., Danforth, C.M., Dodds, P.S.: Twitter reciprocal reply networks exhibit assortativity with respect to happiness. Journal of Computational Science 3(5), 388–397 (9 2012), `http://linkinghub.elsevier.com/retrieve/pii/S187775031200049X`
3. Bolstad, P.: GIS Fundamentals: A First Text on Geographic Information Systems (2012)
4. Bradley, M.M., Lang, P.J.: Affective norms for english words (anew): Instruction manual and affective ratings (1999)
5. Connor, B.O., Balasubramanyan, R., Routledge, B.R., Smith, N.A.: From tweets to polls : Linking text sentiment to public opinion time series (2010)
6. Dodds, P.S., Danforth, C.M.: Measuring the happiness of large-scale written expression: Songs, blogs, and presidents. Journal of Happiness Studies 11(4), 441–456 (7 2009), `http://link.springer.com/10.1007/s10902-009-9150-9`
7. Dodds, P.S., Harris, K.D., Kloumann, I.M., Bliss, C.a., Danforth, C.M.: Temporal patterns of happiness and information in a global social network: hedonometrics and twitter. PloS one 6(12), e26752 (1 2011), `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3233600&tool=pmcentrez&rendertype=abstract`
8. Frank, M.R., Mitchell, L., Dodds, P.S., Danforth, C.M.: Happiness and the patterns of life: a study of geolocated tweets. Scientific reports 3, 2625 (1 2013), `http://www.ncbi.nlm.nih.gov/pubmed/24026340`
9. Gallup - Healthway: State of American Well-Being - 2012 State, Community, and Congressional District Analysis (2013)
10. Harrower, M., Brewer, C.a.: Colorbrewer.org: An online tool for selecting colour schemes for maps. The Cartographic Journal 40(1), 27–37 (6 2003), `http://www.maneyonline.com/doi/abs/10.1179/000870403235002042`
11. Kloumann, I.M., Danforth, C.M., Harris, K.D., Bliss, C.a., Dodds, P.S.: Positivity of the english language. PloS one 7(1), e29484 (1 2012), `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3256157&tool=pmcentrez&rendertype=abstract`
12. Layard, R., Sachs, J.: World happiness report 2013. World Happiness Report 2013 pp. 2 – 153 (2013)
13. Miller, G.: Social scientists wade into the tweet stream (2011)
14. Mitchell, L., Frank, M.R., Harris, K.D., Dodds, P.S., Danforth, C.M.: The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place plos one, 8(5): e64417, 2013 8(5), 1–20 (2013)

15. Nollenburg, M.: Geographic visualization. Human-Centered Visualization Environments pp. 257–294 (2007)
16. Pearson, K.: Note on regression and inheritance in the case of two parents (2006)
17. Redondo, J., Fraga, I., Padrón, I., Comesaña, M.: The spanish adaptation of anew (affective norms for english words). Behavior research methods 39(3), 600–5 (8 2007), http://www.ncbi.nlm.nih.gov/pubmed/17958173
18. Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Dziurzynski, L., Agrawal, M., Park, G.J., Lakshmikanth, S.K., Jha, S., Seligman, M.E.P., Ungar, L., Lucas, R.E.: Characterizing geographic variation in well-being using tweets (June 2009) (2013)
19. Smith, A., Brenner, J.: Twitter use 2012 (2012)
20. The New Economics Foundation: Happy planet index (2012), http://www.happyplanetindex.org/assets/happy-planet-index-report.pdf
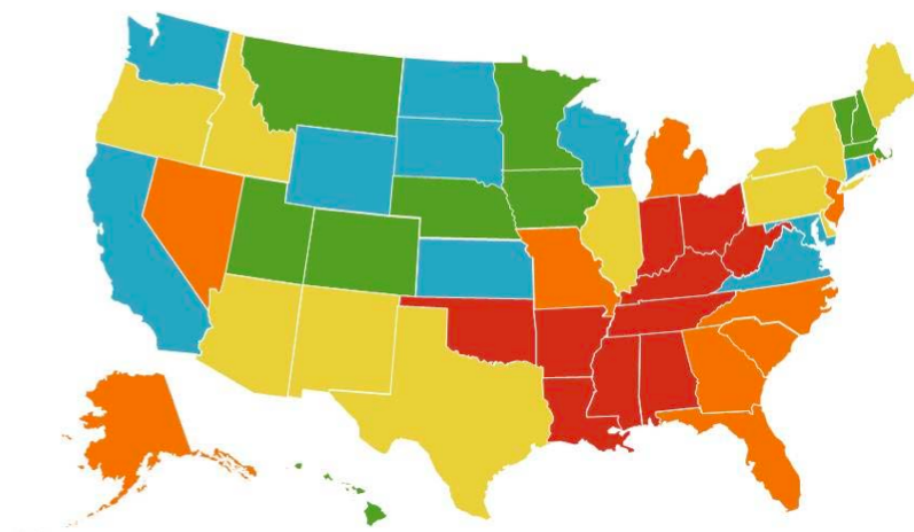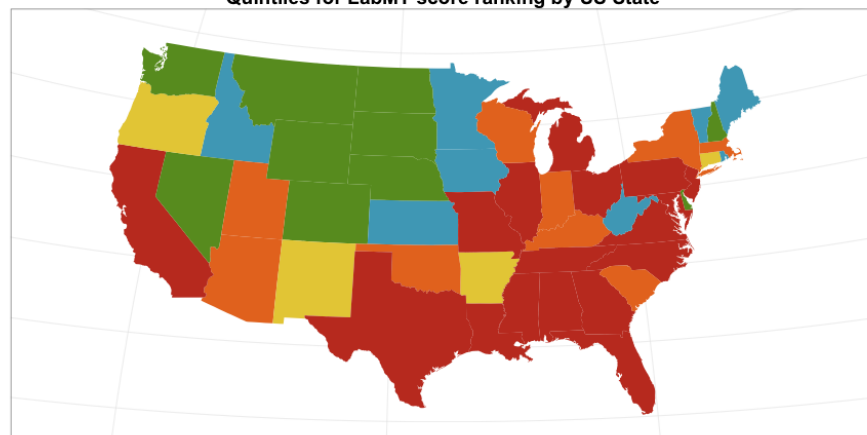
Fig. 4: Gallup-Healthway Well-Being Index choropleth vs Our own choropleth using the same quintile schema as well as the same color code.The happiest 5 states for Gallup-Healthway study are, in order, are: Colorado, Minnesota, Utah, Vermont, Nebraska. The saddest 5 states, in order, are: Arizona, Tennessee, Mississippi, Kentucky, West Virginia. The full state ranking is shown in Appendix section. Source [9]
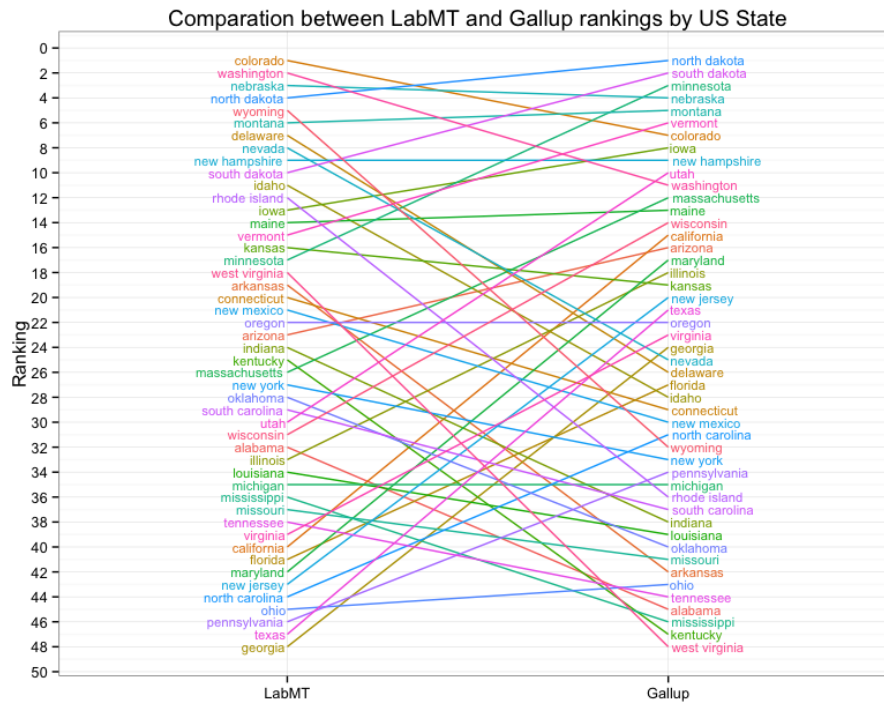
Fig. 5: Bump chart that compares the changes in ranking between the LabMT approach and Gallup-Healthway score. On the right column you have the ranking from our approach using LabMT word list. On the left column I have the Gallup-Healthway ranking for the calendar year of 2012 [9], minus the states that I didn't considered in our study, i.e., Alaska and Hawaii.
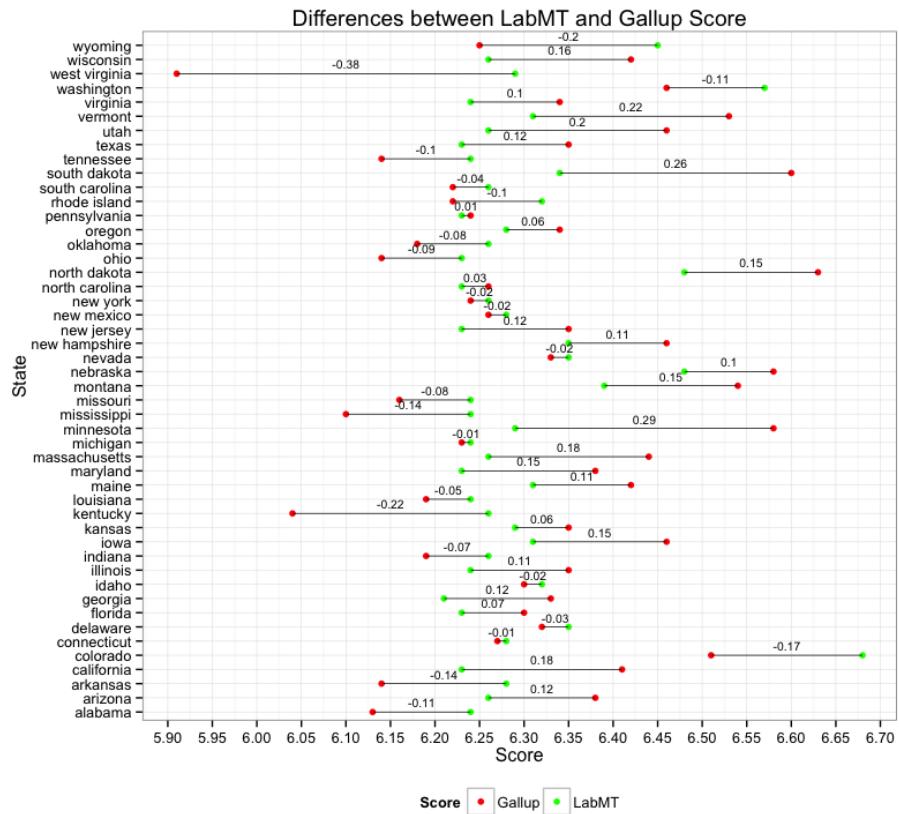
Fig. 6: Difference in the score using our LabMT approach, and Gallup-Healthway Well-Being Index, rescaled to the range used in [6, 7] method, i.e. 1 to 9. I can see - just like it happen with the ranking - that when the difference is small, it is very small, and when the difference is large, it's significantly large.

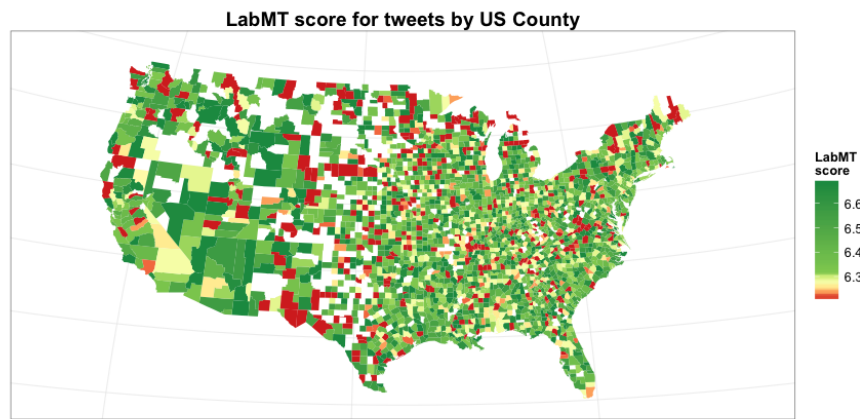**LabMT score for tweets by US County**

Fig. 7: Choropleth showing *mode* word happiness for geotagged tweets in all Continental U.S. counties collected during the calendar year 2012. The scale is the same as the one used in Fig. 3, with values capped to the range of scores of the state.