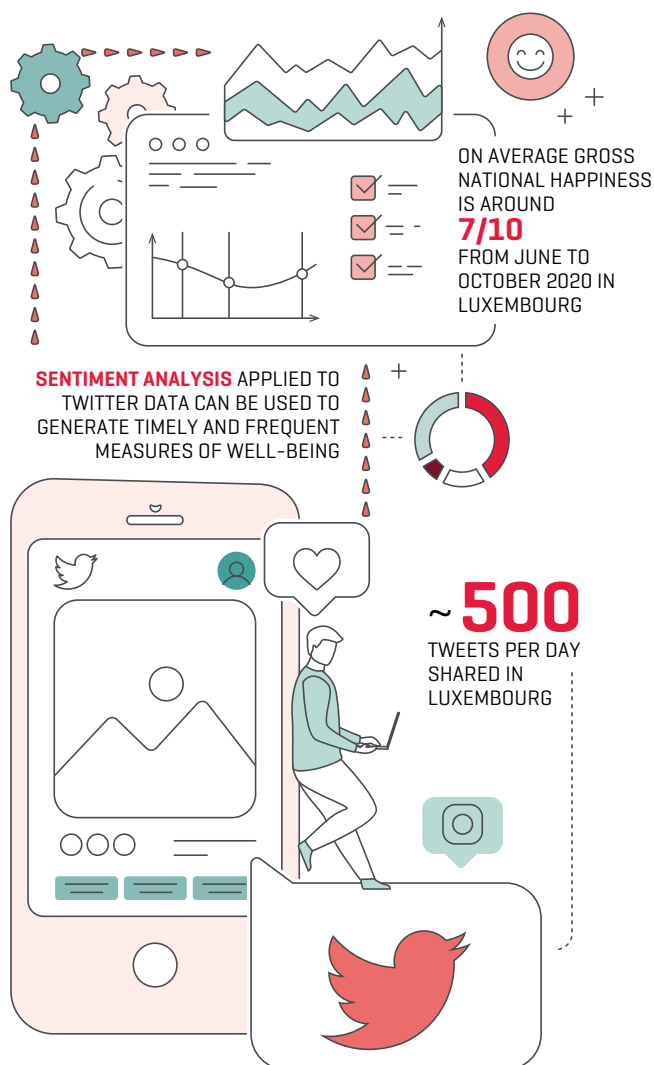# From tweets to statistics[1]

Sentiment analysis of tweets reveals how people's feelings have changed during the pandemic in Luxembourg.

Sentiment analysis applied to Twitter data provides timely insights into how people in Luxembourg fared during the COVID-19 crisis. This short article reports some preliminary results from the project "Preferences through Twitter" and showcases some of the possibilities and limitations made available by the new techniques. STATEC Research leads the project with the support of Fonds National de la Recherche.

ON AVERAGE GROSS NATIONAL HAPPINESS IS AROUND **7/10** FROM JUNE TO OCTOBER 2020 IN LUXEMBOURG

SENTIMENT ANALYSIS APPLIED TO TWITTER DATA CAN BE USED TO GENERATE TIMELY AND FREQUENT MEASURES OF WELL-BEING

~500 TWEETS PER DAY SHARED IN LUXEMBOURG

Improving people's well-being has several beneficial consequences, besides being a desirable outcome per se. Happier people tend to live longer and healthier lives, have better employment outcomes, be more productive and collaborative, and make for less absent workers. Additionally, higher levels of past and present happiness predict higher levels of compliance during COVID-19 [Krekel et al., 2020]. However, traumatic events – such as the COVID-19 pandemic – can alter well-being in rapid and persistent ways, thus triggering medium to long-run social and economic consequences. How can we monitor the changes in people's well-being during the pandemic? Researchers at STATEC, University of Johannesburg (South Africa) and Auckland University of Technology (New Zealand) retrieve timely and frequent data from applying sentiment analysis to tweets. Such data are relevant for effective decision-making because people's well-being can affect the success of health policies, "exit" strategies to ease lockdowns, and recovery plans.

Well-being data are usually collected via large scale surveys that take time to administer, and are thus only available after some delay. For instance, the European Commission administers Eurobarometer surveys multiple times a year in each of the member states. In 2019, the Eurobarometer assessed people's well-being twice; whereas in 2020, the survey was only administered once (see table 1). The Eurobarometer indicates that life satisfaction, a valid and reliable measure of well-being, decreased by eight percentage points between Autumn 2019 and Summer 2020. Survey data alone can not explain the decrease.

It is possible that it is due to the pandemic, but how much of an influence has it had and through which mechanisms? Even the size of well-being impacts is somewhat unknown. It could be that the decrease was actually larger than eight percentage points, but individuals could have recovered somewhat by the time they participated in the survey. Or, perhaps, the survey was administered at a specifically low point in their life, but individuals recovered quickly. New sources of timely information can valuably complement survey data.

TABLE 1. LIFE SATISFACTION IN LUXEMBOURG FROM SPRING 2019 TO SUMMER 2020.

|	Spring 2019	Autumn 2019	Summer 2020
people not satisfied with their lives (%)	4	6	14
people satisfied with their lives (%)	96	94	86

Source: Eurobarometer data (European Commission 2019, 2020). The original variable is organized into four categories. For ease of interpretation, the bottom and top two categories have been collapsed.

Thus, researchers turned to Twitter data. Every day a large number of people around the world share their opinions, reactions, discoveries, worries, questions, and decisions via tweets. In Luxembourg, people share approximately 500 tweets per day, which is nearly 3 500 tweets per week. This wealth of short texts provides a real-time source of information that can be transformed into usable data using sentiment analysis. Sentiment analysis is an automated process that uses natural language processing to determine the feelings and attitudes of a written text's author (Hailong et al. 2014). It has been used by many researchers, for instance, to predict future events, such as the result of elections or stock markets, to track the influenza rate, and to measure people's well-being (Iacus et al., 2020).

STATEC Research collaborates with researchers from the University of Johannesburg (UJ) and Auckland University of Technology (AUT), with the funding support of the FNR, South Africa's URC and AUT, to compute a real-time measure of well-being – the Gross National Happiness (GNH) which was developed by Greyling et al. (2019) – in Luxembourg and in a selected group of European countries during the COVID-19 pandemic.

The GNH in Luxembourg (solid line in Figure 1) increased from the end of June to mid-July, perhaps because COVID-19 and the lockdown seemed to be in the past. However, as people became aware of a resurgence in new cases, their well-being declined suddenly and continued to do so. In fact, **the changes in GNH are consistent with the changes in the number of new positive cases, though in opposite direction** (dashed line)**.** Notice in particular the sharp decline in GNH and the large increase of new positive cases in the month of October.

FIGURE 1. GROSS NATIONAL HAPPINESS AND NUMBER OF NEW POSITIVE COVID-19 CASES OVER TIME IN LUXEMBOURG.
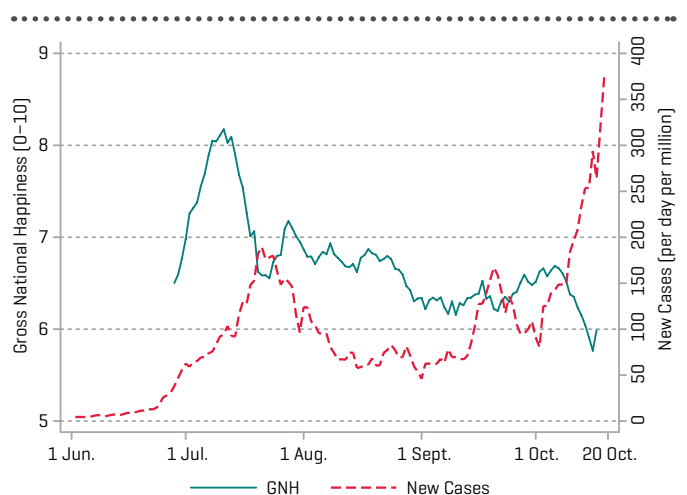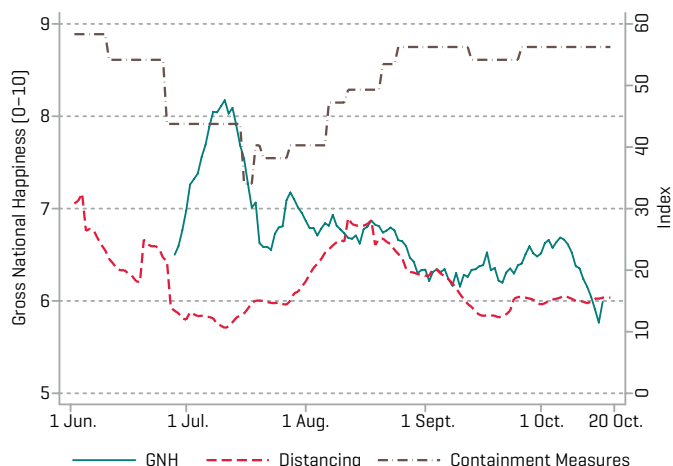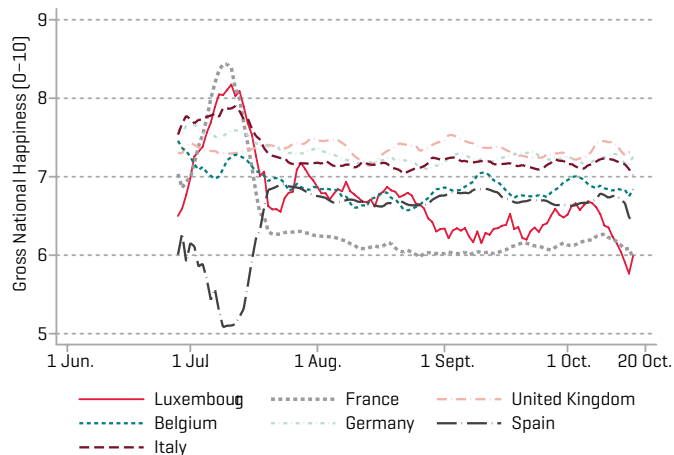

Source: GNH data (Greyling et al., 2019) are sourced from the project "Preferences Through Twitter" with the support of FNR, UJ and AUT. COVID-19 data are from Roser et al. (2020). Both variables are smoothed using a centered, weekly moving average.

How did well-being change in relation to the containment and health policies adopted by the government? And did people comply with the containment policies? Figure 2 shows the rise in GNH during July coincides with a period of relaxation of containment measures (see the dash-dotted line in Figure 2), whereas the subsequent decline in GNH mirrors the progressively more stringent policies adopted afterwards.

The good news is that people generally complied with containment measures. This is shown by the close relationship between the changes in people's distancing behavior[2] (the dashed line) and the changes in containment measures (the dash-dotted line). **This evidence suggests people followed the provisions of the government in Luxembourg. Indeed the correlation between the two measures is 80%.** For the sake of comparison, the correlation is 71% in France, 68% in both Belgium and the United Kingdom, 55% in Germany, and 64% in Spain. The correlation in Luxembourg is second only to Italy (86%) among our sample countries.

---

2   The measure of distancing is based on O'Connor (2020) and uses Google mobility data.

Another advantage of sentiment analysis applied to Twitter data is that it allows researchers to compute GNH for any country in the world (with sufficient Twitter users). In Figure 3 the trend of GNH in Luxembourg is contrasted with those from six other European countries that were severely affected by COVID-19 during the spring of 2020: Italy, Spain, the United Kingdom, France, Belgium, and Germany. **The comparison reveals that since July, Luxembourg experienced greater changes than most of the other EU countries considered, as well as a greater loss of well-being: by the end of October, Luxembourg had the lowest GNH score.** Well-being declined in each country as we moved into autumn, but the changes in Italy, Germany, the UK, and Belgium are not as severe as those in Luxembourg, France, and Spain.

What drove these changes? And what explains the different levels of well-being across countries? Unfortunately, it is still too early to say. The country-specific mix of containment policies, contagion rates, and economic performance may contribute to answering these questions.

At present the research team is focused on producing the GNH score for each day since the first of January 2020, for Luxembourg and the other six European countries. Additionally, the team is working to widen the number of variables derived from Twitter to monitor the changes in economic insecurity, trust in others and institutions, and feelings towards immigrants. Importantly, the team is evaluating whether this data can be considered as reliable and accurate.

The researchers have gathered two pieces of evidence which support the validity of the GNH as a measure of well-being. First, the average GNH score correlates strongly with average life satisfaction as measured by the Eurobarometer (see Figure 4).[3] Second, GNH correlates meaningfully with a measure of negative emotions over time. Indeed, for a measure that has the benefit of timeliness and frequency, the real test of validity is correlation over time. To this end, the researchers created an index of negative emotions (dotted line in Figure 5) by averaging daily Google search scores for three topics: fear, sadness, and anger.[4] If GNH varies in a meaningful way over time, then it should correlate negatively with the index of negative emotions. The correlation between the two indices is indeed -69%. To ascertain whether the correlation is coincidental, the researchers run a falsification test: they compared the GNH with Google search scores for a "neutral" topic, one that should not be correlated with GNH. In this case, the chosen topic is Google itself, one of the most searched for terms. The dashed line in Figure 5, corresponding to searches for Google, is not correlated with GNH: the correlation coefficient is 9% and not statistically significant. This indicates that the GNH meaningfully correlates over time with a variable that is expected to correlate with and that this correlation is likely not driven by chance. In other words, GNH appears to be a valid measure of well-being over time.

---

3 The averages are computed by country over the period mid-July to the end of August.

4 Such search results are available daily by country and have been used in numerous research projects ranging from the assessment of economic conditions to individuals' feelings (e.g., Brodeur et al., 2020).
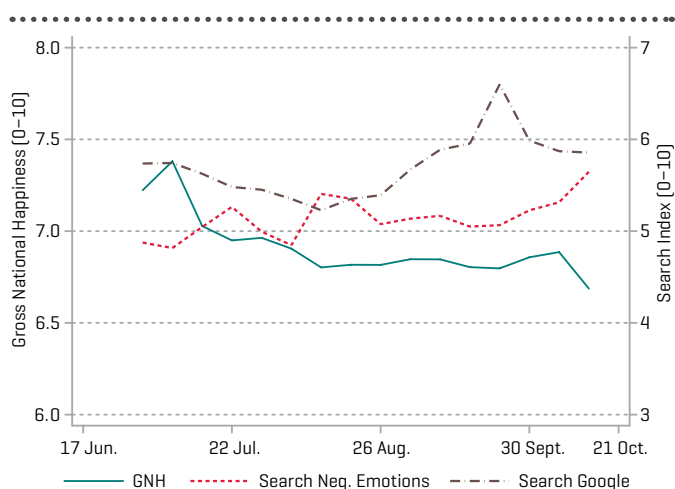
FIGURE 4. AVERAGE GROSS NATIONAL HAPPINESS CORRELATES POSITIVELY WITH AVERAGE LIFE SATISFACTION.
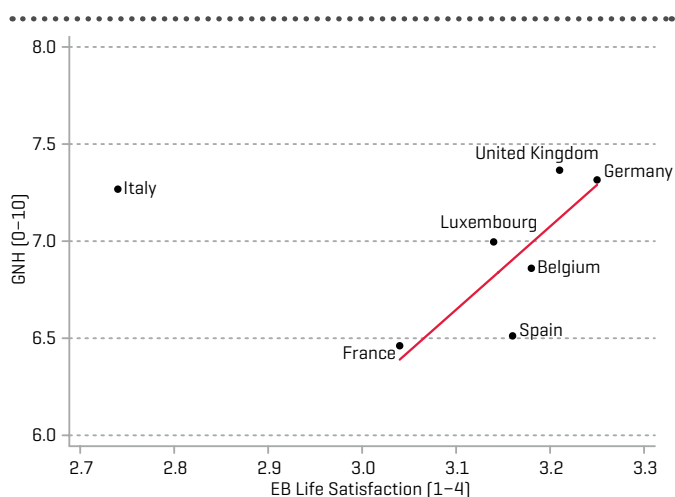
Source: GNH data (Greyling et al., 2019) are sourced from the project "Preferences Through Twitter" with the support of FNR, UJ and AUT. Life satisfaction data are from the Eurobarometer (European Commission, 2020), Summer 2020. The GNH score is the average by country over the period mid-July to the end of August, i.e. the same time the Eurobarometer was collected.

In summary, sentiment analysis applied to Twitter data can be used to generate timely and frequent measures of well-being. These figures will be extended to cover the whole of 2020, enriched with additional explanatory variables, and further tested for validity and reliability. GNH promises to deliver a timely measurement of how well-being changed throughout 2020 in Luxembourg and in international comparison. Such information can supplement the evidence based on survey data to provide invaluable insight for the general public, research community, and policy makers. Building upon and extending the data presented here, the team will deliver new results in the coming months.

FIGURE 5. AVERAGE GROSS NATIONAL HAPPINESS CORRELATES MEANINGFULLY WITH GOOGLE SEARCH ON NEGATIVE EMOTIONS.

Source: GNH data (Greyling et al., 2019) are sourced from the project "Preferences Through Twitter" with the support of FNR, UJ and AUT. Search data are from Google Trends (Google LLC 2020b). The three variables are the average (across seven countries) of the weekly averages within countries.

References:

Brodeur, A., Clark, A. E., Flèche, S., and Powdthavee N. (2020) COVID-19, Lockdowns and Well-Being: Evidence from Google Trends. *IZA Discussion Paper No. 13204*.

De Neve, J. E., & Oswald, A. J. (2012). Estimating the influence of life satisfaction and positive affect on later income using sibling fixed effects. *Proceedings of the National Academy of Sciences*, *109*(49), 19953-19958.

European Commission (2019) Public Opinion in the European Union: Standard Eurobarometer 92 Annex.

European Commission (2020) Public Opinion in the European Union: Standard Eurobarometer 93 Annex.

Google LLC (2020a) Google COVID-19 Community Mobility Reports. Published online at www.google.com/covid19/mobility. Accessed: 28 October 2020.

Google LLC (2020b) Google Trends. Published online at www.google.com/trends. Accessed: 28 October 2020.

Greyling, T., Rossouw, S. & Afstereo. (2020). Gross National Happiness Index. The University of Johannesburg and Afstereo.

Hailong, Z., Wenyan, G., & Bo, J. (2014, September). Machine learning and lexicon based methods for sentiment classification: A survey. In 2014 11th web information system and application conference (pp. 262-265). IEEE.

Hale, T., Webster, S., Petherick, A., Phillips, T., & Kira B. (2020). Oxford COVID-19 Government Response Tracker. Blavatnik School of Government. Accessed: 28 October 2020.

Iacus, S. M., Porro, G., Salini, S., & Siletti, E. (2020). How to exploit big data from social networks: a subjective well-being indicator via Twitter. Social Indicators Research, online first.

Krekel, C., Swanke, S., De Neve, J. E., & Fancourt, D. (2020). Are Happier People More Compliant? Global Evidence From Three Large-Scale Surveys During Covid-19 Lockdowns.

O'Connor K. J. (2020) Physical distancing worked to combat the spread of Covid-19 Using Google Mobility data. *Économie et Statistiques: Working Papers du STATEC* No 116.

Roser M., H. Ritchie, E. Ortiz-Ospina, & J. Hasell (2020) Coronavirus Pandemic (COVID-19). Published online at OurWorldInData.org. Accessed: 28 October 2020.