

Seminar: Mogelijkheden van twitter als bron voor openbare statistieken!

The use of sentiment analysis: Monitoring emotions in a society based on twitterdata

18 March 2021



UNIVERSITY
OF
JOHANNESBURG

A hand is shown holding a glowing, wireframe globe. The globe is surrounded by a network of white dots connected by thin lines, suggesting a digital or data network. The background is a dark blue gradient with a blurred cityscape at night.

‘The use of sentiment analysis: Monitoring emotions
in a society based on twitterdata’.

Prof. Talita Greyling
University of Johannesburg, School of Economics,
Director: Gross National Happiness.Today Project
talitag @uj.ac.za

1. Problems with the metrics of well-being

2. Big data

3. Twitter

- The choice of Twitter
- Twitter facts
- Twitter limitations and benefits

4. Gross National Happiness Today Project

- Method
- Validation
- Results
- Research
- Limitations
- Awards
- Future Projects



Problems with metrics of well-being (1/4)

GNP as a metric of well-being:

“The welfare of a nation can scarcely be inferred from the measurement of national income”

(Simon Kuznets, “father of modern GDP metrics” 1934)

“It measures everything , in short, except that which makes life worthwhile” -

(Robert Kennedy, University of Kansas, March 18, 1968)

Problems with metrics of well-being (2/4)

“What you **measure affects what you do“**

“If you cannot **measure it, you cannot manage it”**

Well-being/ Quality of life (QoL) is multi-dimensional and depends on people’s objective conditions and capabilities:

- health,
- education,
- personal activities (employment),
- environmental conditions.
- social connections,
- political voice,

The objective conditions predicts their life satisfaction (subjective well-being) (Stiglitz et al. 2019)

Problems with metrics of well-being (3/4)

Subjective well-being (SWB) is
'a person's cognitive and affective valuations of his or her life'
– Diener, Lucas, & Oishi, 2002, p. 63

The **cognitive** element
refers to what one thinks
about his or her life – an
evaluation of life

Survey (cognitive):
How things are these days?

The **affective** element
refers to emotions, moods
and feelings

Survey (hedonic levels of
affect):
How do you feel today?

Problems with metrics of well-being (4/4)

Survey data – problems (has been mentioned by Dries in previous presentation)

The major two concerns are **timeliness** and the **cost** of survey data.

Survey data is most likely only available two years after the survey, thus it is old news. In an instant world we need instant data to inform policy makers.

Decision: measure well-being using Big Data specifically Twitter

The beginnings of the GNH.Today Project

Benefits of Big Data (organic data?)

1. The **timeliness** of Big Data is the primary benefit.
2. It also allows governments and social scientists to 'listen' and **capture the needs and well-being concerns** of their citizens, rather than relying on answers to pre-defined explicit questions.
3. It offers social scientists the possibility to **observe people's behaviour** and not just opinions.
4. Do not suffer from **non-response bias**.
5. **Costs** are low and data are available

The choice of Twitter (1/3)

Why Twitter and not Facebook or Instagram

- Twitter is different as it's more open and it's asymmetric. Traditional social networks are mostly restricted to people whom you know.
- Twitter Networks allow you to build open and unexpected relationships with new people where social networks do not.
- With Twitter a more objective “subjective” view.
- Twitter is easier accessible. Facebook and Instagram only sell information to restricted partners and they impose strict restrictions on how the data can be used.

Twitter facts (2020) (2/3)

- A total of **1.3 billion accounts** have been created.
- 79% of accounts are held outside of the U.S.
- There are **500 million tweets sent each day** - 6,000 tweets every second.
- A day's worth of tweets would fill a 10 million page book.
- The top three countries by user count outside the U.S. are:
 - Brazil (27.7 million users),
 - Japan (25.9 million), and
 - Mexico (23.5 million)

Twitter: Challenges, limitations, benefits and corrections (3/3)

- Social media posts are **not representative of the overall population**; older people and children are underrepresented.
- However, corrections in the style of stratified sampling can be made.
- Iacus et al. (2020), Schwartz et al. (2013), among other, successfully fit their biased samples to representative data.
- Twitter **accommodates many groups of people**: individuals, groups of individuals, organisations and media outlets, representing a kind of disaggregated sample, thus giving access to data of a **vast blend of Twitter users**, not found in survey data.

GNH Today

<https://www.gnh.today>

- GNH.Today Project: Launched in 2019
- Measures the real-time well-being (evaluative mood) of a nation
- Evaluative mood – has both a cognitive and an affect dimension (Dan Gilbert 2019, Dan Haybron 2019)
- Project extended to analyse the underpinning emotions of tweets.
- Eight emotions: *joy, anticipation, trust, disgust, anger, surprise, fear and sadness.*
- Initially South Africa, New Zealand and Australia.
- Since 2020: Luxembourg, France, Italy, Spain, Germany, Great Britain, and Belgium.





Method: GNH.Today

<https://www.gnh.today>

AFSTEREO

AUT

GNH Index

Natural Language Processing (NLP) to uncover the sentiment of tweets.

- Sentiment analysis is the process of determining whether a piece of writing (tweet) is **positive, negative or neutral**.
- Sentiment analysis is driven by an algorithm and is better than text analysis since it **helps you understand an entire opinion** and not merely a word from the text.



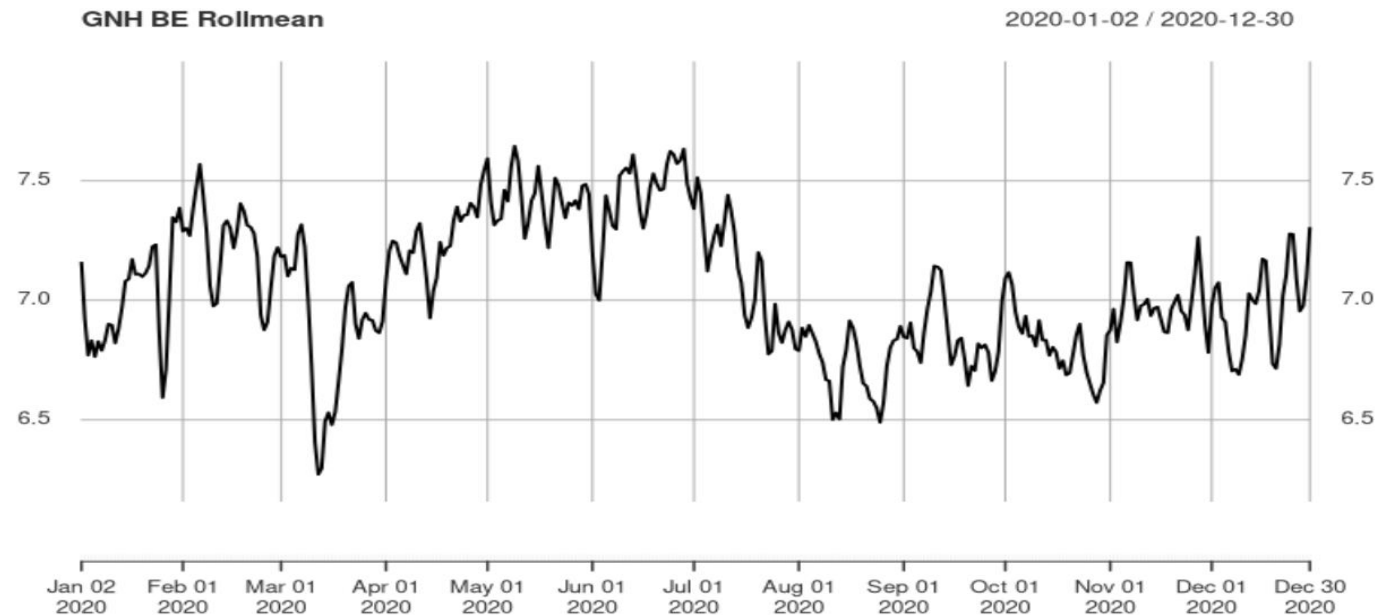
Method: GNH.Today

<https://www.gnh.today>

AFSTEREO

AUT

- After each tweet has been classified, we apply a sentiment balance algorithm to derive a happiness score per hour.
- Scale between 0 😞 and 10, 😊 with 5 being neutral,
- Figure: GNH Belgium





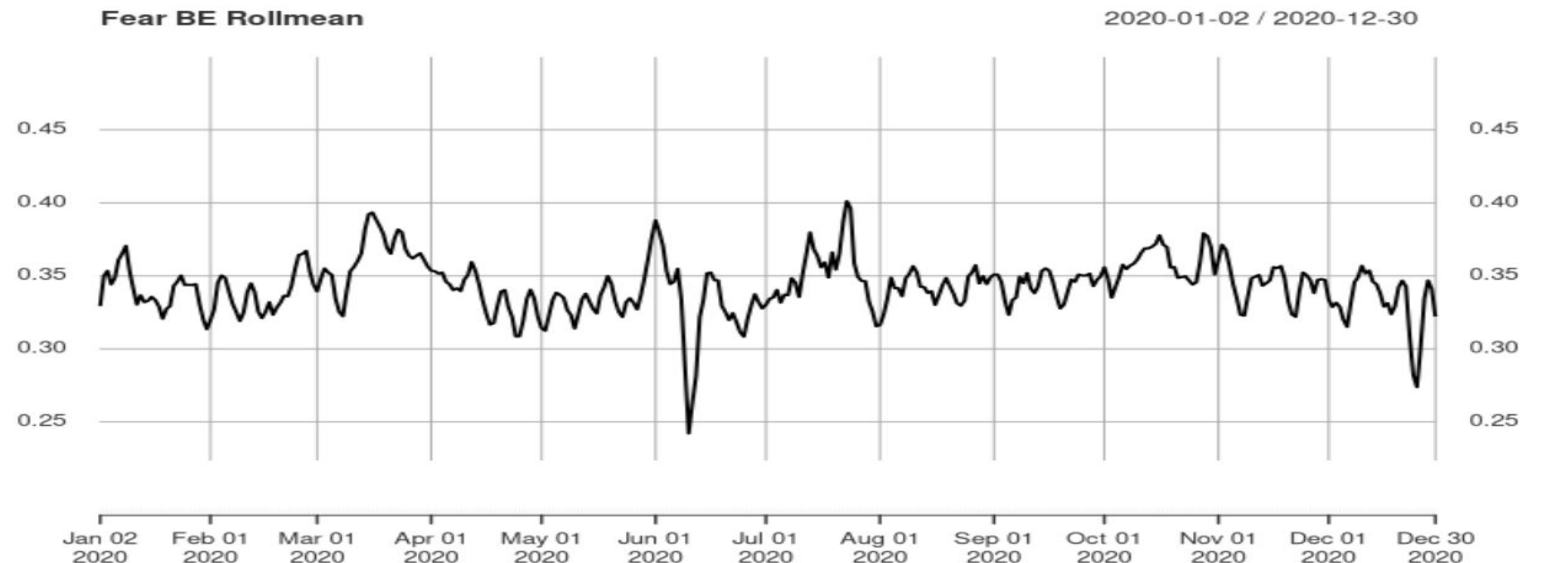
Emotion Time Series

Natural Language Processing – determine underlying emotion

I love dogs they are good companions								
Anger	Fear	Anticipation	Trust	Surprise	Sadness	Joy	Disgust	Trust
0	0	0	1	0	0	4	0	3

To derive emotion time series - calculate mean emotion per day

Figure: Fear - Belgium



Method: GNH.Today

<https://www.gnh.today>

AFSTEREO

AUT



Method: GNH.Today

<https://www.gnh.today>

AFSTEREO

AUT

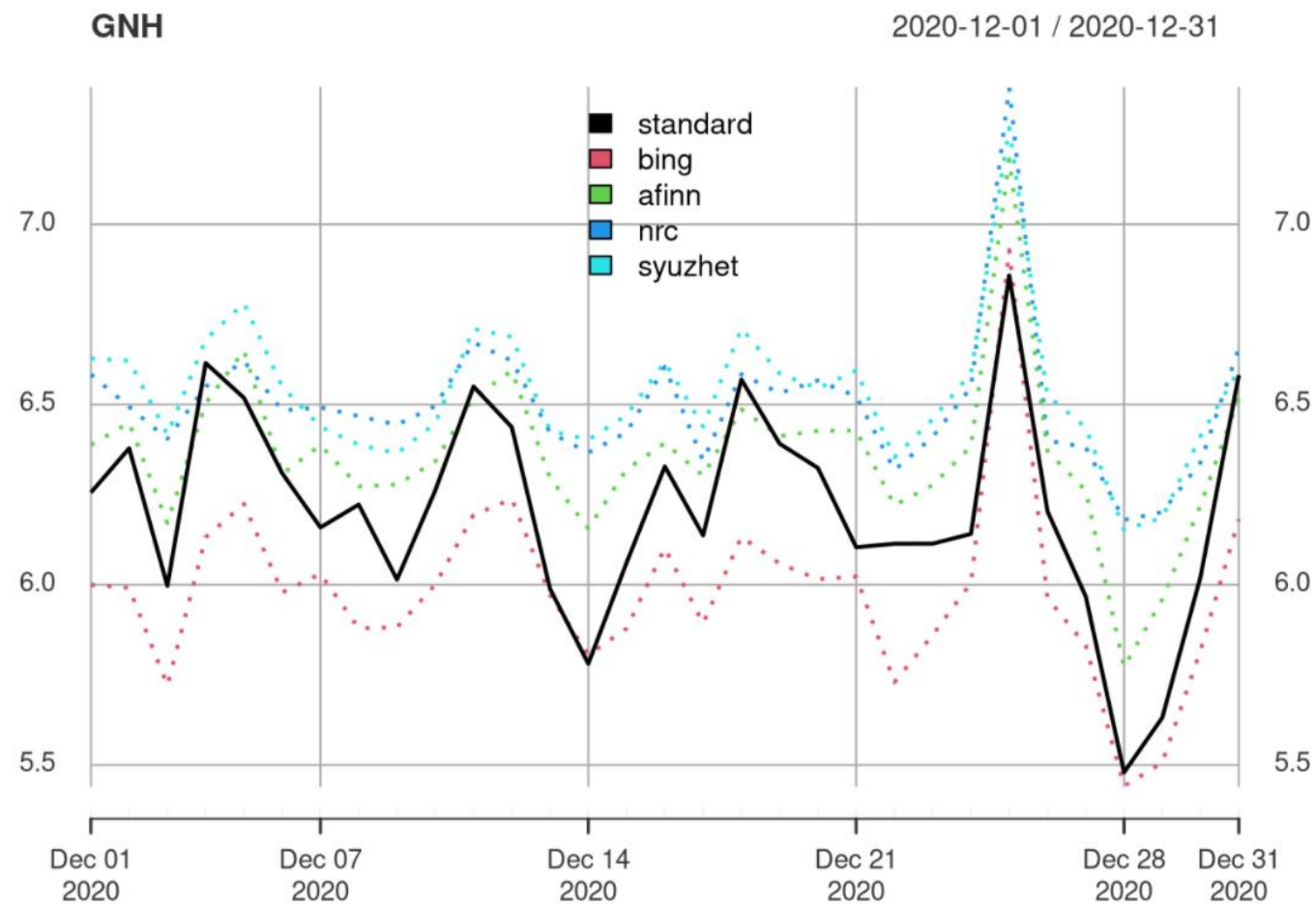
Number of extracted Tweets per day

Country	Tweets per day
BE	6 464
SA	58 131
NZ	4 683
AUS	24 706
NL	18 567
LU	835
IT	27 545
GB	100 867
FR	38 776
ES	57 982
DE	22 427

Validation of the GNH

<https://www.gnh.today>

Using Different Lexicons to calculate GNH



Validation of the GNH

<https://www.gnh.today>

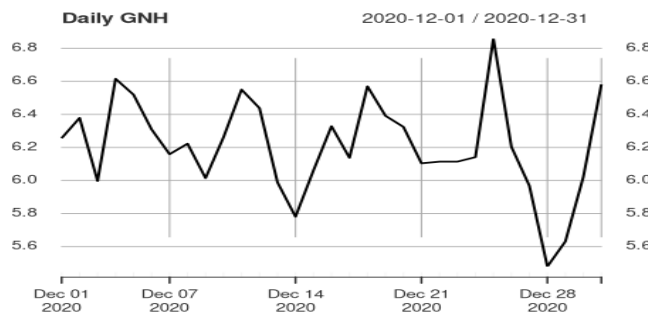
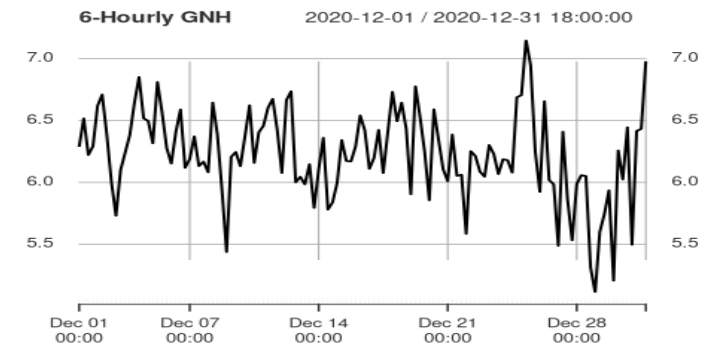
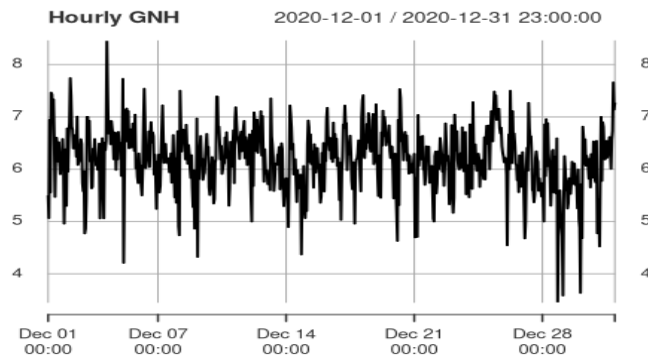
Timescale Invariance

Does **sampling period** significantly influences the GNH values?

We sample the tweets per day, and then calculate the GNH:

(i) per hour, (ii) every 6 hours, and (iii) per day.

If we were to sample tweets per hour and average the result by day, will we obtain the same measure?

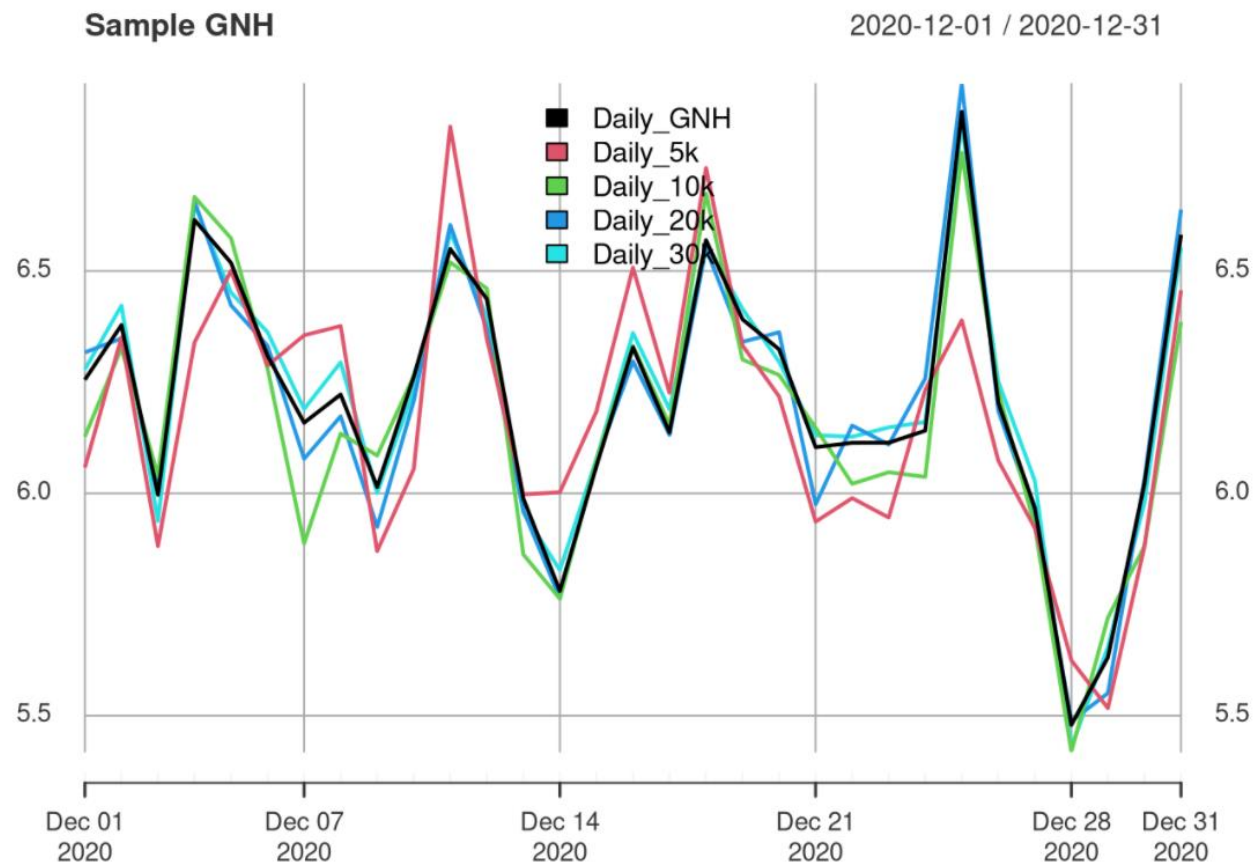


Validation of the GNH

<https://www.gnh.today>

Volume Invariance

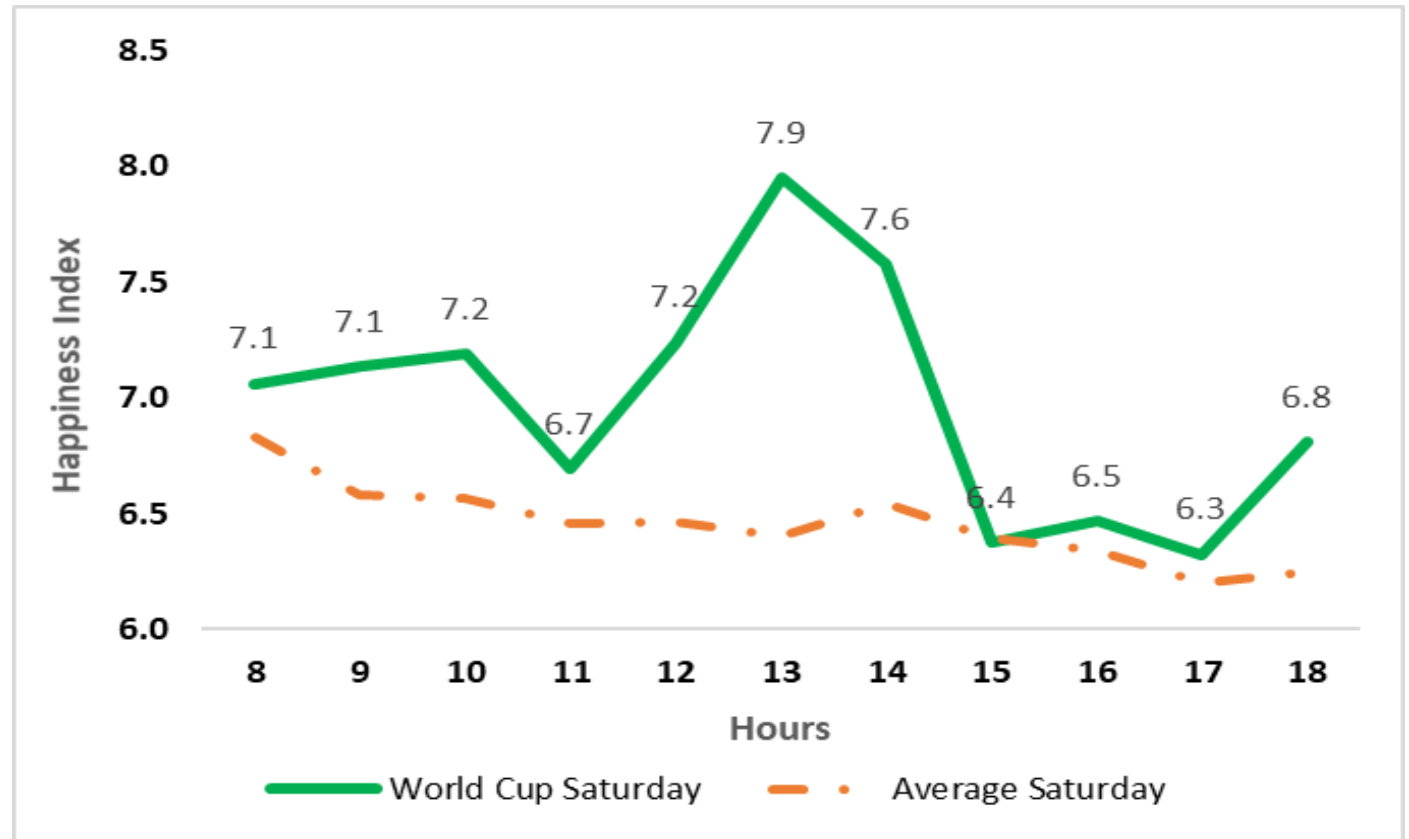
Sample a different number of tweets per day: 5k, 10k, 20k, and 30k



Results of GNH Index and Emotion Time Series

<https://www.gnh.today>

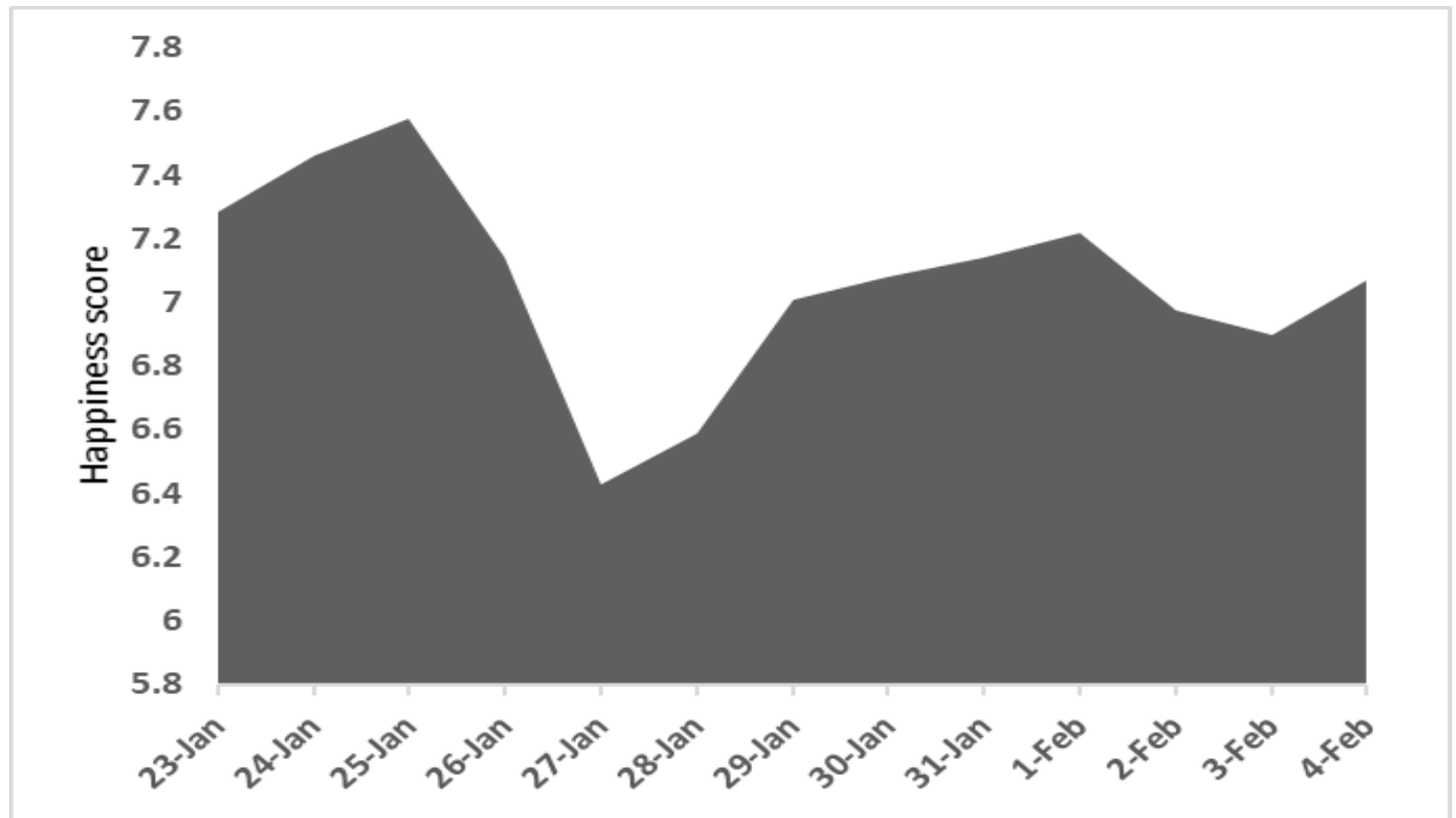
South Africa's intraday (hourly) happiness levels during the Rugby World Cup 2019



Results of GNH Index and Emotion Time Series

<https://www.gnh.today>

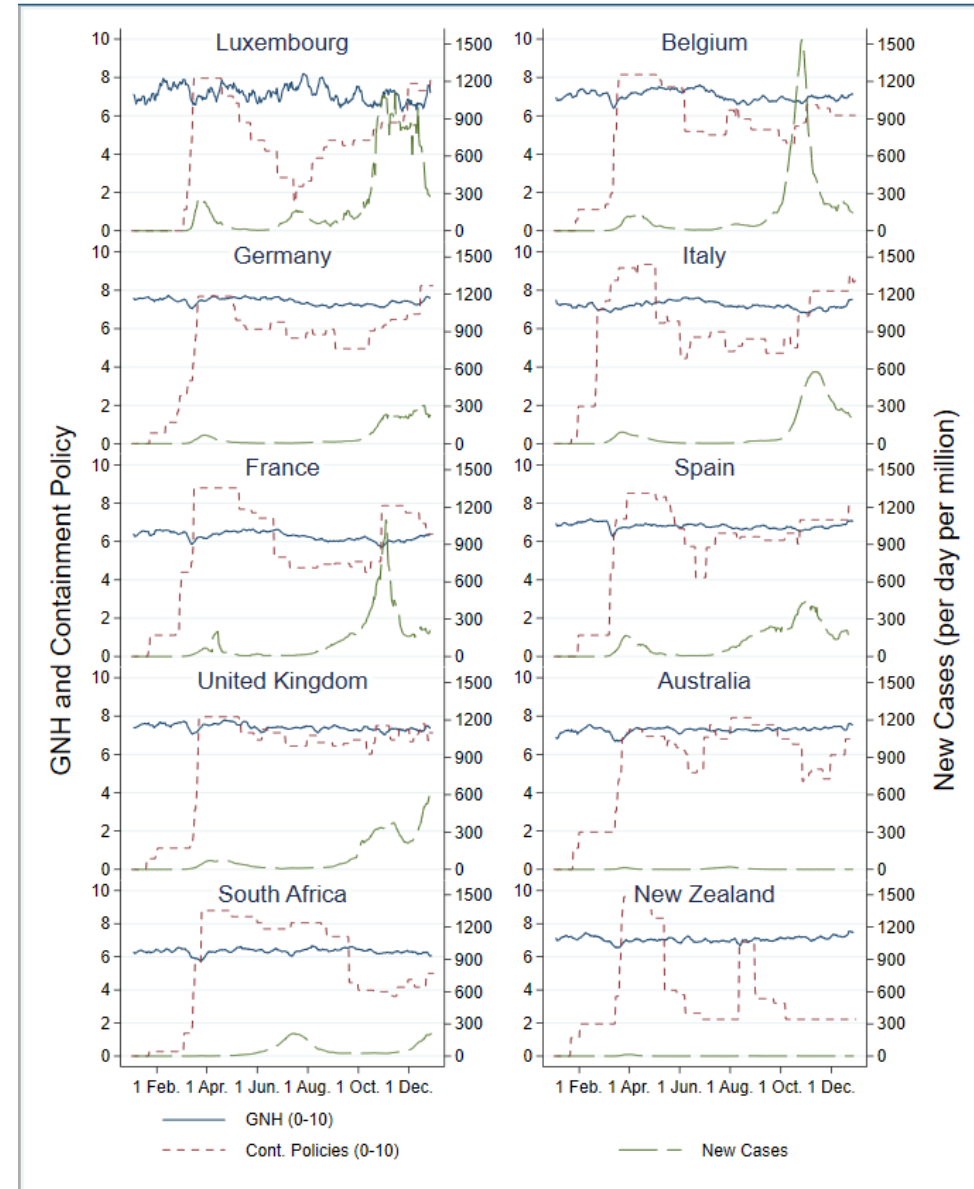
Happiness levels in New Zealand, death of Kobe Bryant



GNH, Government policies to contain COVID and New Cases

Results of GNH Index and Emotion Time Series

<https://www.gnh.today>

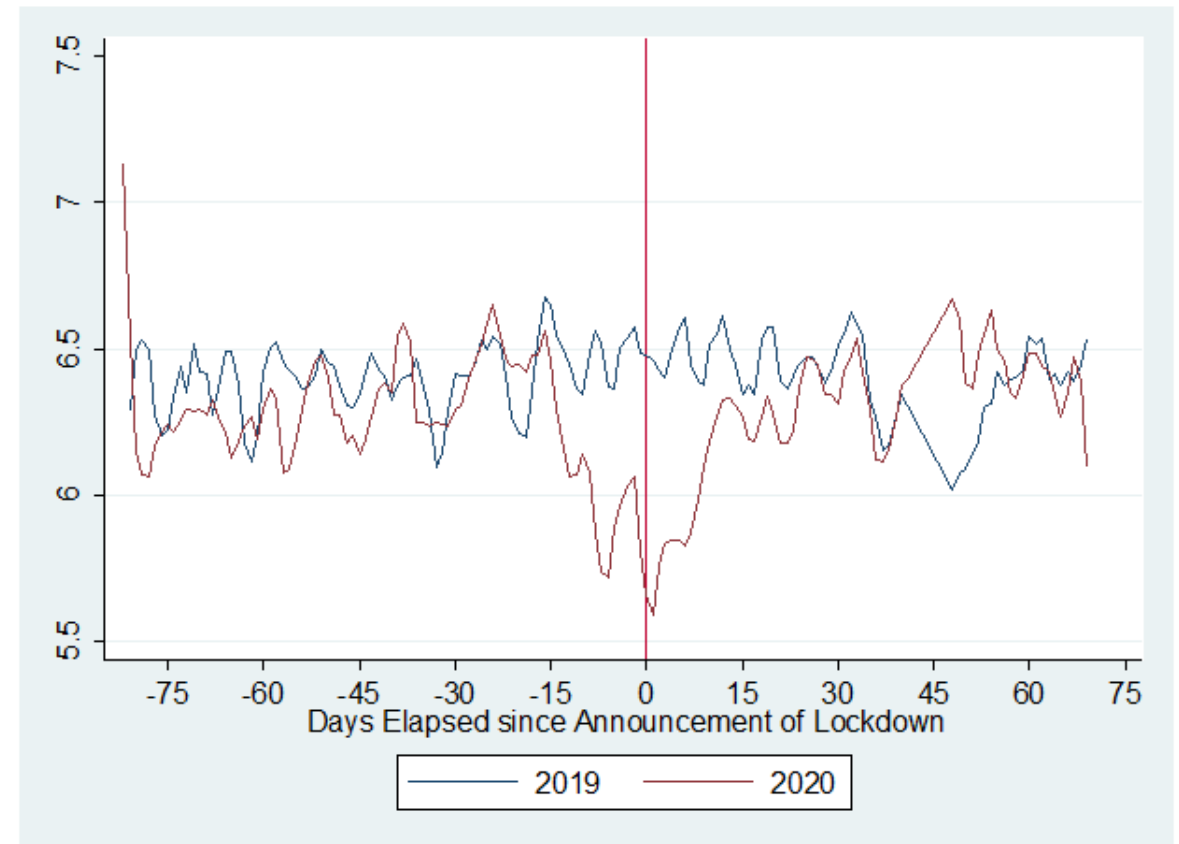


Examples of published research:

The good, the bad and the ugly of lockdowns during Covid-19.

A **difference-in-difference approach** was used to make causal inferences on the lockdown effect on happiness (GNH), and an **OLS estimation** investigated the determinants of happiness after lockdown in South Africa.

The results show that the lockdown had a significant and negative impact on happiness. Determinants of happiness are: stay-at-home orders (positive), lockdown regulations such as a ban on alcohol sales, a fear of becoming unemployed and a greater reliance on social media (negative)



Examples of Published research:

A tale of three countries: How did Covid-19 lockdown impact happiness?

Focused on using the GNH to investigate the *relationship between lockdown and happiness* for South Africa, New Zealand and Australia.

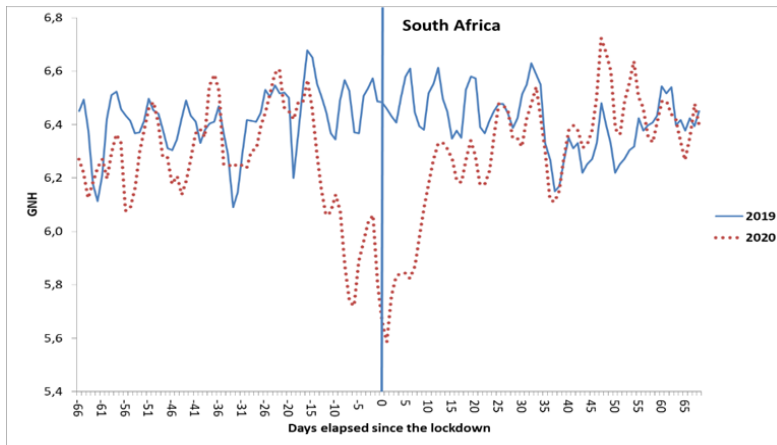
Also investigated the well-being costs *of the different degrees of strictness* of these countries' lockdown regulations

Examples of Published research:

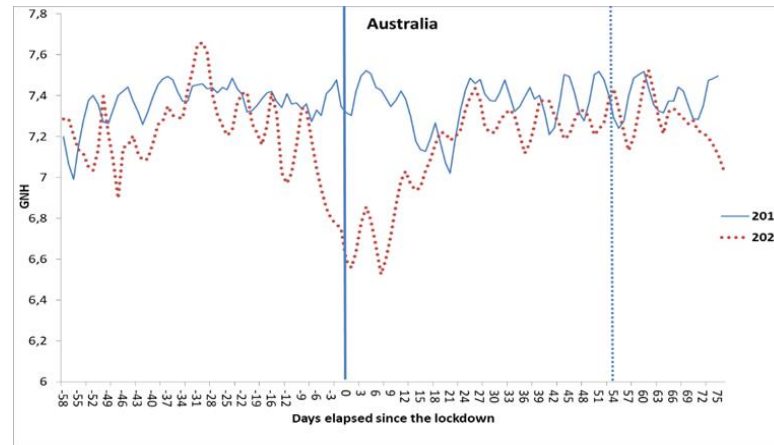
A tale of three countries: How did Covid-19 lockdown impact happiness?

The main results show robust evidence of a **negative relationship between the lockdown regulations and happiness**, notwithstanding the diversity in characteristics and lockdown regulations of the countries included in our sample. Stricter regulations have higher happiness costs.

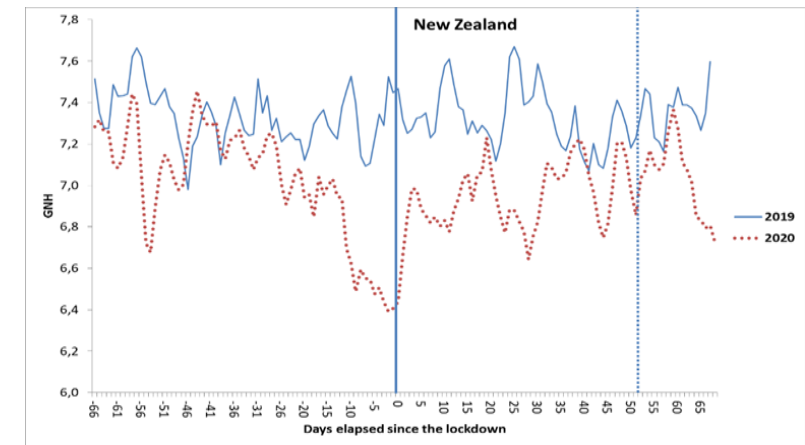
South Africa



Australia



New Zealand



Examples of Published research:

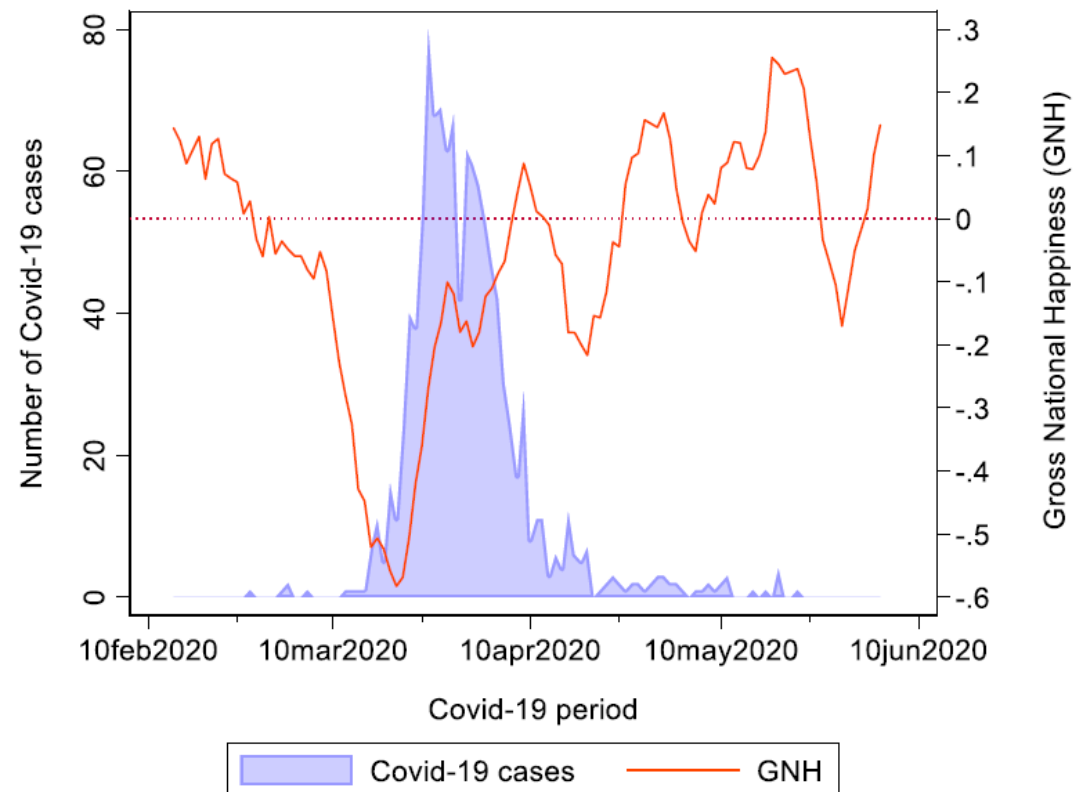
The impact of exogenous shocks on wellbeing. New Zealanders' reaction to Covid-19.

- We explored the sensitivity of a national well-being index (GNH) to the arrival and passage of the COVID-19 epidemic in a small, physically separate, economy in the South Pacific, New Zealand.
- We found **distinct reactions to the epidemic**: a marked drop in national wellbeing.
- The drop in wellbeing was short-lived however due in large part to the rapid and effective response by the New Zealand government.
- In well-being terms the country returned to 'normal' following a rapid and severe lockdown designed to limit domestic transmission of the virus.

Examples of Published research:

The impact of exogenous shocks on wellbeing. New Zealanders' reaction to Covid-19.

GNH and COVID-19 cases – 14 February to 14 June 2020



Published Research Papers

Journal articles:

- Greyling, T., Rossouw, S., & Adhikari, T. 2021. The good, the bad and the ugly of lockdowns during COVID-19. *PLoS ONE*, 16(1), e0245546.
- Rossouw, S., Greyling, T. & Adhikari, T. 2021. Happiness-lost: Was the decision to implement lockdown the correct one? Forthcoming *South African Journal of Economic and Management Sciences*
- Greyling, T., Rossouw, S. & Adhikari, T. 2021. A tale of three countries: What is the relationship between Covid-19, lockdown, and happiness? Forthcoming in *South African Journal of Economics*. 10.1111/saje.12284
- Morrison, P., Rossouw, S. & Greyling, T. 2020. The impact of exogenous shocks on wellbeing. New Zealanders' reaction to COVID-19. Forthcoming. *Applied Research in Quality of Life*
- Rossouw, S. & Greyling, T. 2020. Big Data and Happiness. Invited chapter for the [Handbook of Labor, Human Resources and Population Economics](#). Edited by Klaus F. Zimmermann.

Published Working Papers

Working Papers:

- Sarracino, F., Greyling, T., O'Connor, K., Peroni, C. & Rossouw, S. 2020. Six months of well-being data from Twitter: levels, changes and validity of GNH in ten countries. FNR Discussion Paper.
- Rossouw, S., Greyling, T., Adhikari, T. & Morrison, P. 2020: Markov switching models for happiness during a pandemic: The New-Zealand experience, GLO Discussion Paper, No. 573, Global Labor Organization (GLO), Essen

Awards

Nominated for the prestigious **Human Science Research Council Award** for contributing positively to the Social Sciences and Humanities and COVID research in 2020

Limitations of GNH

<https://www.gnh.today>

- We lack **demographic information**, such as gender age etc.
- We only have the geographical location.
- Our results are interpreted as the mean GNH for the geographical area.

GNH.Today Future

<https://www.gnh.today>

- To further **validate** the data
- Extend to **other social media platforms and include emoji's in analyses**
- Collaborate with **other institutions**
- Find ways to make data **more representative**
- **Combine** it with survey data

<https://www.gnh.today>

A hand is shown reaching out from the left side of the frame towards a glowing, wireframe globe of the Earth. The globe is composed of a network of white lines and dots, representing a digital or networked world. The background is a dark blue, blurred cityscape at night, with lights from buildings visible. The overall aesthetic is futuristic and technological.

Thank you very much for listening.