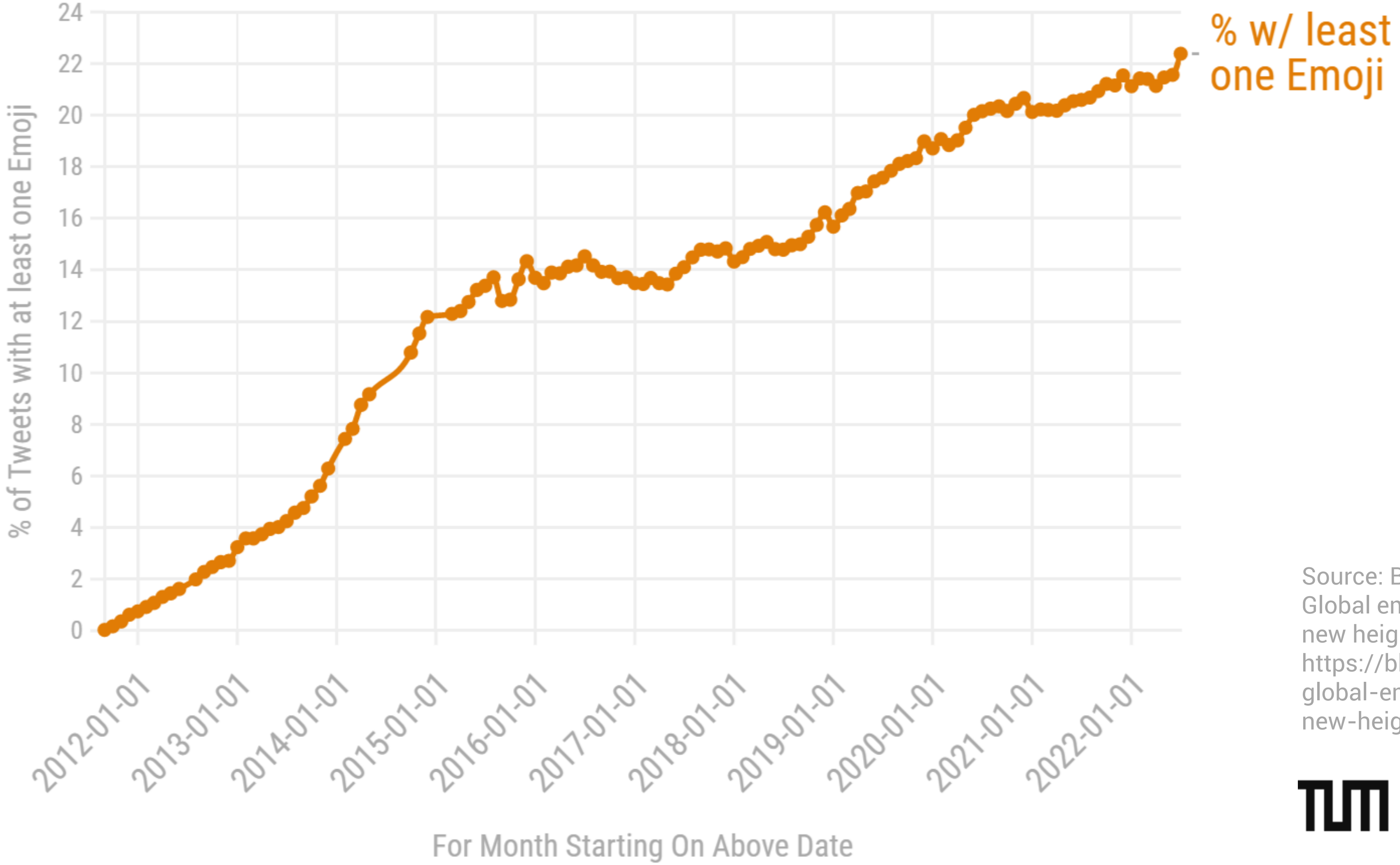
Emojis As Indicators Of Spatial-Temporal-Thematic Developments In Geo-Social Media

Samantha Levi

Supervisors: Dr. Ing. Eva Hauthal and Sagnik Mukherjee (TUD)

Reviewer: Dr. Frank Ostermann (UT)

Emojis As Indicators Of Spatial-Temporal-Thematic Developments In Geo-Social Media



Source: Broni, K. (2022, 7). Global emoji use reaches new heights. Retrieved from https://blog.emojipedia.org/global-emoji-use-reaches-new-heights

Emojis As Indicators Of Spatial-Temporal-Thematic Developments In **Geo-Social Media**

State of the Art

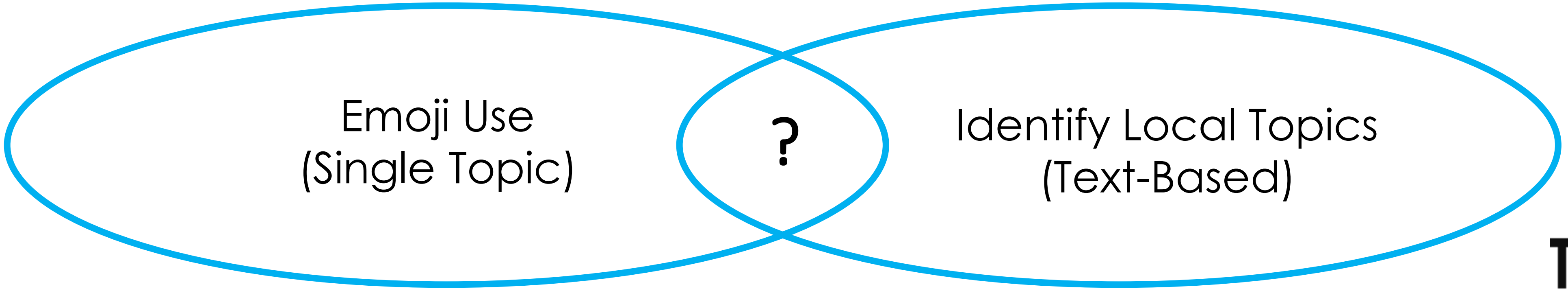
Social Media as a Data Source

- Generated by billions of users
- Provide insights into ideas and topics that attract large numbers of users

Emojis vs. Text

- Emojis increasingly used to convey meaning, not chosen arbitrarily
- Allow for language-independent research
- Circumvent obstacles common in text-based approaches like Natural Language Processing (slang, spelling and grammatical errors)

Existing Research Gap



Objective

Determine whether emojis can be used to identify relevant topics and their spatial-temporal evolution in a non-topic-specific dataset

The Dataset

~4 MILLION POSTS

GEOTAGGED WITHIN EUROPE

DURING 2020*

AT LEAST ONE EMOJI AND ONE HASHTAG

* No data available for November 2020

Image source: https://about.twitter.com/en/who-we-are/brand-toolkit

Research Questions

RQ1

Does the usage of emojis change over time and space?

RQ2

Do changes in emoji usage have thematic connections?

RQ3

How can these spatial/temporal/ thematic developments be visualized?

Methodology

Conceptual Framework: 4 Facet Structure

Dunkel et al. (2019)

TEMPORAL

(WHEN)

Time of post creation

SPATIAL

(WHERE)

Location collected at time of post creation

TOPICAL

(WHAT)

Topics discussed in the post

SOCIAL

(WHO)

User who created the post

Conceptual Framework: 4 Facet Structure

Dunkel et al. (2019)

TEMPORAL
(WHEN)

Time of post creation

post_date

SPATIAL
(WHERE)

Location collected at time of post creation

TOPICAL
(WHAT)

Topics discussed in the post

SOCIAL
(WHO)

User who created the post

Conceptual Framework: 4 Facet Structure

Dunkel et al. (2019)

TEMPORAL (WHEN)

Time of post creation

post_date

SPATIAL (WHERE)

Location collected at time of post creation

Latitude and Longitude

TOPICAL (WHAT)

Topics discussed in the post

SOCIAL (WHO)

User who created the post

Conceptual Framework: 4 Facet Structure

Dunkel et al. (2019)

TEMPORAL (WHEN)

Time of post creation

post_date

SPATIAL (WHERE)

Location collected at time of post creation

Latitude and Longitude

TOPICAL (WHAT)

Topics discussed in the post

Emoji + Hashtag

SOCIAL (WHO)

User who created the post

Conceptual Framework: 4 Facet Structure

Dunkel et al. (2019)

TEMPORAL (WHEN)	SPATIAL (WHERE)	TOPICAL (WHAT)	SOCIAL (WHO)
Time of post creation	Location collected at time of post creation	Topics discussed in the post	User who created the post
post_date	Latitude and Longitude	Emoji + Hashtag	User-IDs…

Conceptual Framework: 4 Facet Structure

Dunkel et al. (2019)

TEMPORAL (WHEN)	SPATIAL (WHERE)	TOPICAL (WHAT)	SOCIAL (WHO)
Time of post creation	Location collected at time of post creation	Topics discussed in the post	User who created the post
post_date	Latitude and Longitude	Emoji + Hashtag	User-IDs…

Social Facet: Privacy-Awareness

HyperLogLog (HLL) Algorithm

- Cardinality estimator, MurMur hash function
- Privacy is a side-effect, comes with other perks (count user days, user count, post count)

Cryptographic Hashing

- Data encryption – pseudonymize user information
- Weak measure by itself, but strengthens the effectiveness of other measures

Spatial Data Aggregation

- Geohashing possible at several levels
- Aggregation level 4 (used in this analysis) “snaps” to a 20 km spatial resolution

Visualizations at Coarse Resolution

- 100 by 100 kilometer grid used for resulting maps to avoid visualizing precise user locations

Social Facet: Privacy-Awareness

HyperLogLog (HLL) Algorithm

- Cardinality estimator, MurMur hash function
- Privacy is a side-effect, comes with other perks (count user days, user count, post count)

Cryptographic Hashing

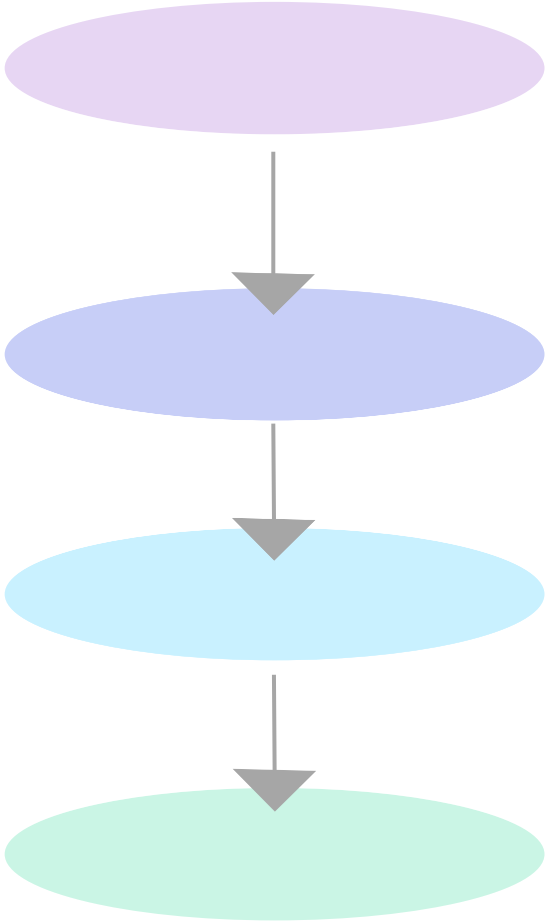
- Data encryption – pseudonymize user information
- Weak measure by itself, but strengthens the effectiveness of other measures

Spatial Data Aggregation

- Geohashing possible at several levels
- Aggregation level 4 (used in this analysis) “snaps” to a 20 km spatial resolution

Visualizations at Coarse Resolution

- 100 by 100 kilometer grid used for resulting maps to avoid visualizing precise user locations



Conceptual Framework: 4 Facet Structure

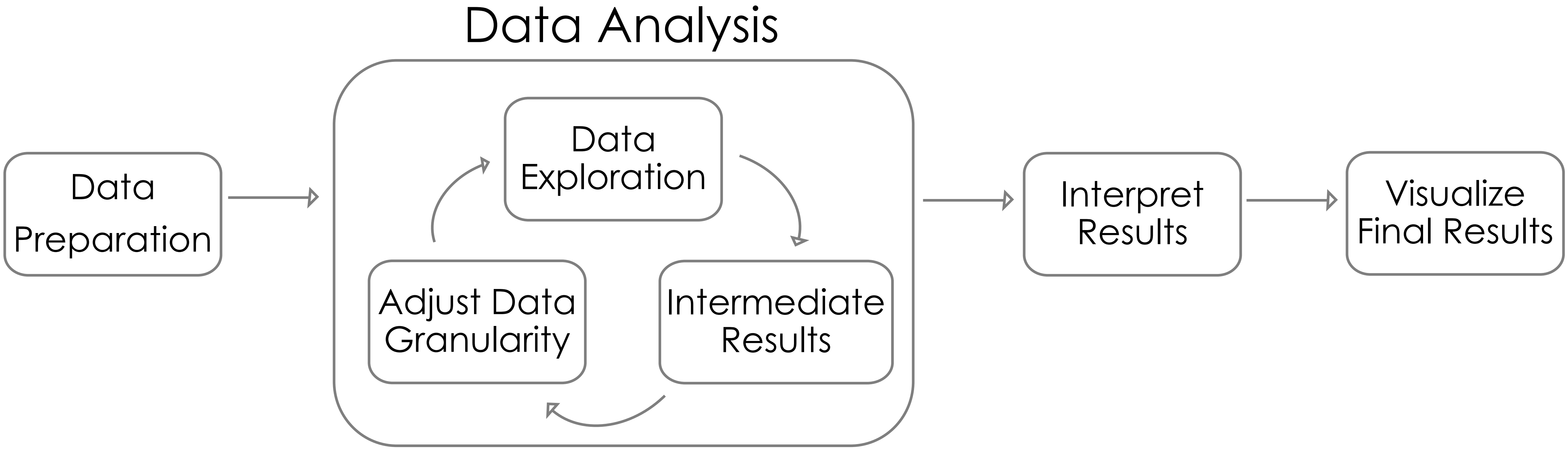
Dunkel et al. (2019)

- **TEMPORAL (WHEN)**
- **SPATIAL (WHERE)**
- **TOPICAL (WHAT)**
- **SOCIAL (WHO)**

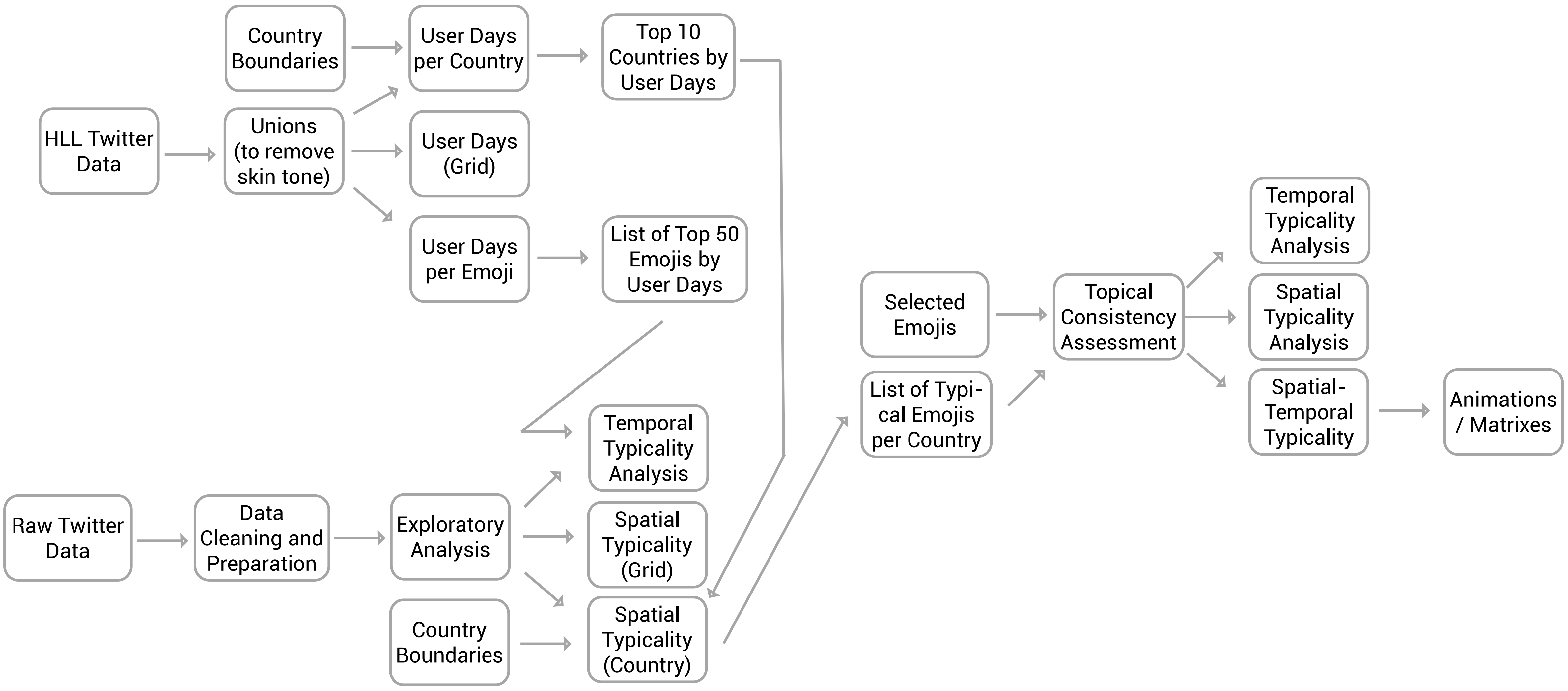
RAW/ORIGINAL DATA

HLL DATA

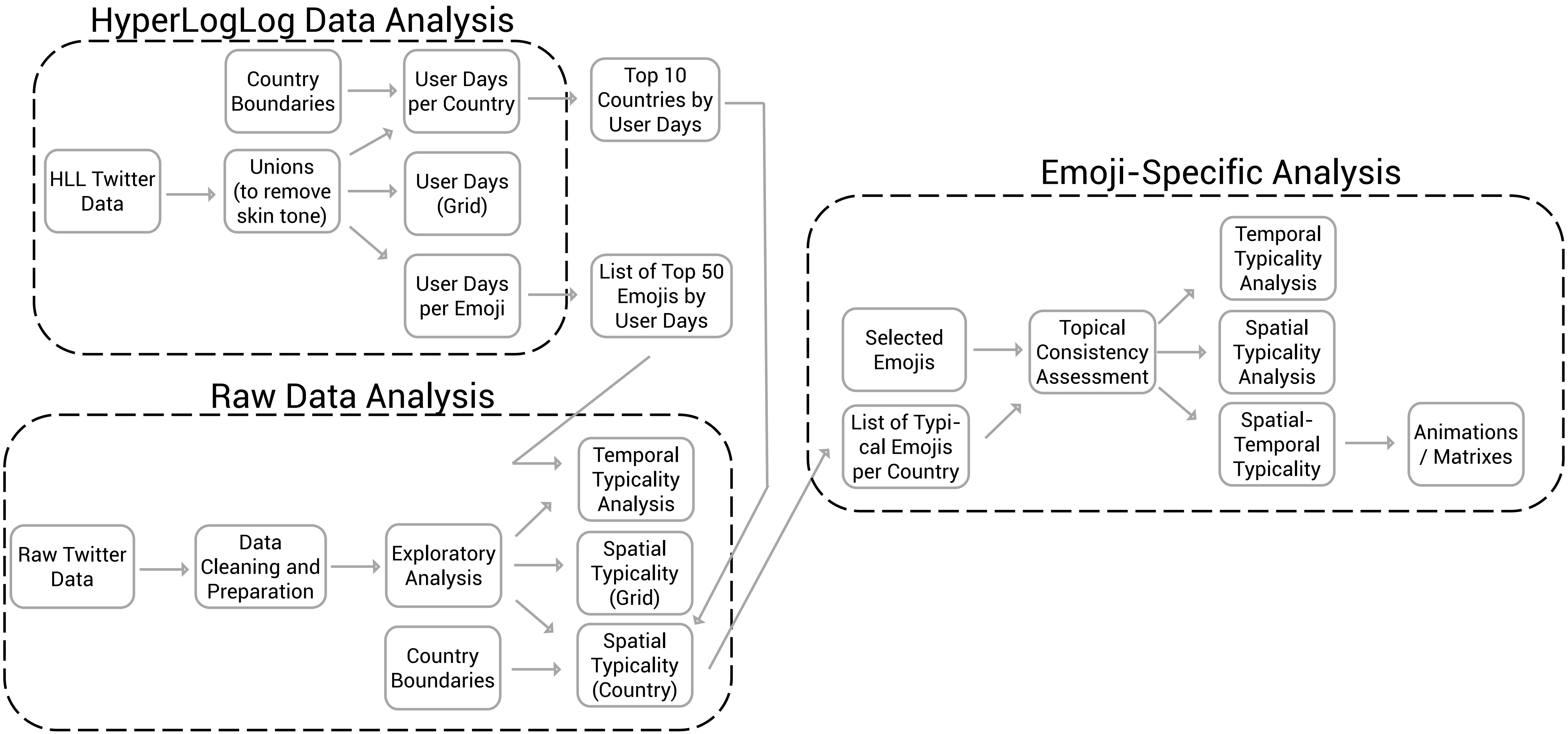
Workflow: Proposed



Workflow: Actual



Workflow: Actual



HyperLogLog
Data Analysis

Benefits/Uses of HyperLogLog (HLL)

Privacy-Awareness

- Investigate social facet of data with increased user privacy
- Sensitive information (like user IDs) is never gathered from remote server

Benefits/Uses of HyperLogLog (HLL)

Privacy-Awareness

- Investigate social facet of data with increased user privacy
- Sensitive information (like user IDs) is never gathered from remote server

Calculate Cardinality, Unions, Intersections

- Efficiently compute number of distinct users, posts, and user days* with error of 2-5%
- Unions allow for joining of emojis with multiple skin tone variations

💪 ∪ 💪🏻 ∪ 💪🏽 ∪ 💪🏾 ∪ 💪🏼 ∪ 💪🏿

⬇

💪

Benefits/Uses of HyperLogLog (HLL)

Privacy-Awareness

- Investigate social facet of data with increased user privacy
- Sensitive information (like user IDs) is never gathered from remote server

Calculate Cardinality, Unions, Intersections

- Efficiently compute number of distinct users, posts, and user days* with error of 2-5%
- Unions allow for joining of emojis with multiple skin tone variations

Reduce Reliance on Absolute Frequency

- Narrow down scope of calculations for raw data analysis by finding top emojis and countries by user days

Benefits/Uses of HyperLogLog (HLL)

Privacy-Awareness

- Investigate social facet of data with increased user privacy
- Sensitive information (like user IDs) is never gathered from remote server

Calculate Cardinality, Unions, Intersections

- Efficiently compute number of distinct users, posts, and user days* with error of 2-5%
- Unions allow for joining of emojis with multiple skin tone variations

Reduce Reliance on Absolute Frequency

- Narrow down scope of calculations for raw data analysis by finding top emojis and countries by user days

Identify Emojis Often Used by Bots/Hyper-Active Users

- Calculate the difference between post count and distinct user days, find which emojis are used frequently by only a few users

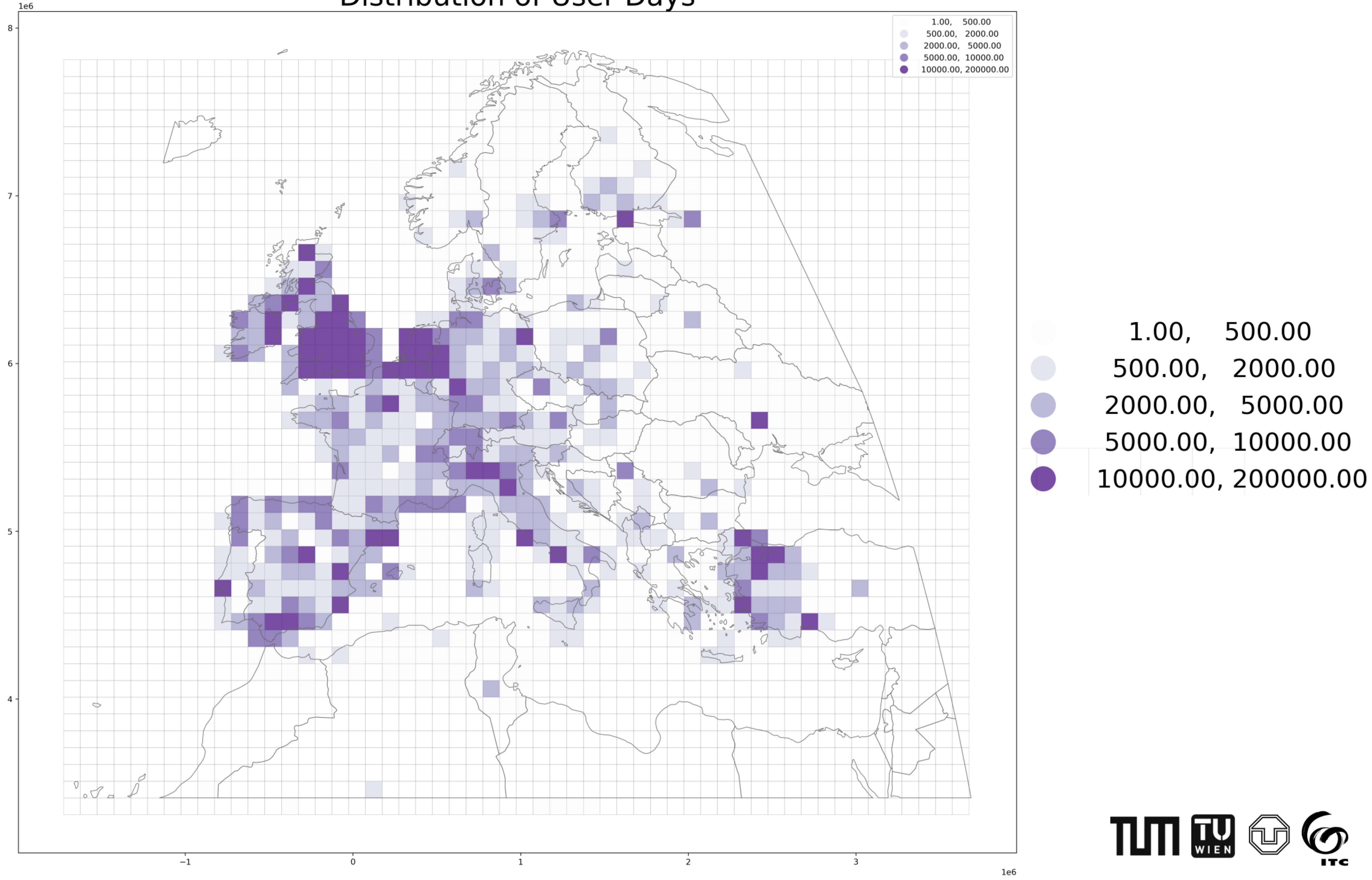
User Days

"[U]ser-days are defined as the total number of days, across all users, that each person took at least one photograph within each site" – Wood et. al (2013)

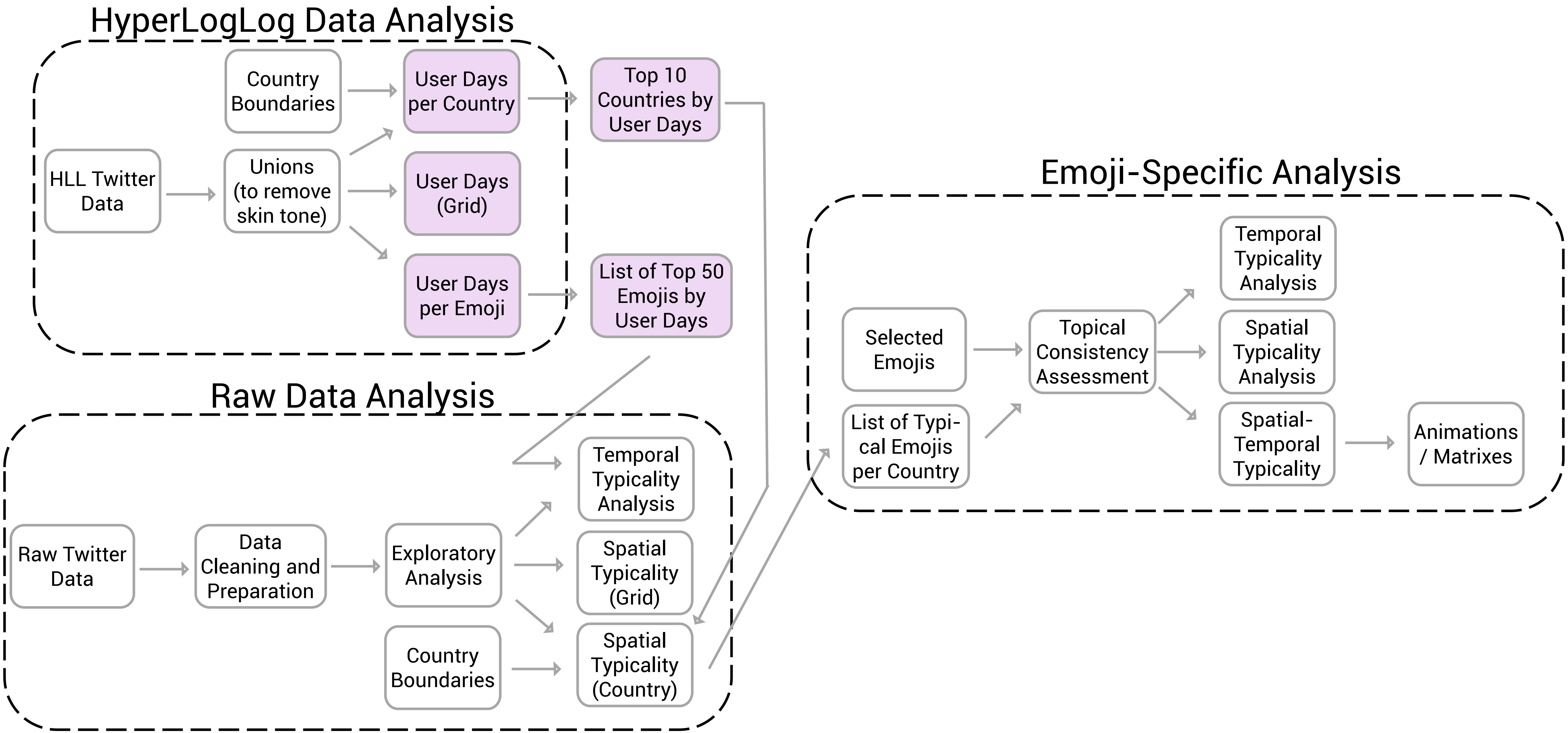
In this context:

The number of days, across all users, that a distinct user posted at least one tweet within the study area

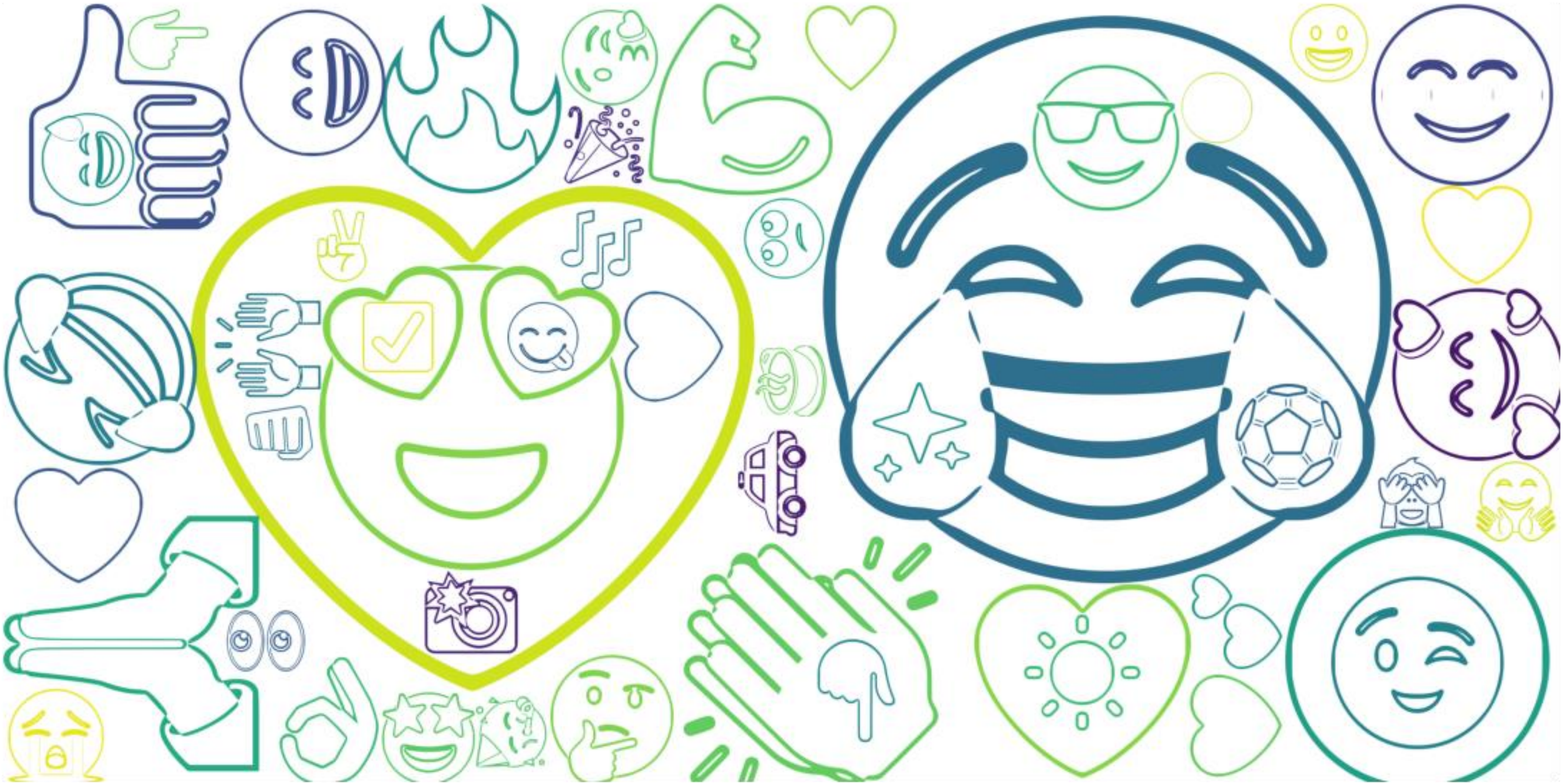
Distribution of User Days

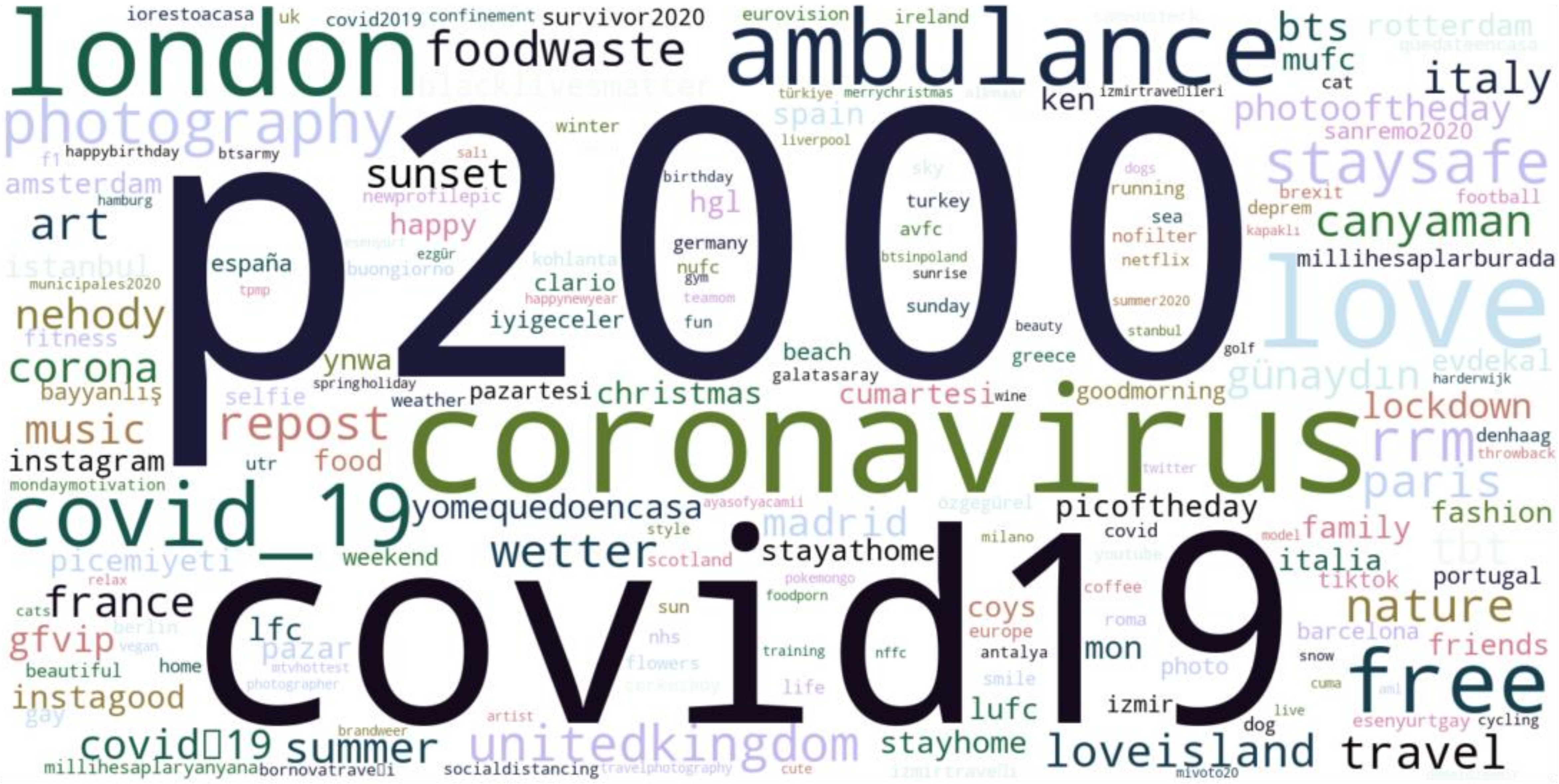


Workflow: Actual



Raw Data
Exploratory Analysis




p2000
covid19
coronavirus
london
ambulance
covid_19
love
photography
free
foodwaste
staysafe
italy
bts
mufc
photooftheday
canyaman
unitedkingdom
loveisland
travel
summer
repost
wetter
stayathome
yomequedoencasa
picoftheday
christmas
lockdown
rrm
paris
nature
corona
nehody
music
instagram
art
sunset
france
gfvip
lfc
instagood
stayhome
lufc
mon
italia
fashion
family
portugal
friends
barcelona
tiktok
sanremo2020
millihesaplarburada
günaydın
goodmorning
cumartesi
pazartesi
iyigeceler
clario
españa
amsterdam
happy
hgl
spain
ken
survivor2020
iorestoacasa
uk
covid2019
confinement
eurovision
ireland
winter
happybirthday
btsarmy
birthday
turkey
germany
avfc
running
brexit
deprem
football
sea
nofilter
netflix
btsinpoland
sunrise
sunday
beauty
golf
stanbul
summer2020
greece
beach
galatasaray
ynwa
springholiday
weather
selfie
bayyanlış
fitness
municipales2020
hamburg
nufc
fun
happynewyear
teamom
buongiorno
ezgür
mondaymotivation
utr
food
picemiyeti
weekend
scotland
style
madrid
milano
covid
coys
europe
coffee
antalya
izmir
sun
foodporn
training
nffc
beautiful
home
cats
covid19
millihesaplaryanyana
socialdistancing
travelphotography
cute
artist
brandweer
dog
cycling
esenyurtgay
snow
denhaag
throwback
harderwijk
evdekal
cat
liverpool
merrychristmas
türkiye
wine
pazar
life
photo
roma
nhs
smile
flowers
dogs
twitter
model
live
mlvoto20

Most Frequently Used Emojis Per Month

January	February	March	April	May	June	July	August	September	October	December
😂	❤️	😂	😂	❤️	❤️	❤️	❤️	😂	😂	❤️
❤️	😂	❤️	❤️	😂	😂	😂	😂	❤️	❤️	🎄
😍	😍	😍	😍	😍	😍	😍	😍	🔴	🔴	😂
🔴	🔴	🔴	👏	🔴	🔴	🔴	🔴	😍	😍	😍
🤣	💙	👏	🙏	💙	💙	💙	💙	💙	💙	🔴
💙	🤣	🤣	🤣	🤣	🤣	🔥	🔥	🤣	🤣	🤣
🔥	🔥	🙏	💙	😊	😊	🤣	😎	🔥	🔥	💙
👍	👍	💙	👍	👍	👍	😎	🤣	👍	👍	🔥
😊	😊	👍	😊	🖤	🔥	😊	☀️	😊	😊	🥰
😜	🖤	😊	🔴	👏	😎	👍	😊	🥰	💛	🖤

Most Frequently Used Emojis Per Country

United Kingdom	Spain	France	Germany	Italy	Turkey	Netherlands	Belgium	Switzerland	Austria
😂	❤️	❤️	❤️	❤️	❤️	🔴	❤️	❤️	❤️
❤️	😍	😂	😂	😂	🙏	🚓	😂	😂	😍
👍	👏	😍	☔	😍	😂	🚑	😍	😍	😂
👏	💪	🙏	😍	💪	💙	🚒	😉	💙	👍
😍	😂	👍	👍	💙	😍	❤️	🔥	👍	🙏
🤣	💙	🔥	🙏	🤣	💛	😂	😎	🙏	💪
💙	🤣	😉	😉	♥️	👏	👍	👍	😎	😎
😊	👉	💪	😎	🔥	😊	💪	🙏	💪	🥰
🙏	👇	😭	💪	🙏	🔥	😉	💪	☀️	😊
🙌	🔥	🤣	😊	🖤	💋	😍	🖤	😉	💙

The Problem with Absolute Frequency: Bots

United Kingdom	Spain	France	Germany	Italy	Turkey	Netherlands	Belgium	Switzerland	Austria
😂	❤️	❤️	❤️	❤️	❤️	🔴	❤️	❤️	❤️
❤️	😍	😂	😂	😂	🙏	🚓	😂	😂	😍
👍	👏	😍	☔	😍	😂	🚑	😍	😍	😂
👏	💪	🙏	😍	💪	💙	🚒	😉	💙	👍
😍	😂	👍	👍	💙	😍	❤️	🔥	👍	🙏
🤣	💙	🔥	🙏	🤣	💛	😂	😎	🙏	💪
💙	🤣	😉	😉	♥️	👏	👍	👍	😎	😎
😊	👉	💪	😎	🔥	😊	💪	🙏	💪	🥰
🙏	👇	😭	💪	🙏	🔥	😉	💪	🌞	😊
🙌	🔥	🤣	😊	🖤	💋	😍	🖤	😉	💙

Typicality

Hauthal et al. (2021)

Meaning

For a **designated subset of a larger dataset,** how 'characteristic' or 'typical' of that subset is a given emoji?

Calculation

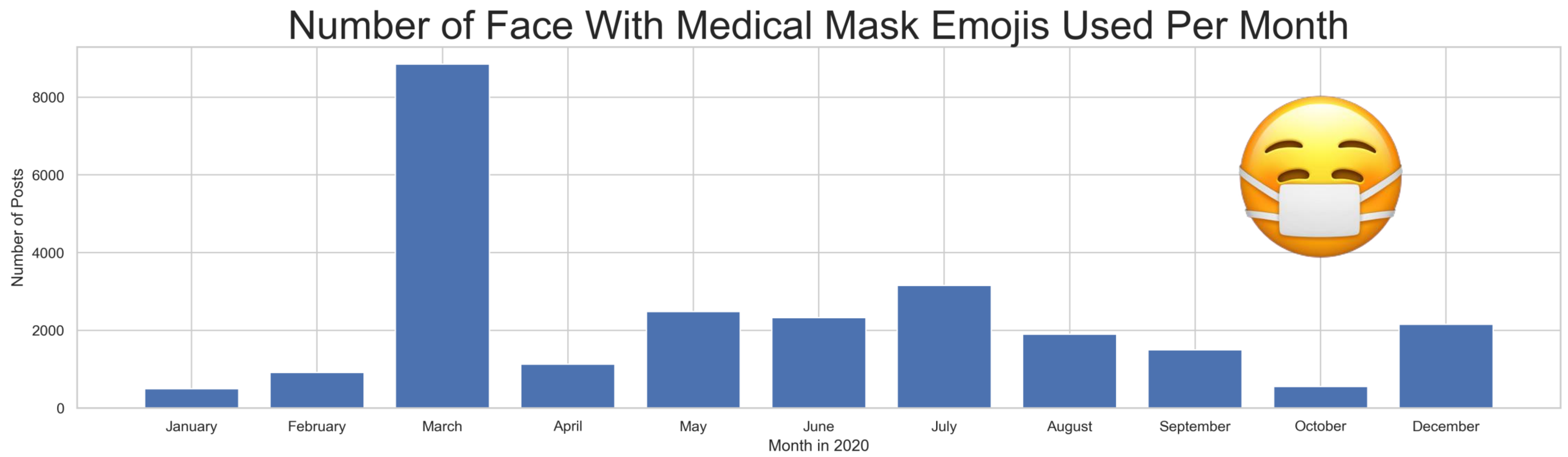
$$\text{Typicality} = \frac{ns/Ns - nt/Nt}{nt/Nt} = \frac{\text{Rel. freq. within the subset} - \text{rel. freq. within the total dataset}}{\text{rel. freq. within the total dataset}}$$

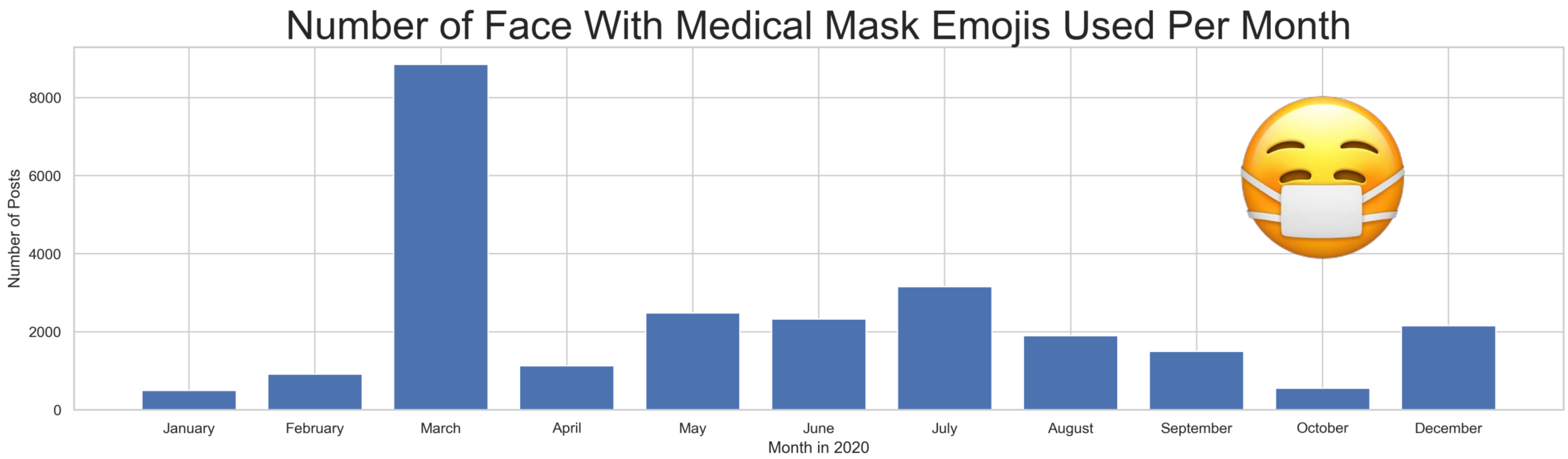
Interpretation

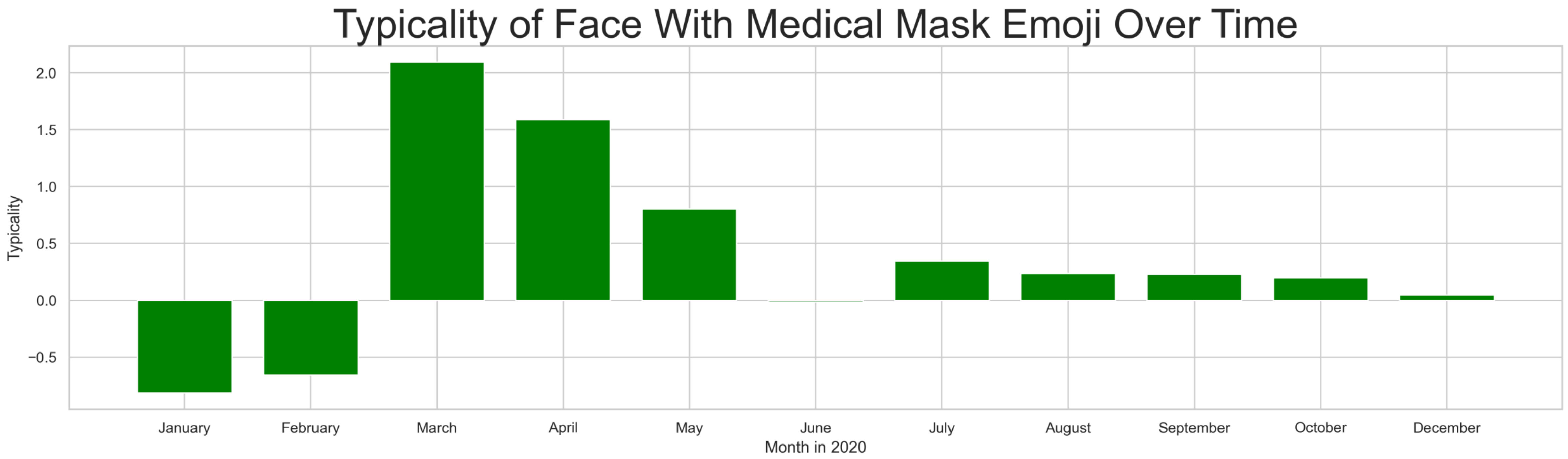
Positive: an occurrence is **typical** for the subset

Negative: an occurrence is **atypical** for a subset

Temporal Typicality


Number of Face With Medical Mask Emojis Used Per Month
Number of Posts
0
2000
4000
6000
8000
January
February
March
April
May
June
July
August
September
October
December
Month in 2020

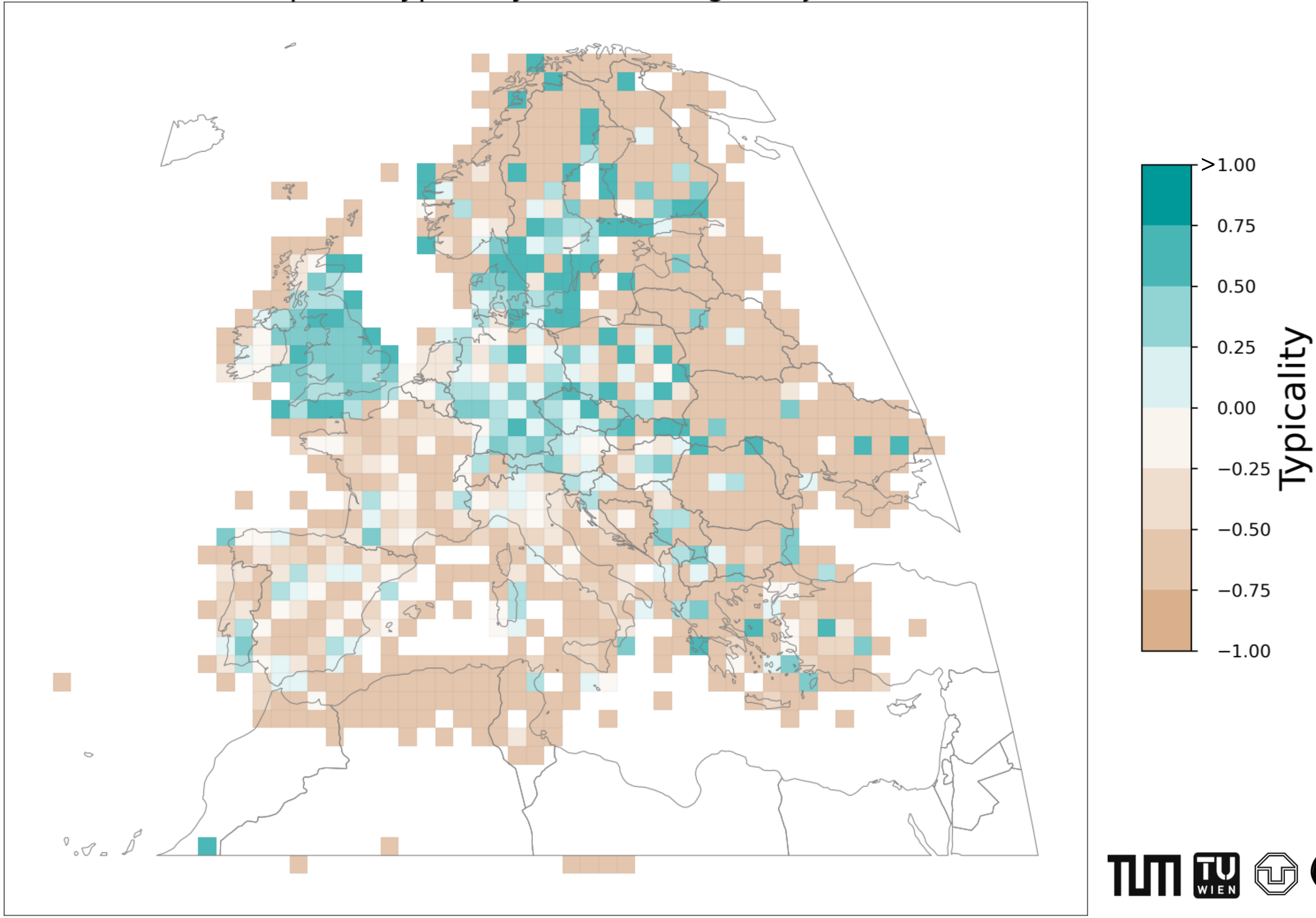

Number of Face With Medical Mask Emojis Used Per Month
Number of Posts
0
2000
4000
6000
8000
January
February
March
April
May
June
July
August
September
October
December
Month in 2020


Typicality of Face With Medical Mask Emoji Over Time
Typicality
−0.5
0.0
0.5
1.0
1.5
2.0
January
February
March
April
May
June
July
August
September
October
December
Month in 2020

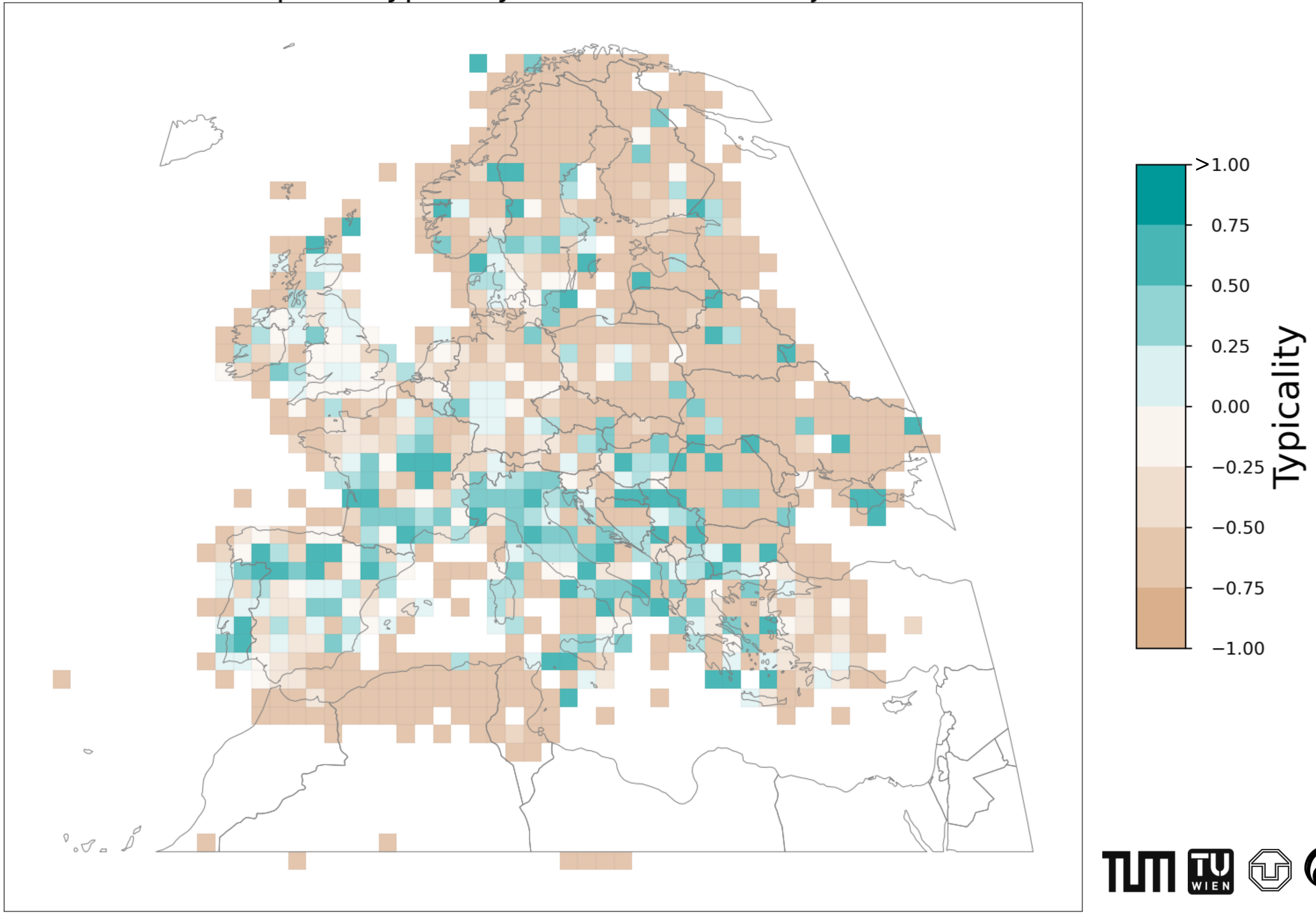
Spatial Typicality

100x100 km Grid Superimposed on Study Area

Spatial Typicality of Beer Mug Emoji



Spatial Typicality of Wine Glass Emoji



Research Questions

RQ1

Does the usage of emojis change over time and space?

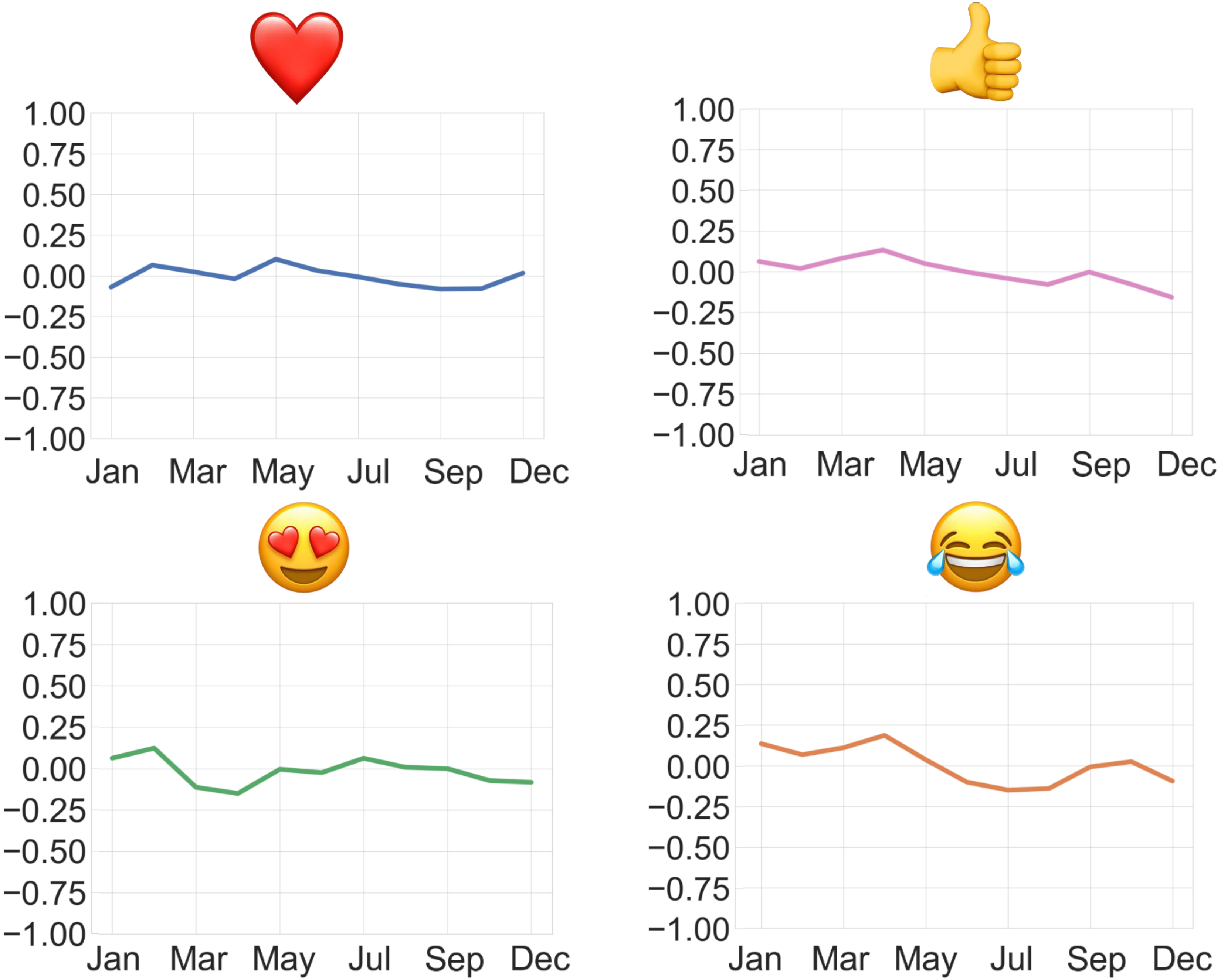
RQ2

Do changes in emoji usage have thematic connections?

RQ3

How can these spatial/temporal/ thematic changes be visualized?

❤️ 👍 😍 😂



1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
Jan
Mar
May
Jul
Sep
Dec
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
Jan
Mar
May
Jul
Sep
Dec
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
Jan
Mar
May
Jul
Sep
Dec
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
Jan
Mar
May
Jul
Sep
Dec

Too generic to
be useful

Much more
interesting for
analysis

Emoji-Specific Analysis

Example: Masked Face Emoji

Example: Masked Face Emoji

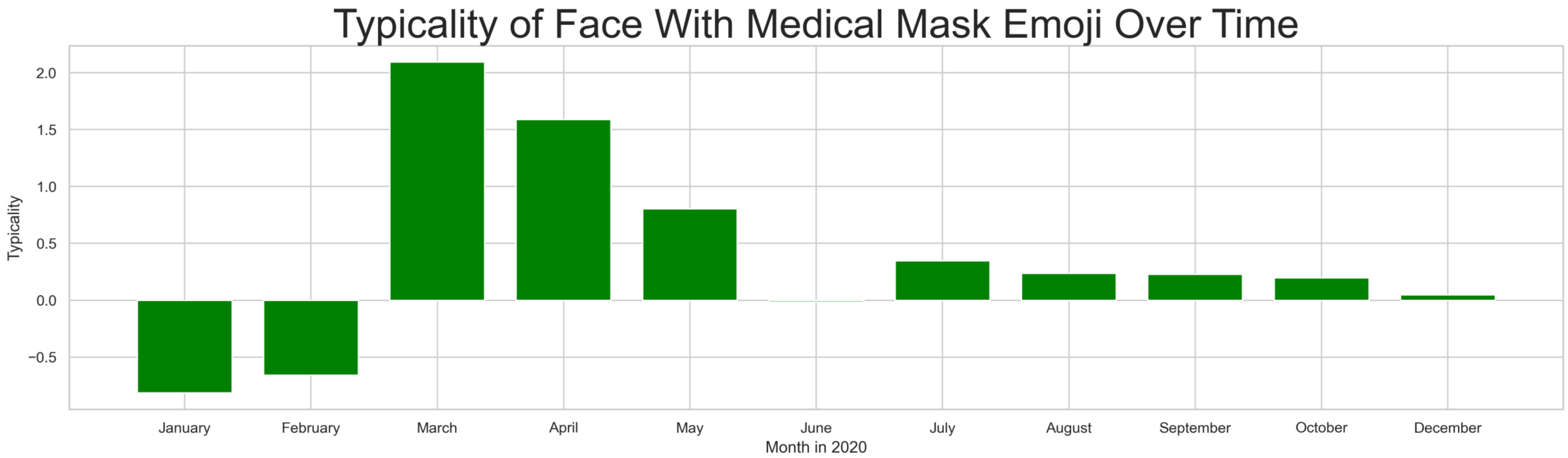
?

Example: Masked Face Emoji

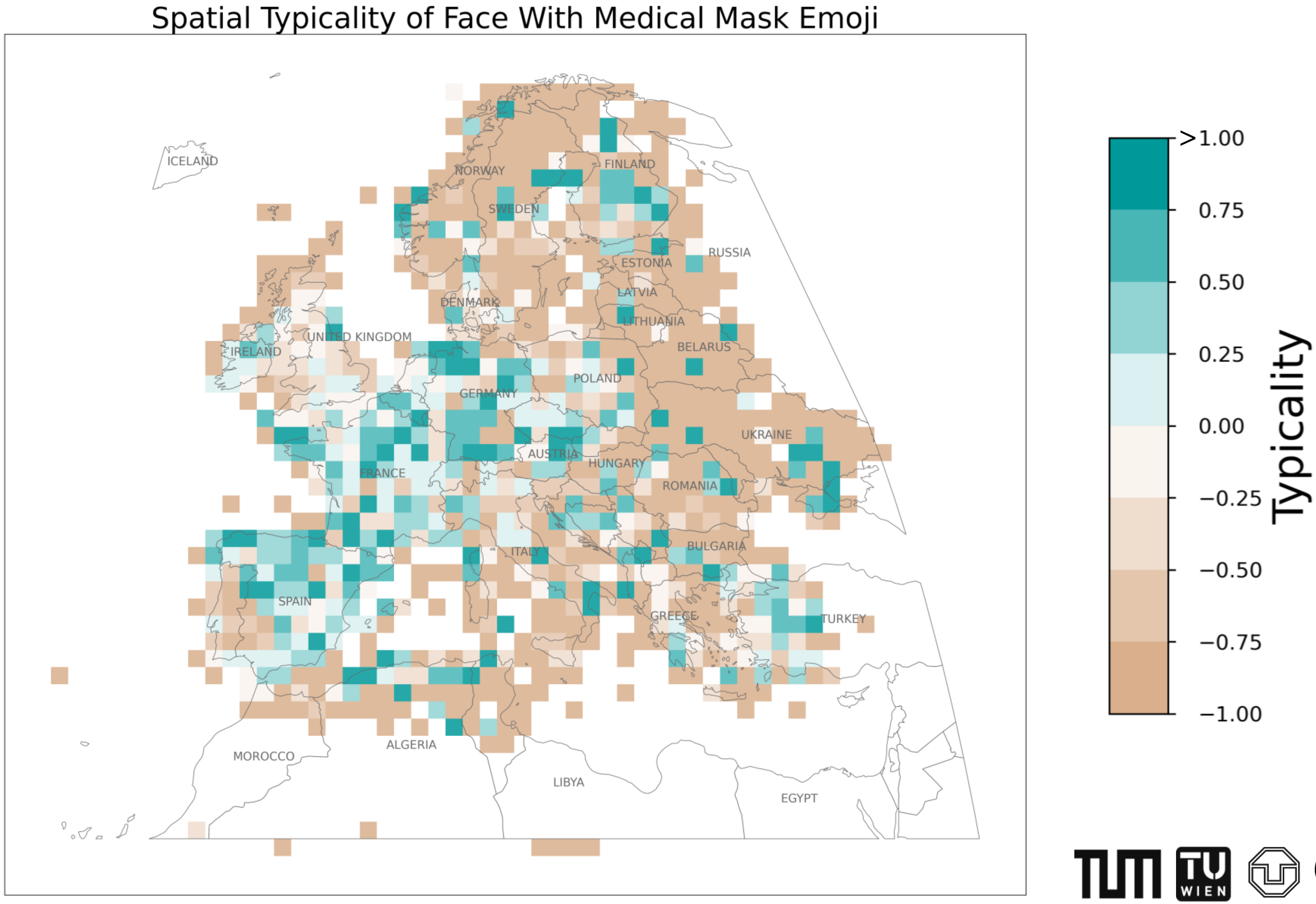
Rank	Hashtag	Uses
1	coronavirus	2186
2	covid19	1857
3	covid_19	906
4	awareness	609
5	corona	553
6	covid—19	447
7	staysafe	426
8	wearamask	343
9	nowwashyourhands	341
10	evdekal	291
11	yomequedoencasa	266
12	covid	217
13	covid2019	208
14	stayhome	205
15	mask	191
16	stayathome	178
17	quedateencasa	173
18	facemask	169
19	lockdown	166
20	catalunya	158

$\approx 92\%$

Example: Masked Face Emoji



Example: Masked Face Emoji



Research Questions

RQ1

Does the usage of emojis change over time and space?

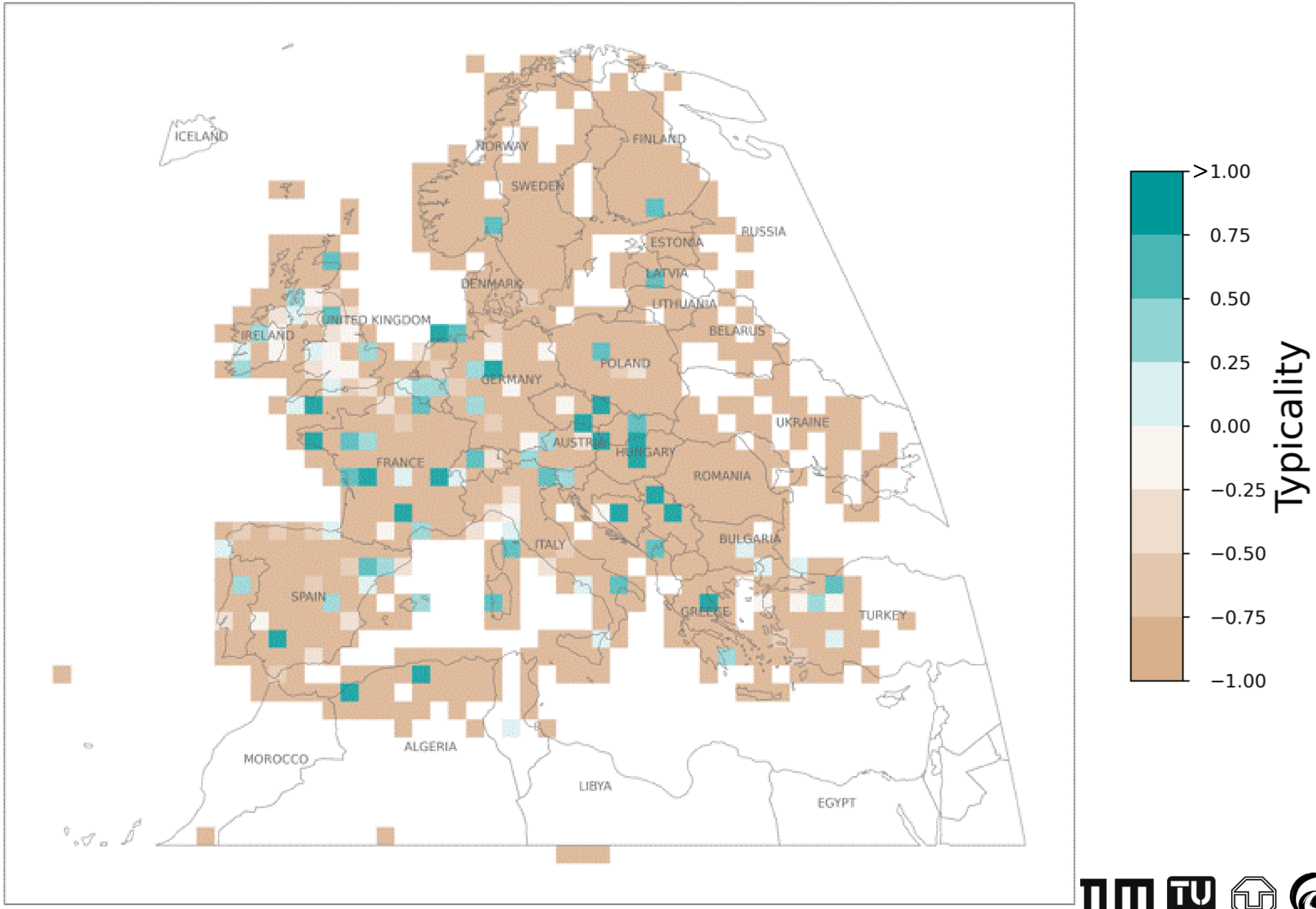
RQ2

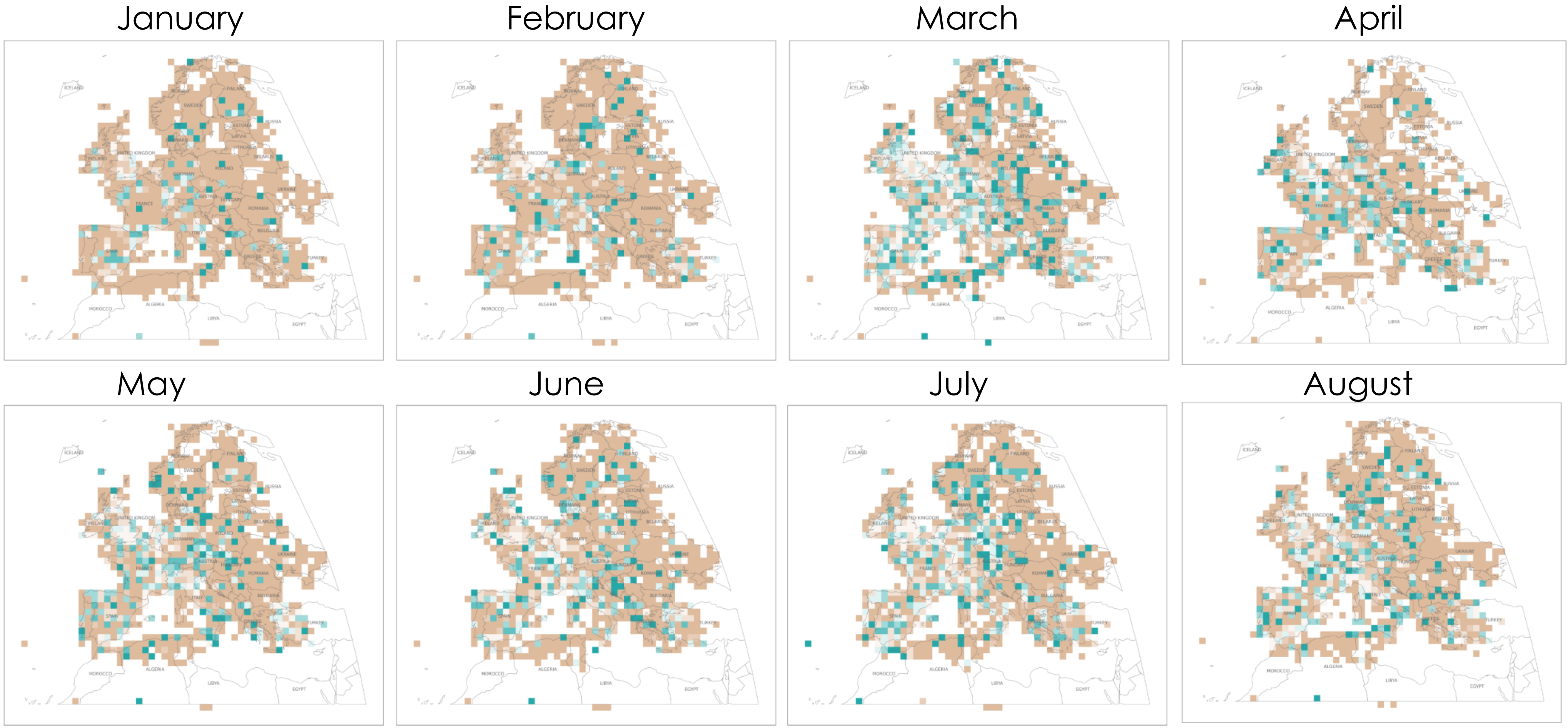
Do changes in emoji usage have thematic connections?

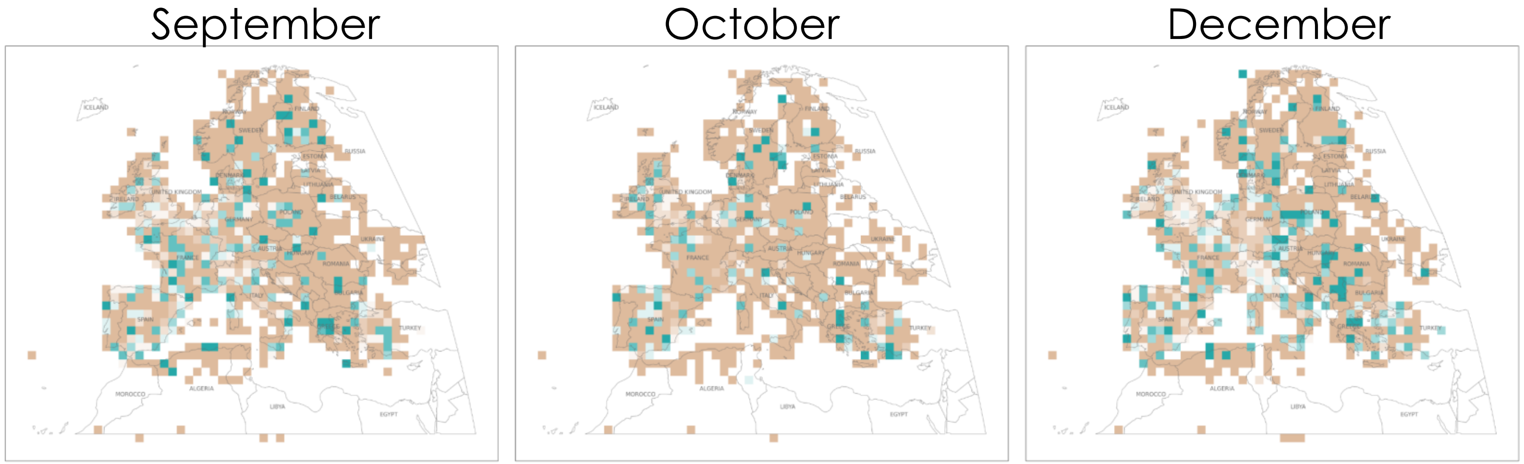
RQ3

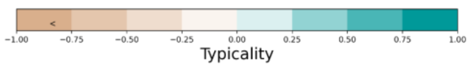
How can these spatial/temporal/thematic changes be visualized?

Typicality of Face With Medical Mask Emoji (January)









One grid cell represents 10,000 square kilometers.

Data Sources: Natural Earth, Twitter database of the Technical University of Dresden Institute of Cartography

Maps by Samantha Levi

Results

Selected Emojis

Baseline

🍺 🍷

Predicted Topical Consistency

🙏 😷 ✊ 🌈 💉 🎄 🦠 👏

Typical within 1 of 10 Top Countries by User Days

U.K.	Spain	France	Germany	Italy	Turkey	Netherlands	Belgium	Switzerland	Austria
🤱	🕓	🗳️	🤮	💣	👄	🚁	🖤	💙	😍
🧁	👩‍⚕️	💬	☝️	🤡	🔞	🚒	🤔		
🏌️	🏥	⤵️	🤓	💎	🎀		😎		

Detected Topics

COVID-19
Pandemic

Activist
Movements

Lifestyle /
Leisure

Detected Topics

COVID-19
Pandemic

Virus

Healthcare worker appreciation

Lifestyle /
Leisure

Breastfeeding

Golf

Black Lives Matter

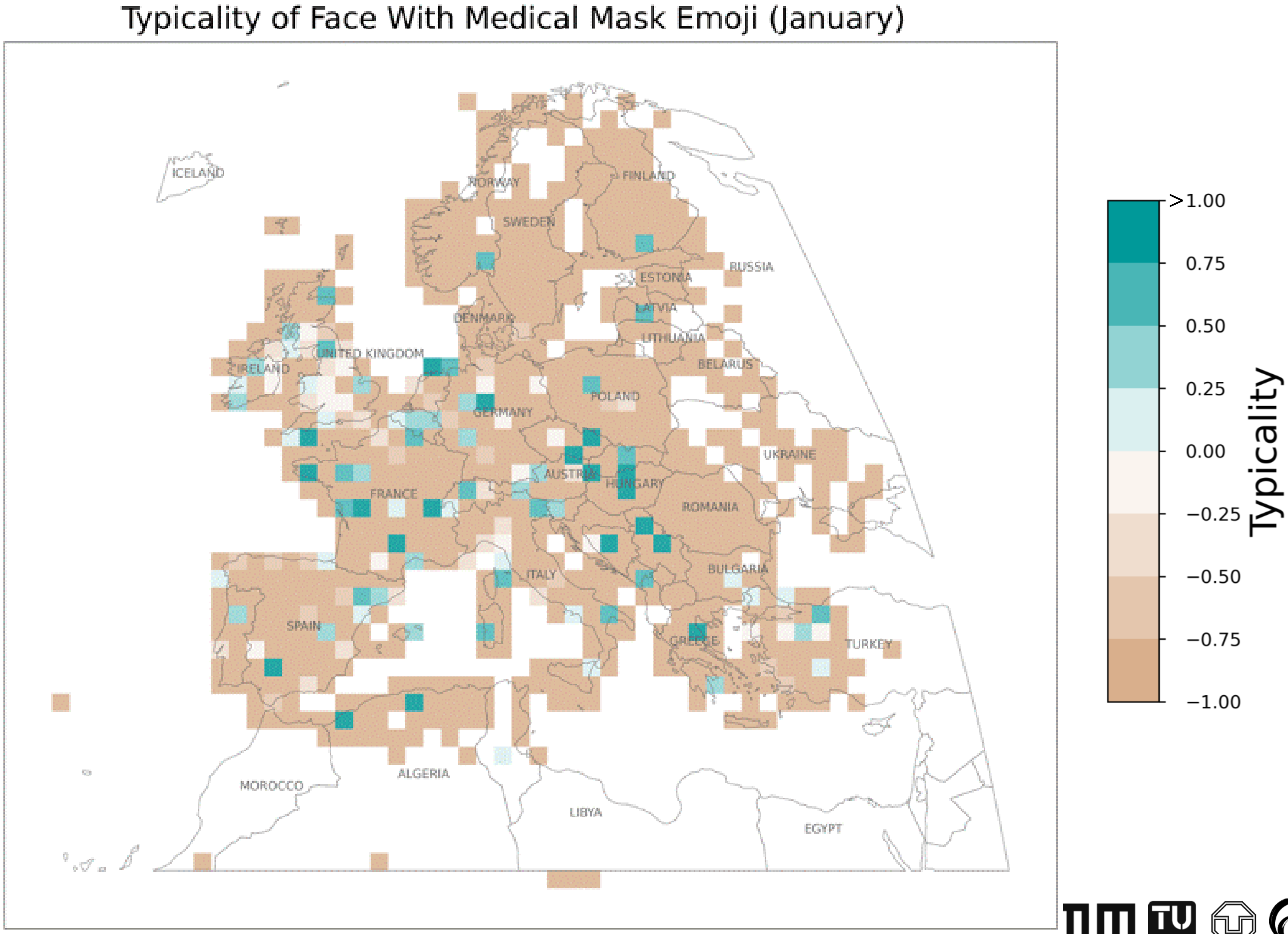
LGBTQIA+ Pride

Elections

Activist
Movements

Topic:
COVID-19

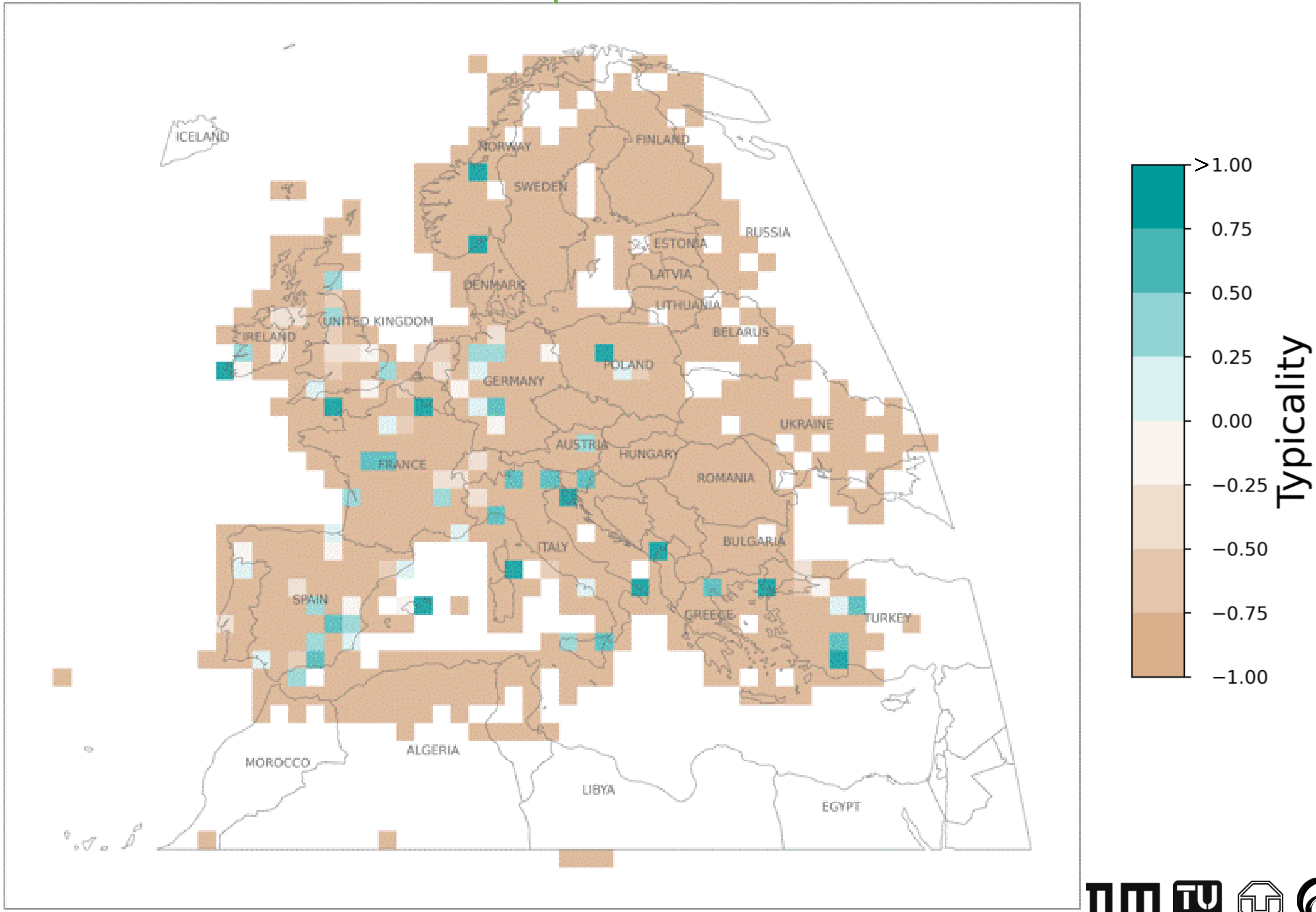
Topical Consistency:
92.2%



Typicality of Microbe 🦠 Emoji (January)

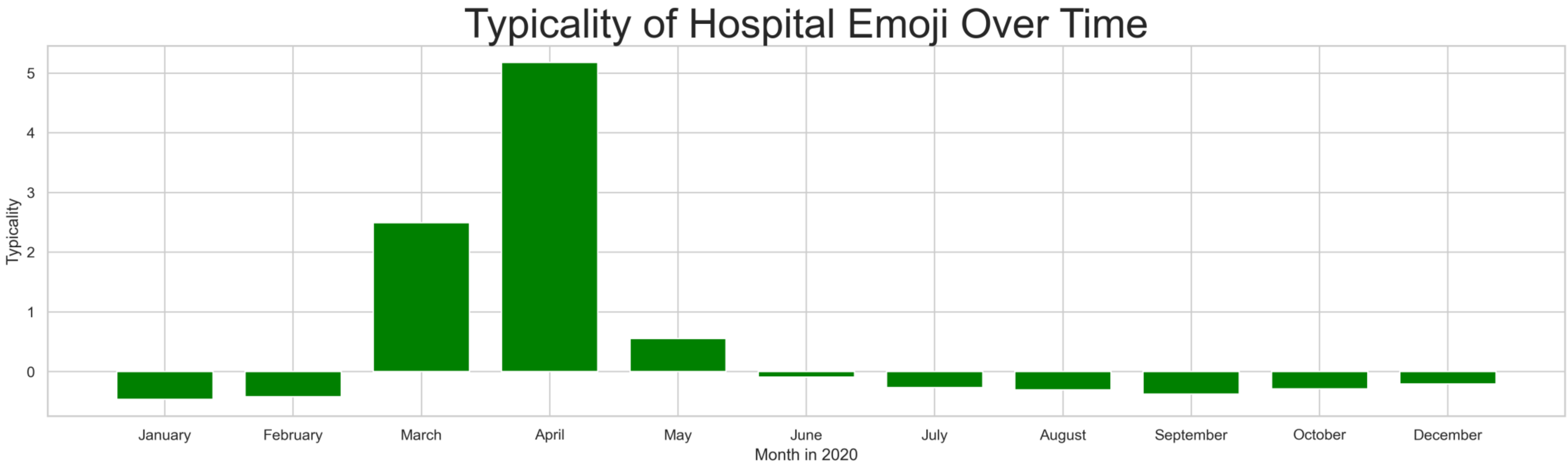
Topic:
COVID-19

Topical Consistency:
98.7%



Topic:
COVID-19

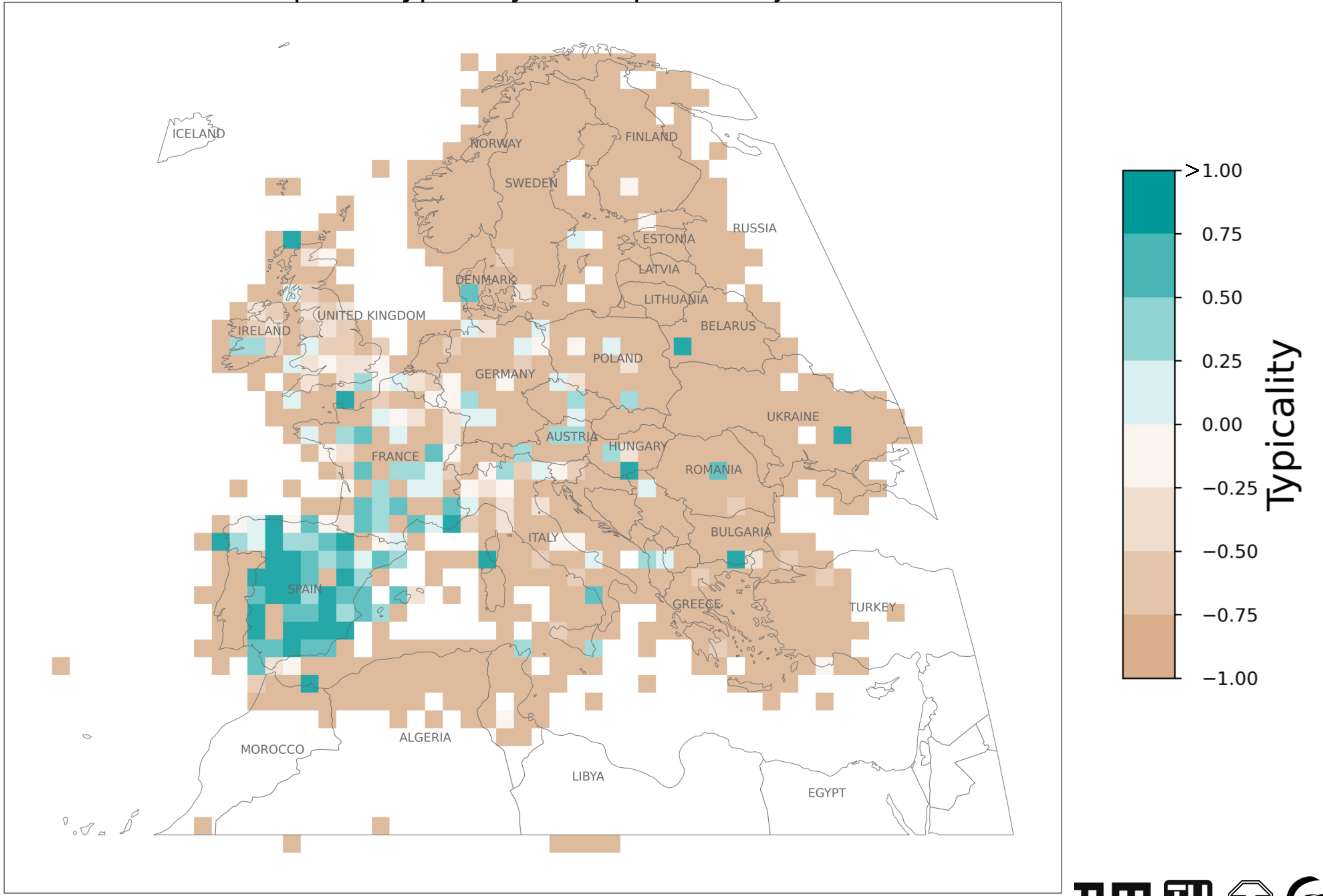
Topical Consistency:
88.2%



Topic:
COVID-19

Topical Consistency:
88.2%

Spatial Typicality of Hospital Emoji



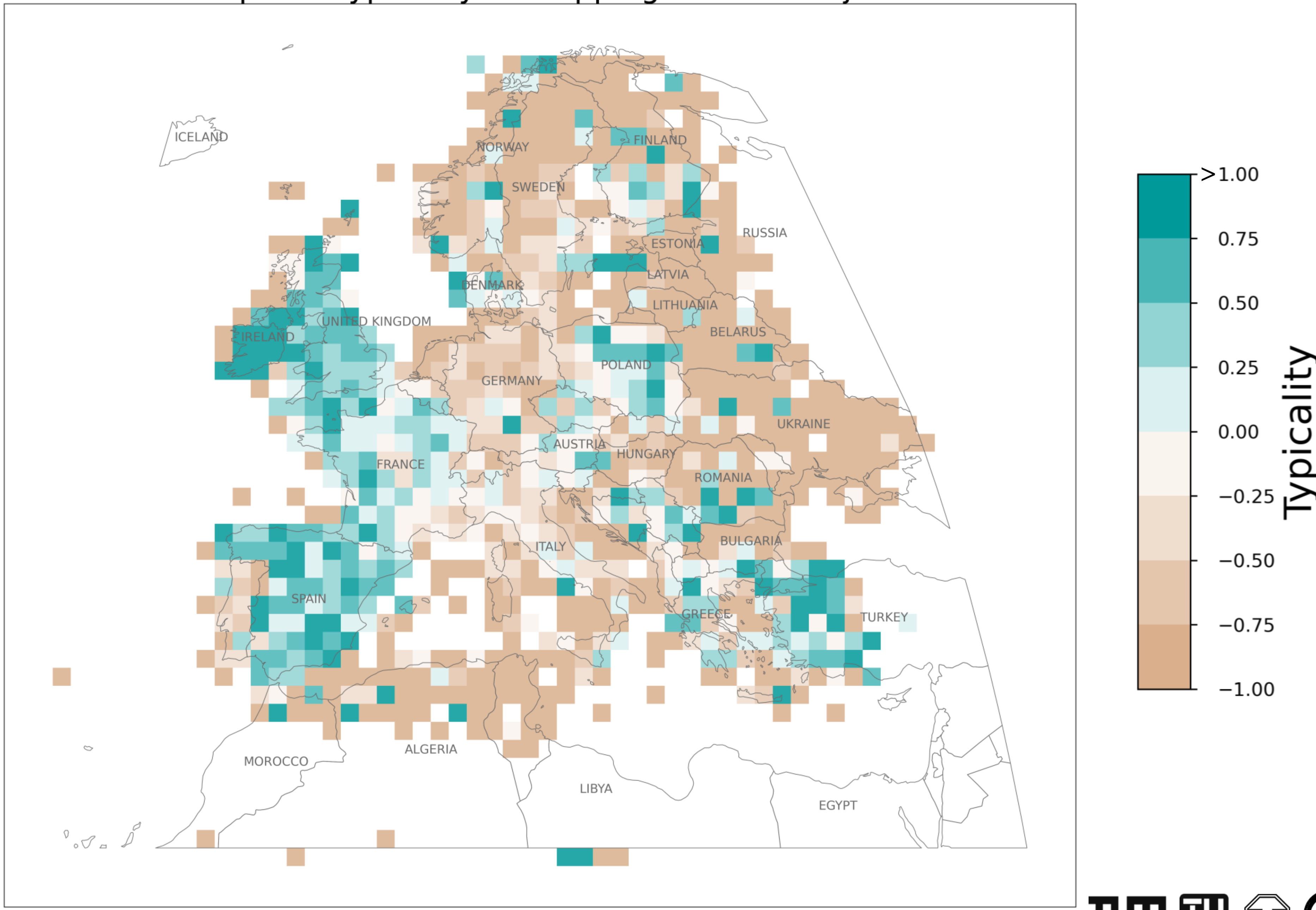
Topic:

Healthcare Worker Appreciation

Topical Consistency:

89.7%

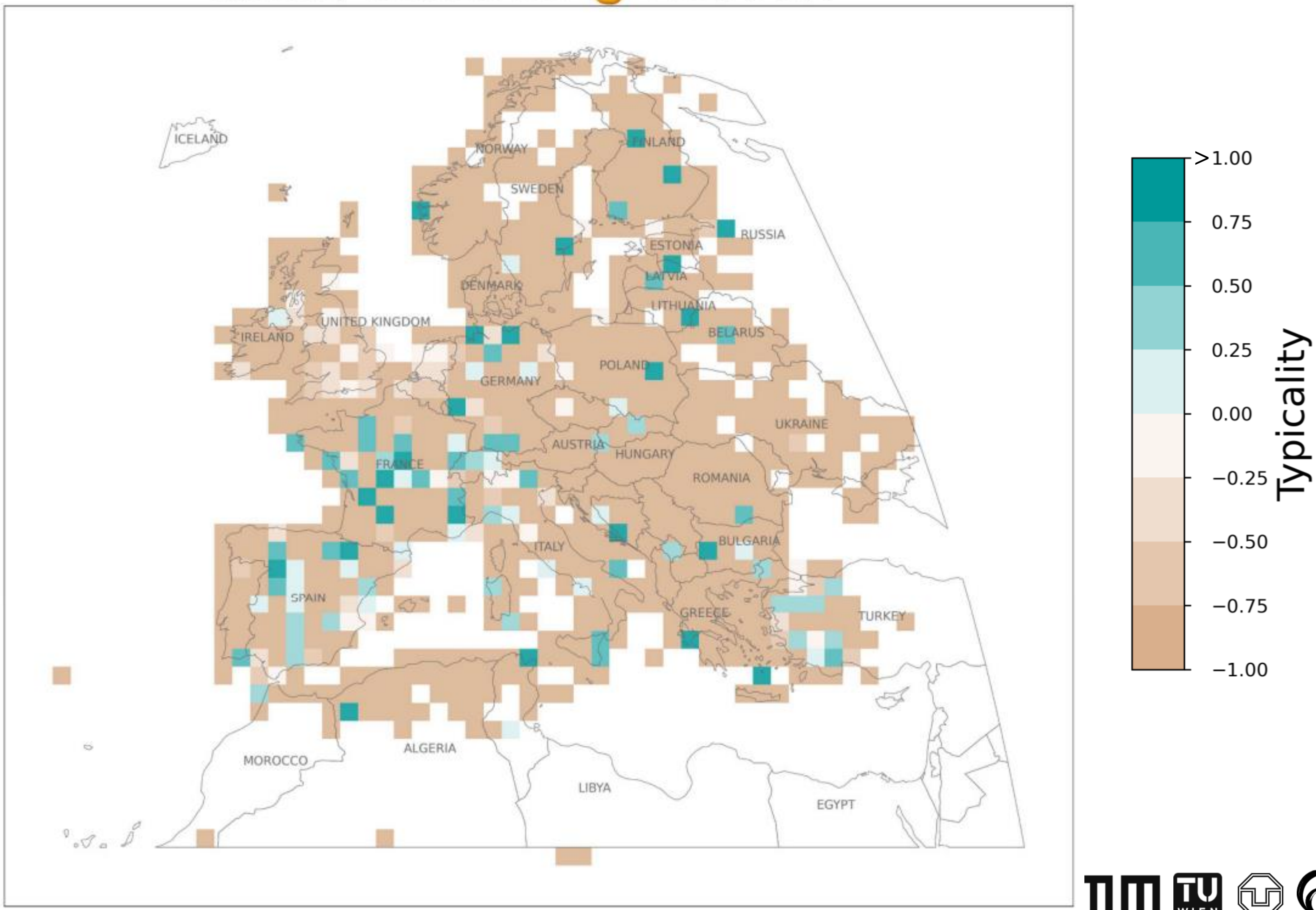
Spatial Typicality of Clapping Hands Emoji



Topic:
Black Lives Matter

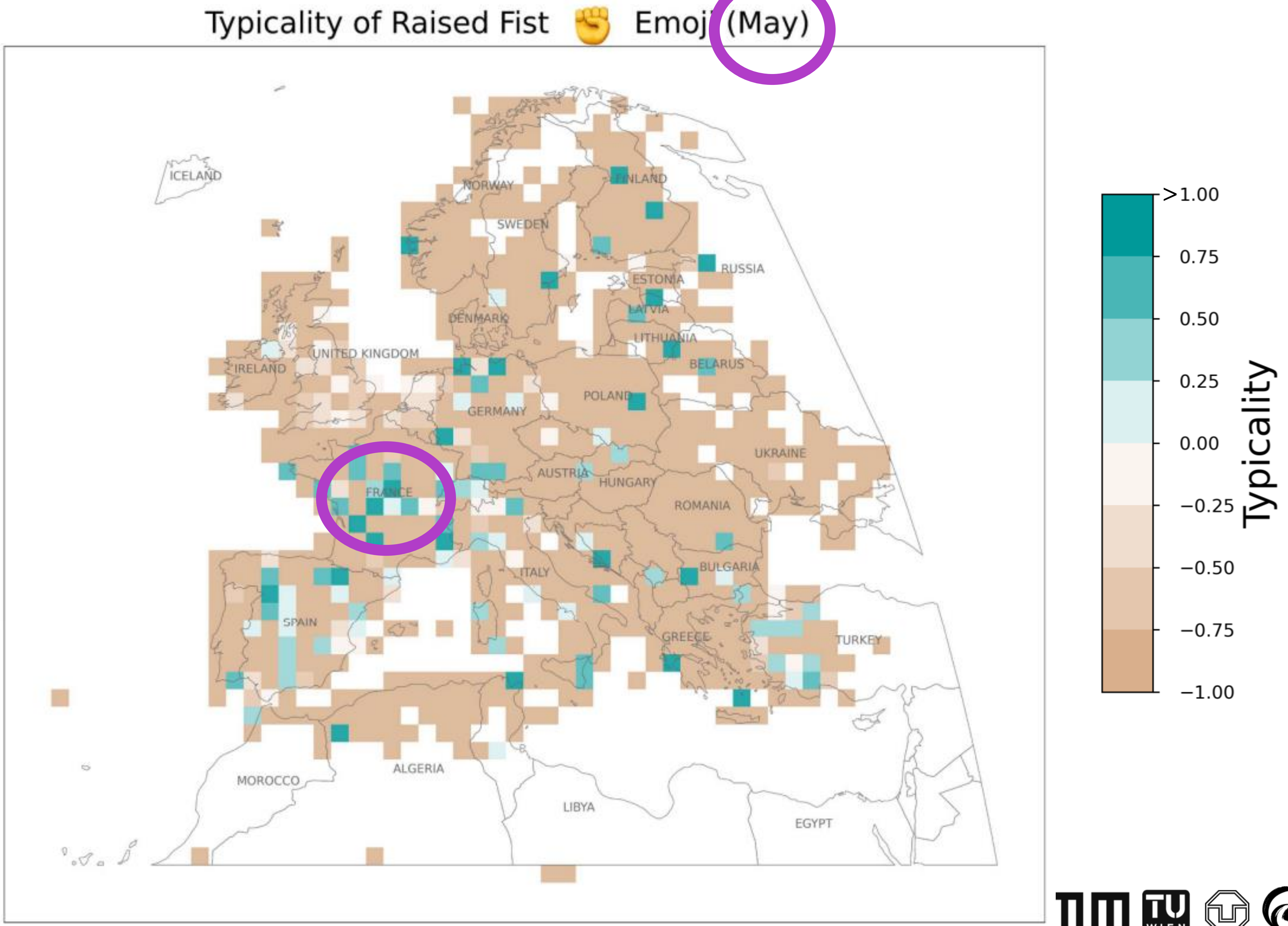
Topical Consistency:
90.9%

Typicality of Raised Fist ✊ Emoji (May)

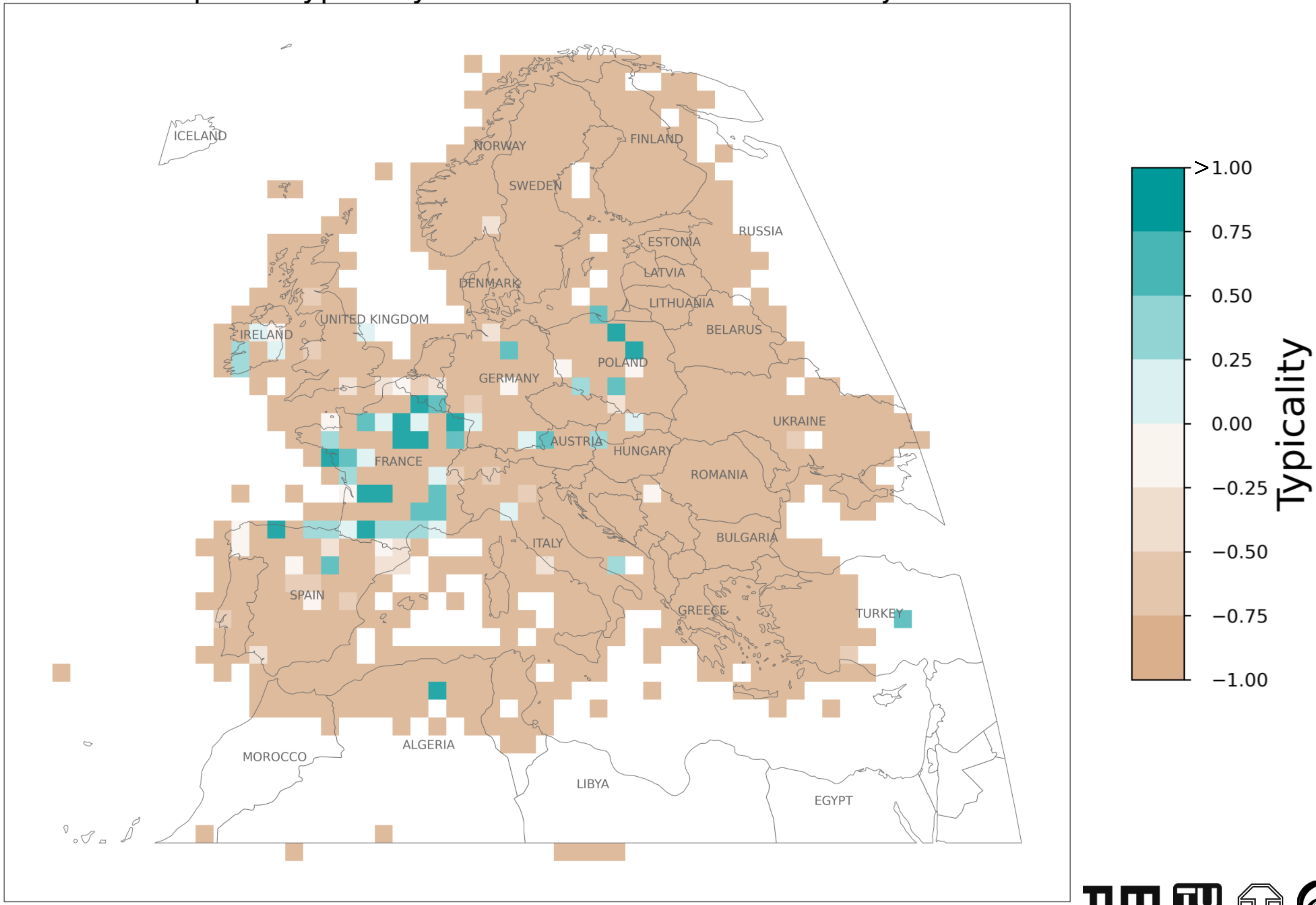


Topic:
Black Lives Matter

Topical Consistency:
90.9%



Spatial Typicality of Ballot Box With Ballot Emoji



Topic:
Elections

Topical Consistency:
72%

Conclusions

Objective

Determine whether emojis can be used to identify relevant topics and their spatial-temporal evolution in a non-topic-specific dataset

Objective

Determine whether emojis can be used to identify relevant topics and their spatial-temporal evolution in a non-topic-specific dataset

Conclusions

Topical Consistency

- Emojis with less topical consistency are less reliable proxies
- Labor-intensive to calculate, subject to human bias and error

Metrics Matter

- Absolute, relative frequencies are easily skewed by hyper-active users
- Metrics like typicality and user days help to mitigate this influence

Privacy Awareness

- Strive to protect user privacy wherever possible
- Degree of necessary privacy depends on the applications of results

Limitations of HLL Data

- Benefits: cardinality, unions, privacy for investigation of social facet
- Drawbacks: cannot analyze multiple facets simultaneously, emojis and hashtags are separated

Future Work

Quantitative Spatial-Temporal Analysis

- Use clustering algorithms, space-time scan statistics to detect statistically significant clusters of emoji usage

Head-Tail Breaks for Visualizations

- Account for skew towards negative values in spatial typicality maps

Emoji Expansion

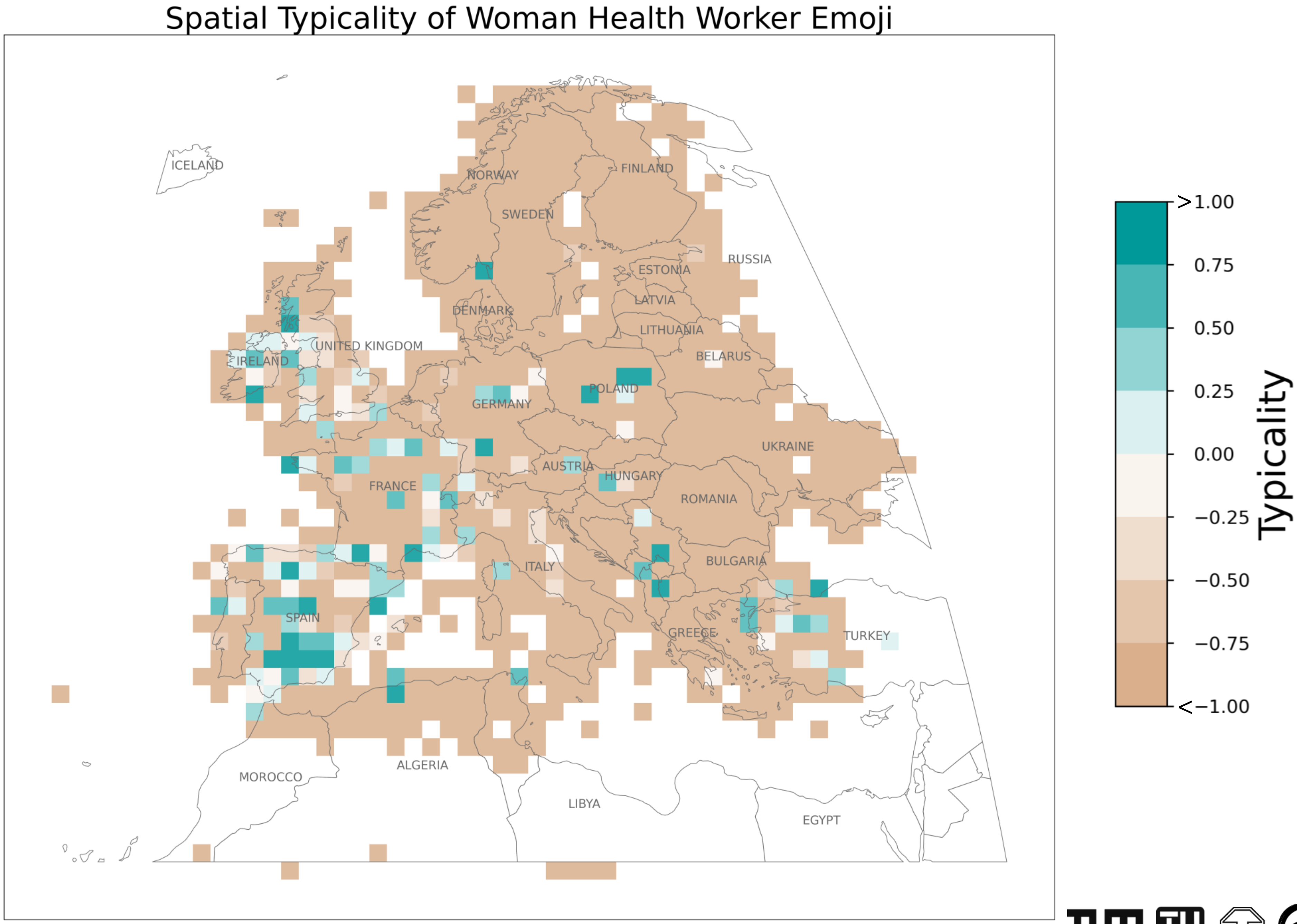
- Flag emojis, skin tone modifiers

Thank You!

Questions? Comments?

Topic:
Healthcare Worker Appreciation

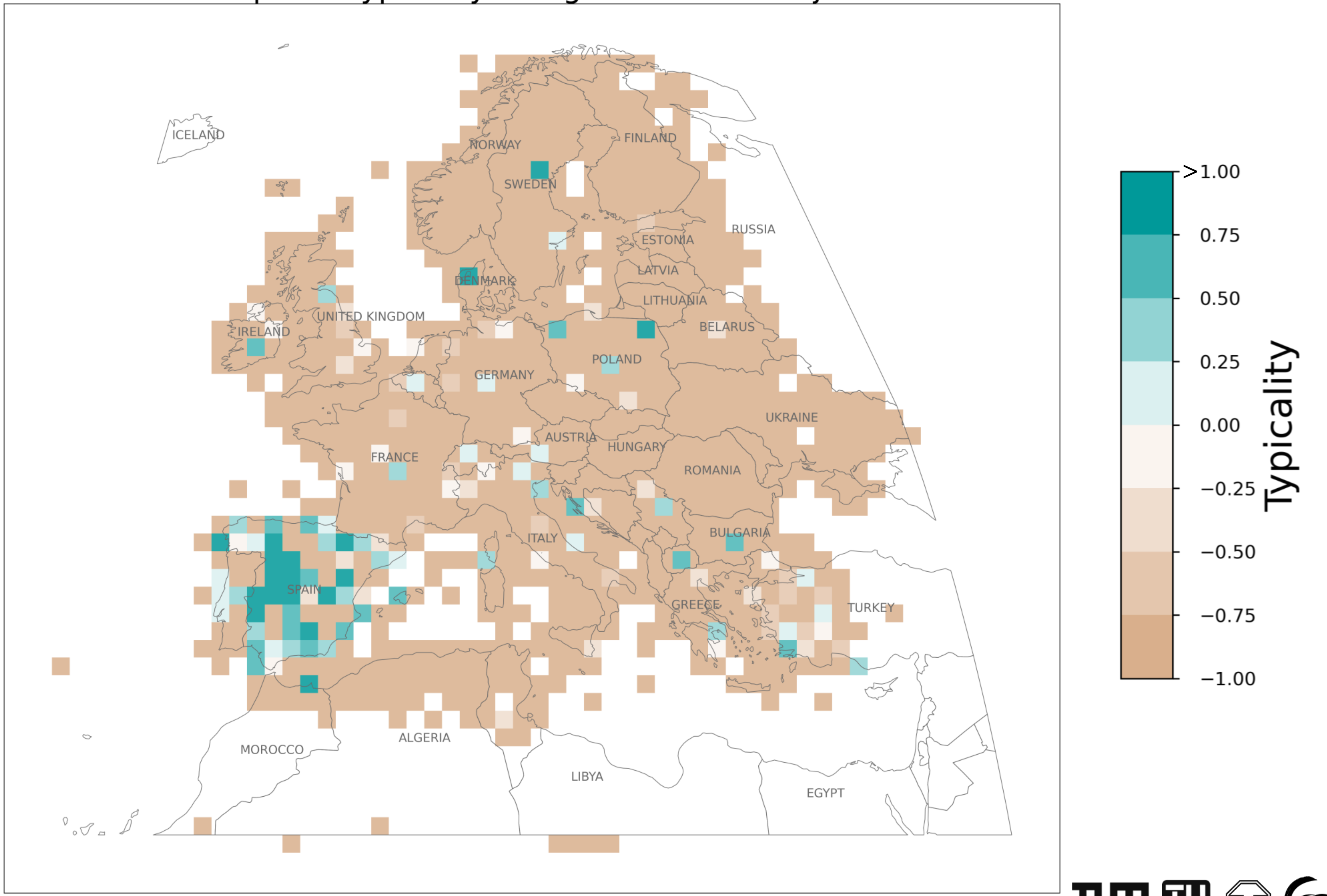
Topical Consistency:
88.2%



Topic:
Healthcare Worker Appreciation

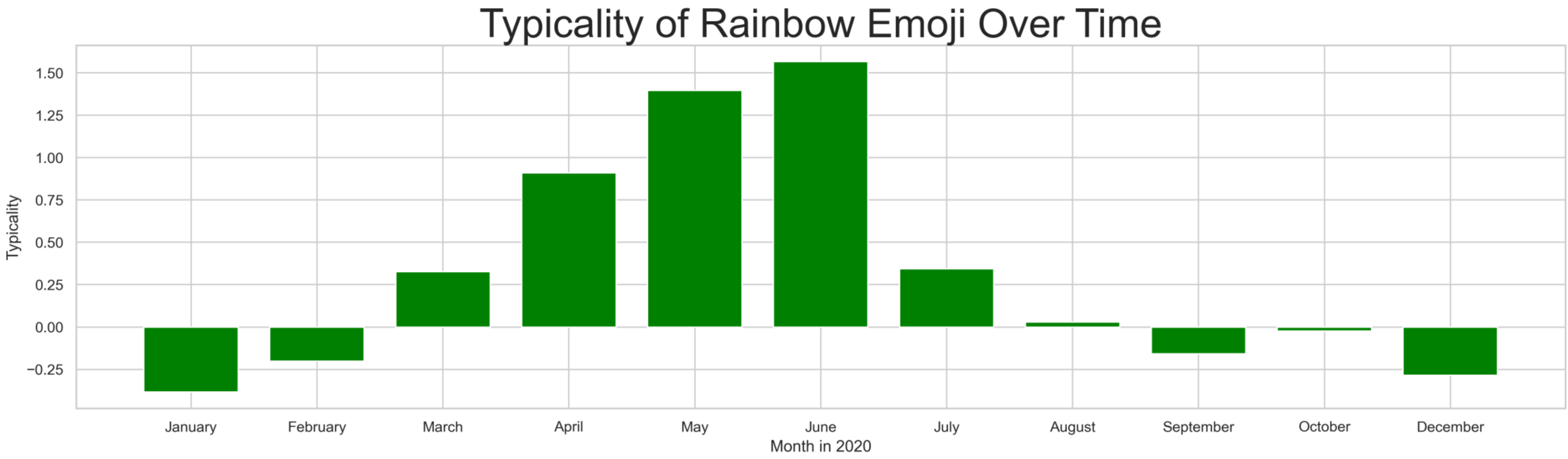
Topical Consistency:
78%

Spatial Typicality of Eight O'Clock Emoji



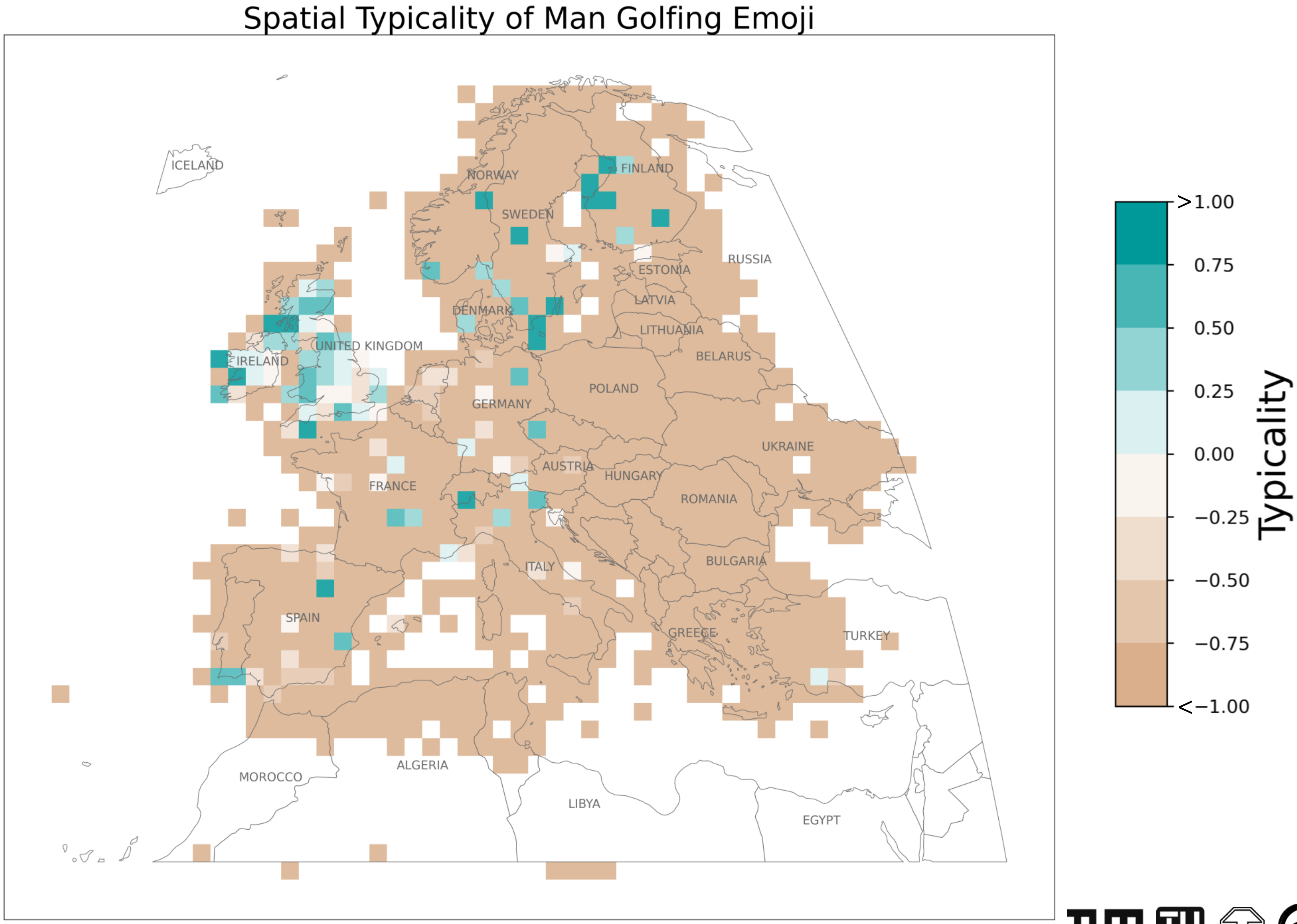
Topic:
LGBTQIA+ Pride

Topical Consistency:
71.7%



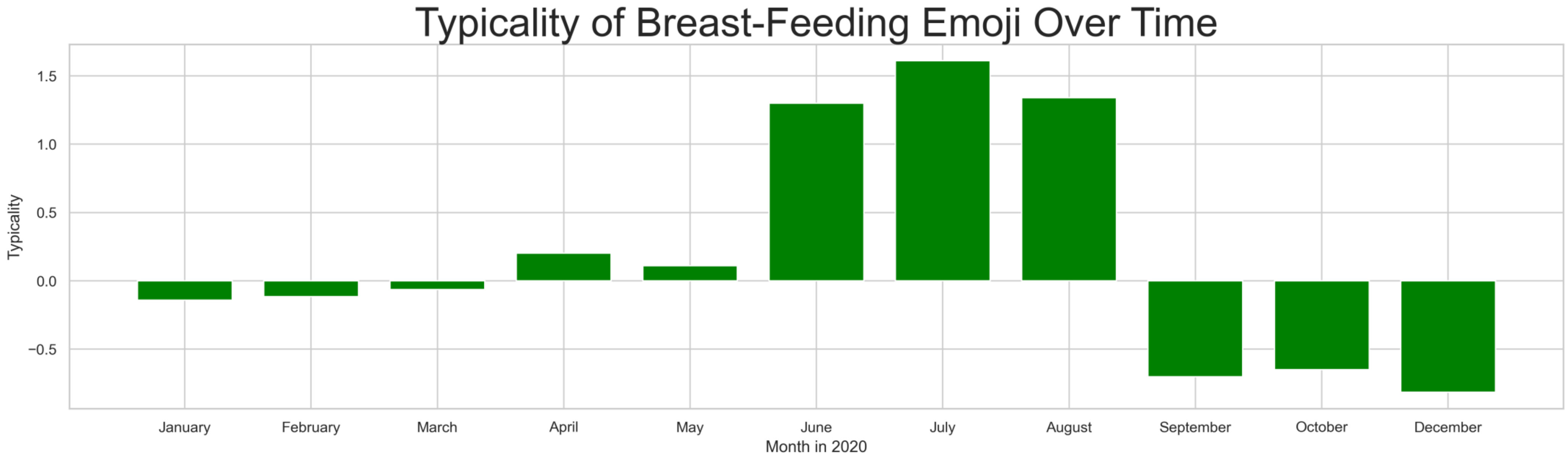
Topic:
Golf

Topical Consistency:
86.9%

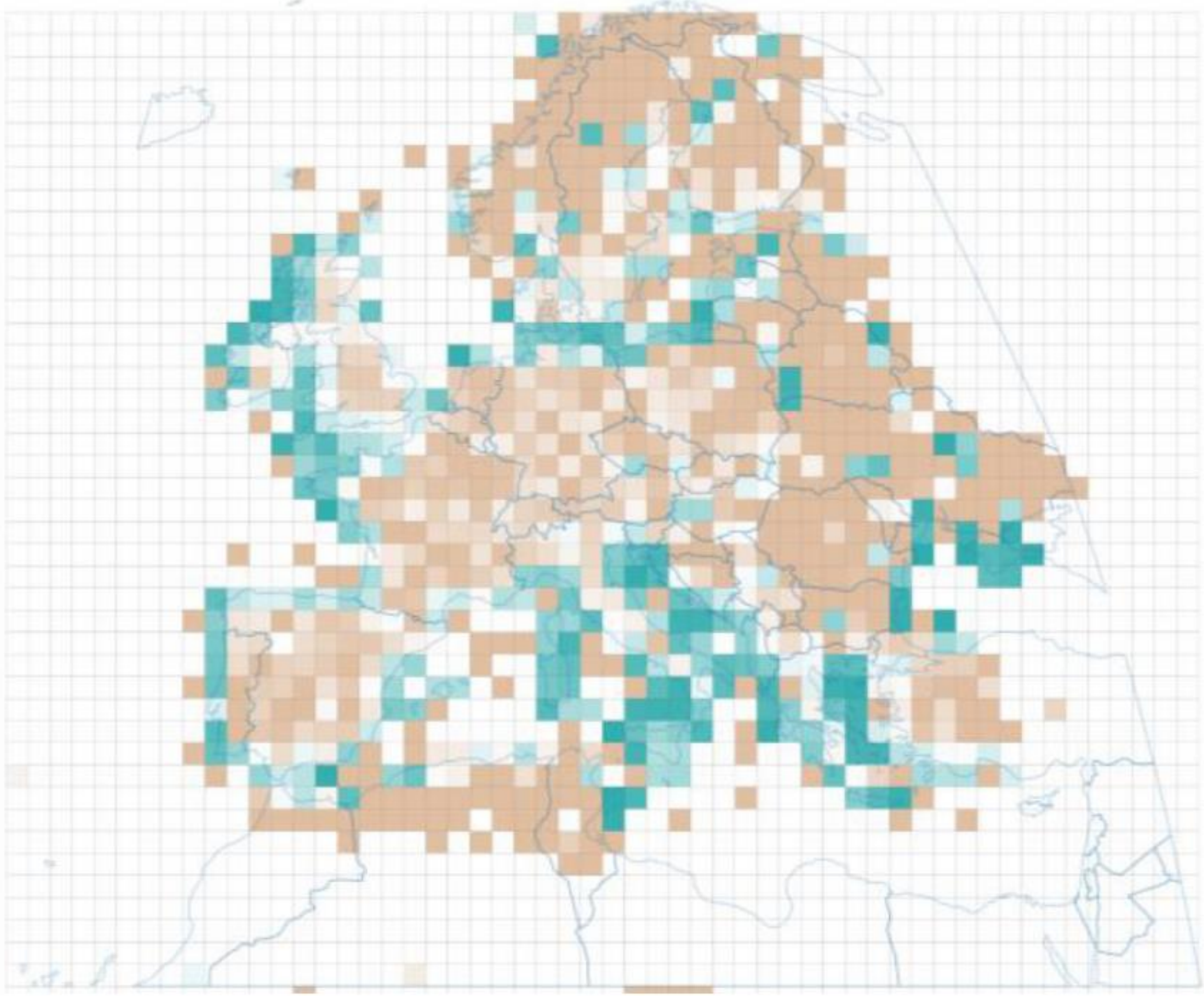


Topic:
Breast-feeding

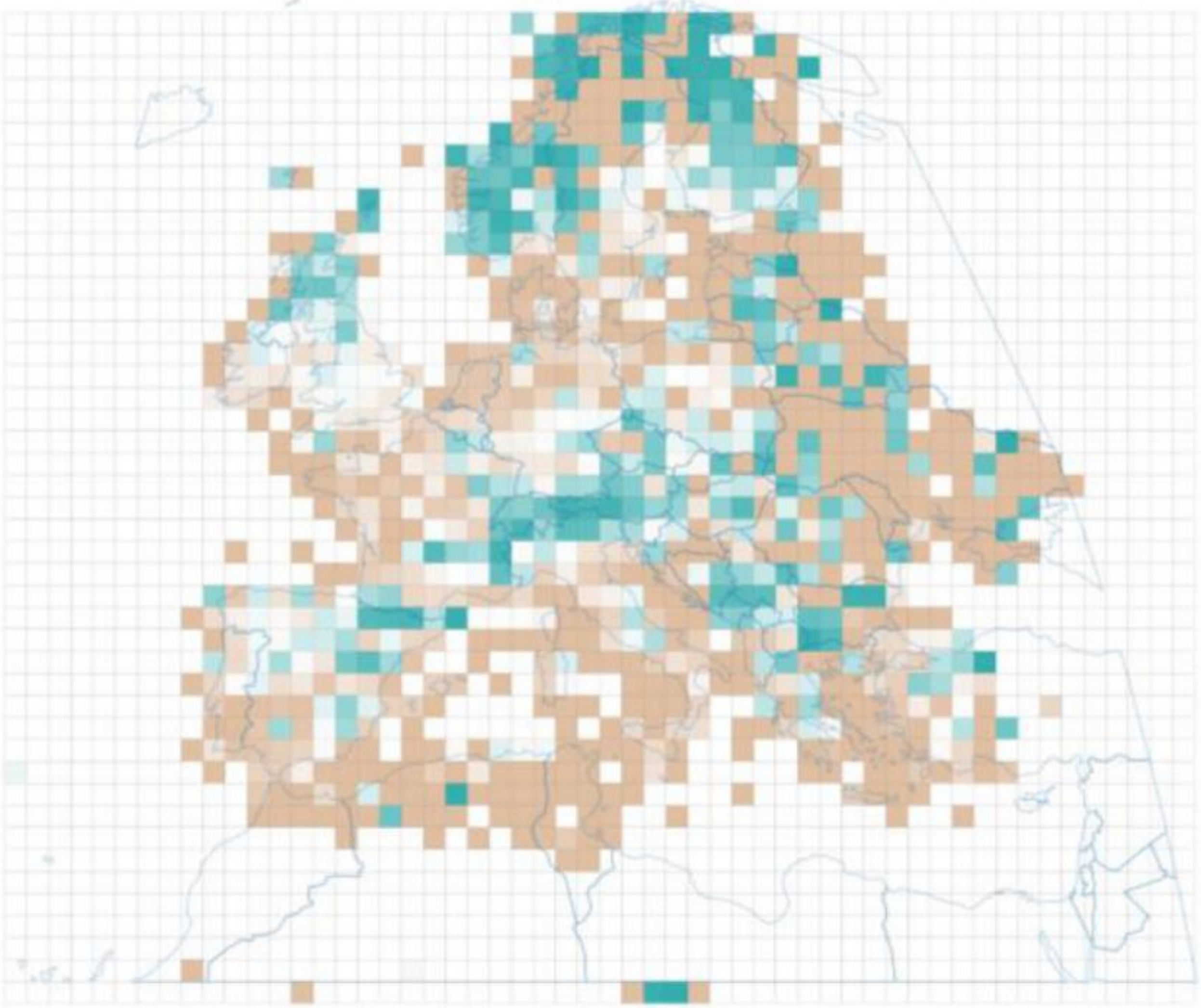
Topical Consistency:
88.8%



Spatial Typicality of the Water Wave Emoji 🌊

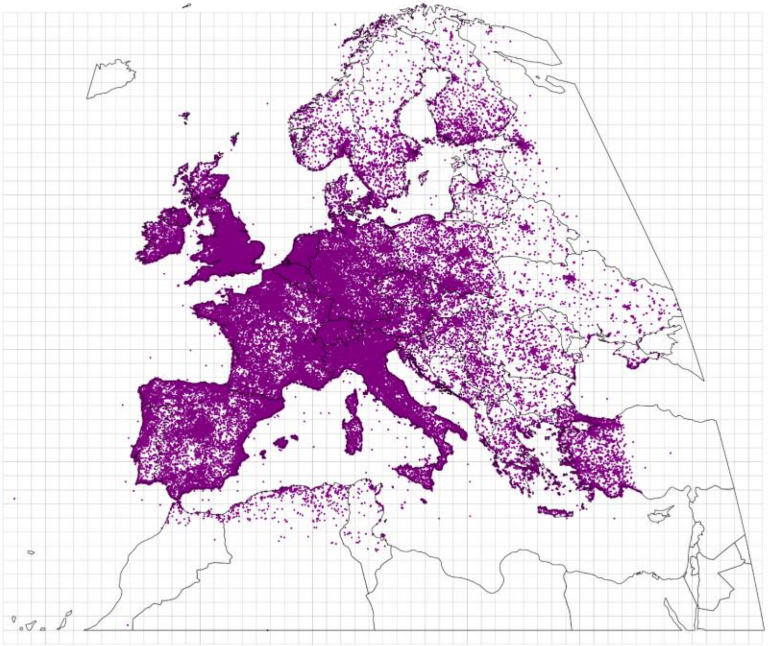


Spatial Typicality of the Snowflake Emoji ❄️

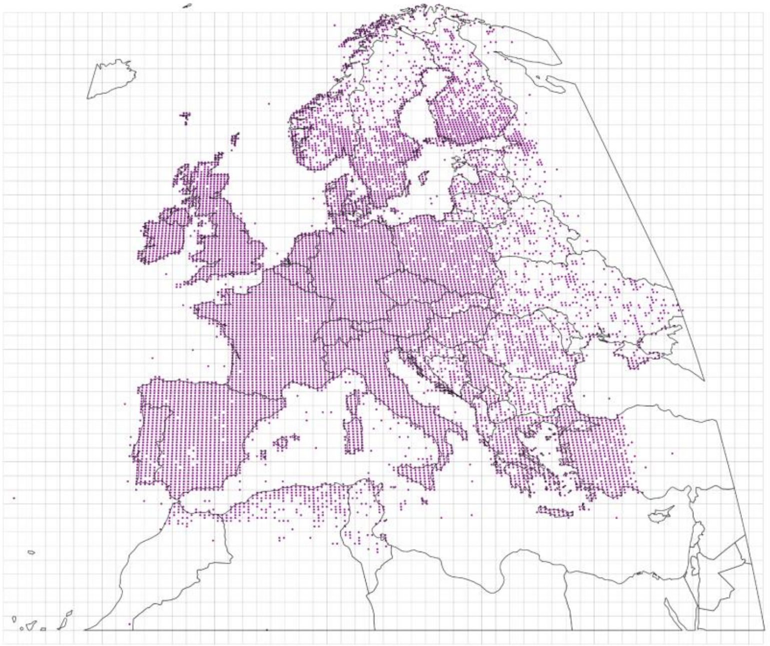


Geo-Hashing Aggregation Levels

Aggregation Level 5



Aggregation Level 4



Aggregation Level 3

MAUP

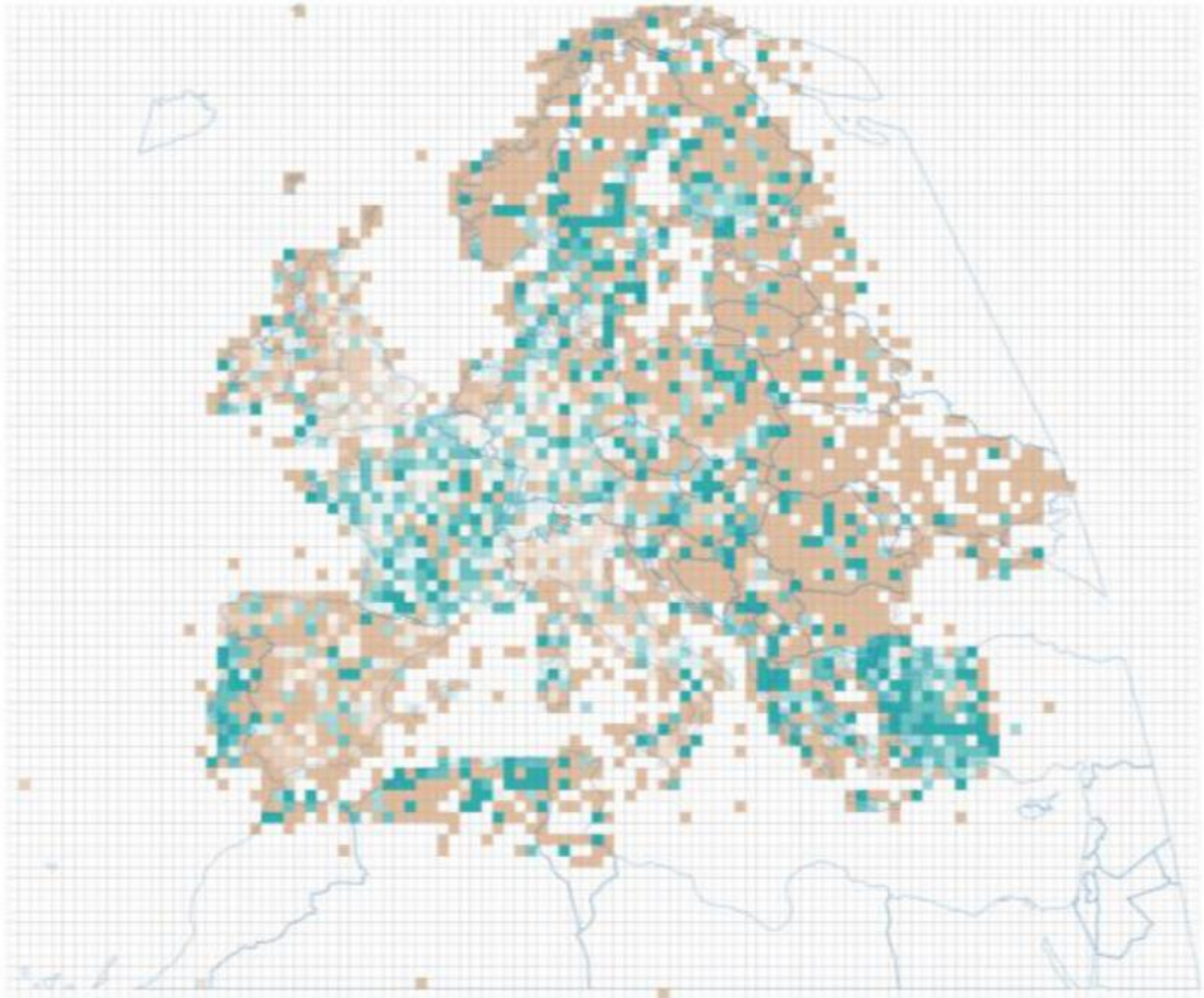
Country	User Days (Agg. Level 4)	User Days (Agg. Level 5)	Difference
United Kingdom	811956	810535	0.18%
Spain	288547	288819	-0.09%
France	228942	230294	-0.59%
Germany	143224	142974	0.17%
Italy	143012	141807	0.85%
Turkey	111138	108351	2.57%
Netherlands	76856	73083	5.16%
Belgium	40219	39852	0.92%
Switzerland	20624	23061	-10.57%
Austria	17732	18070	-1.87%
Portugal	13177	12699	3.76%
Czech Republic	10485	10711	-2.11%

Design Considerations

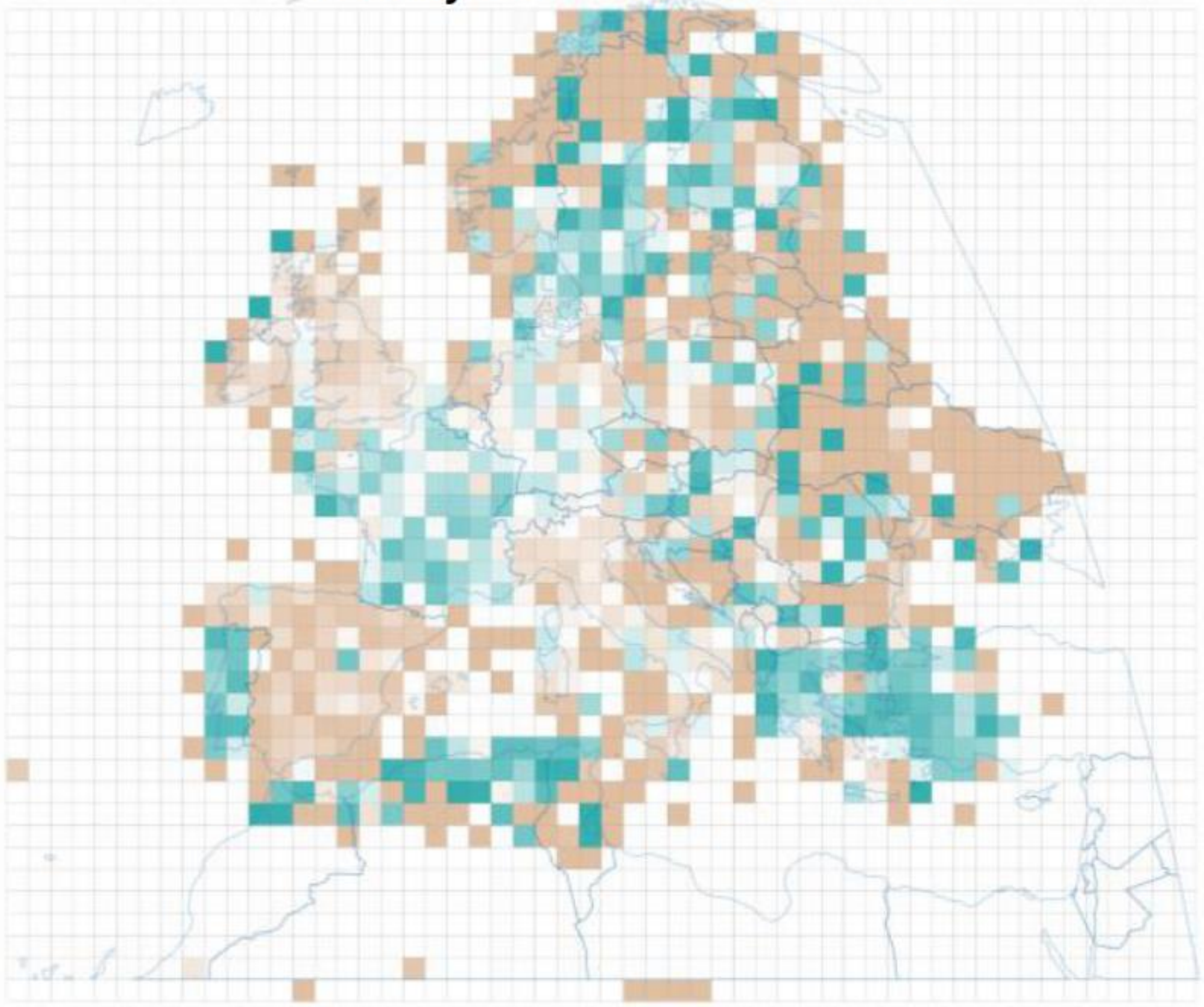
Grid Size

Spatial Typicality of the Folded Hands Emoji 🙏

50 by 50 Kilometer Grid



100 by 100 Kilometer Grid



Cropping the Data