Covid-19 World Tweet visualization

Visual analytics project

Gabriele Marcozzi, Nicolò Palmiero

Introduction

Why visualize Covid-19 related Tweets?

- Temporal correlation between Tweets and Covid-19 cases
- Spatial correlation between Tweets and Covid-19 cases

The goal of this project is to find out what relationships exist between the number of tweets related to covid-19 and the spread of the pandemic.

Related Work

Visualizing the Covid-19 Twitter chatter dataset

Similar solutions, objectives and views

- heat map showing the different number of tweets according to the country.
- bar-chart showing the comparison of number of tweets by country in a bar view.

Design and analysis of a large-scale COVID-19 tweets dataset

Similar solutions and views

- representing the total number of tweets in a month-by-month representation, according to the country.

Related Work

Topics, Trends, and Sentiments of Tweets About the COVID-19 Pandemic: Temporal Infoveillance Study

Similar solutions and objectives

- examines key themes and topics of English-language COVID-19–related tweets posted by individuals and explores the trends and variations of the COVID-19–related tweets.

Public risk perception and emotion on Twitter during the Covid-19 pandemic

Similar objectives and views

- comparison of the proximity of each country

Dataset

The Coronavirus (Covid-19) geo-tagged tweets dataset from IEEE-DataPort contains information from March 2020 to January 2021 about covid-19 related tweets from all over the world.

Preprocessing

- we decided to shrink the dimensions;
- we decided to drop some columns which contained sparse data;
- we created another column which indicates the ISO 3166 standard id of each country

Moreover, we added two more columns from two different datasets that are *population*, that is the number of people that live in each country, and *cities* that is the number of cities of each country, to have more contextual informations to conduct our analysis.

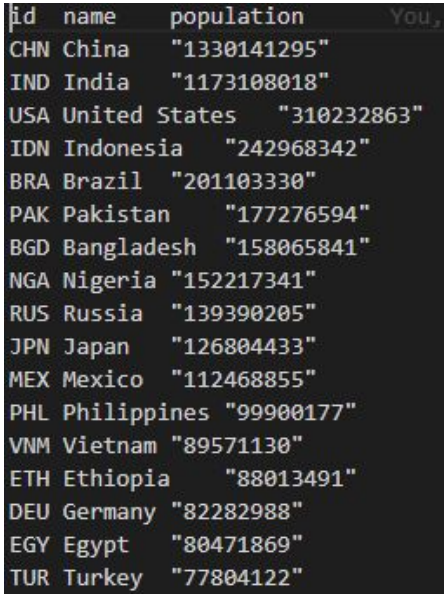
Dataset

At the end of this preprocessing final dataset is composed by 7 columns and 274242 rows. So respecting the AngeliniSantucci index with **AS = 1,919,694**.

	coordinates	created_at	place	retweet_count	text	user_friends_count	country_id
0	7.81314396,60.45918237	Thu Nov 26 05:57:26 +0000 2020	Hol, Norge	0	Bruh.... can 202	640	NOR
1	76.20987122,10.52470892	Thu Nov 26 05:45:02 +0000 2020	Trissur, India	0	Modi Govt	134	IND
2	-73.47935661,45.6677507	Thu Nov 26 10:23:02 +0000 2020	Varennes, QuÃ©bec	0	Fighting Stigma	1306	CAN
3	29.40201953,36.47161396	Thu Nov 26 08:09:54 +0000 2020	Fethiye, MuÄŸla	0	#tbt to sitting	1073	TUR
4	-2.5280328,51.4774408	Thu Nov 26 06:29:20 +0000 2020	Bristol, England	0	Bristol tattoo	795	GBR
5	133.68232727,-32.12074612	Thu Nov 26 05:39:01 +0000 2020	Ceduna, South Australia	0	COVID-19: New	114	AUS
6	-85.6557,42.9613	Thu Nov 26 06:31:58 +0000 2020	Grand Rapids, MI	0	Mr. Clean Bean	1204	USA
7	145.06548209,-37.98121944	Thu Nov 26 04:29:45 +0000 2020	Melbourne, Victoria	0	These vintage s	1291	AUS
8	-0.12485951,51.50538967	Thu Nov 26 10:16:02 +0000 2020	London, England	1	#WHO says adu	445	GBR
9	13.401251,52.518391	Thu Nov 26 09:28:01 +0000 2020	Berlin, Germany	0	When other	77	DEU
10	-74.0007613,40.7207559	Thu Nov 26 09:50:49 +0000 2020	Manhattan, NY	1	BLACK PANDEM	281	USA
11	-100,40	Thu Nov 26 05:47:09 +0000 2020	Kansas, USA	0	Rage against th	69	USA

Dataset

Two auxiliary datasets:

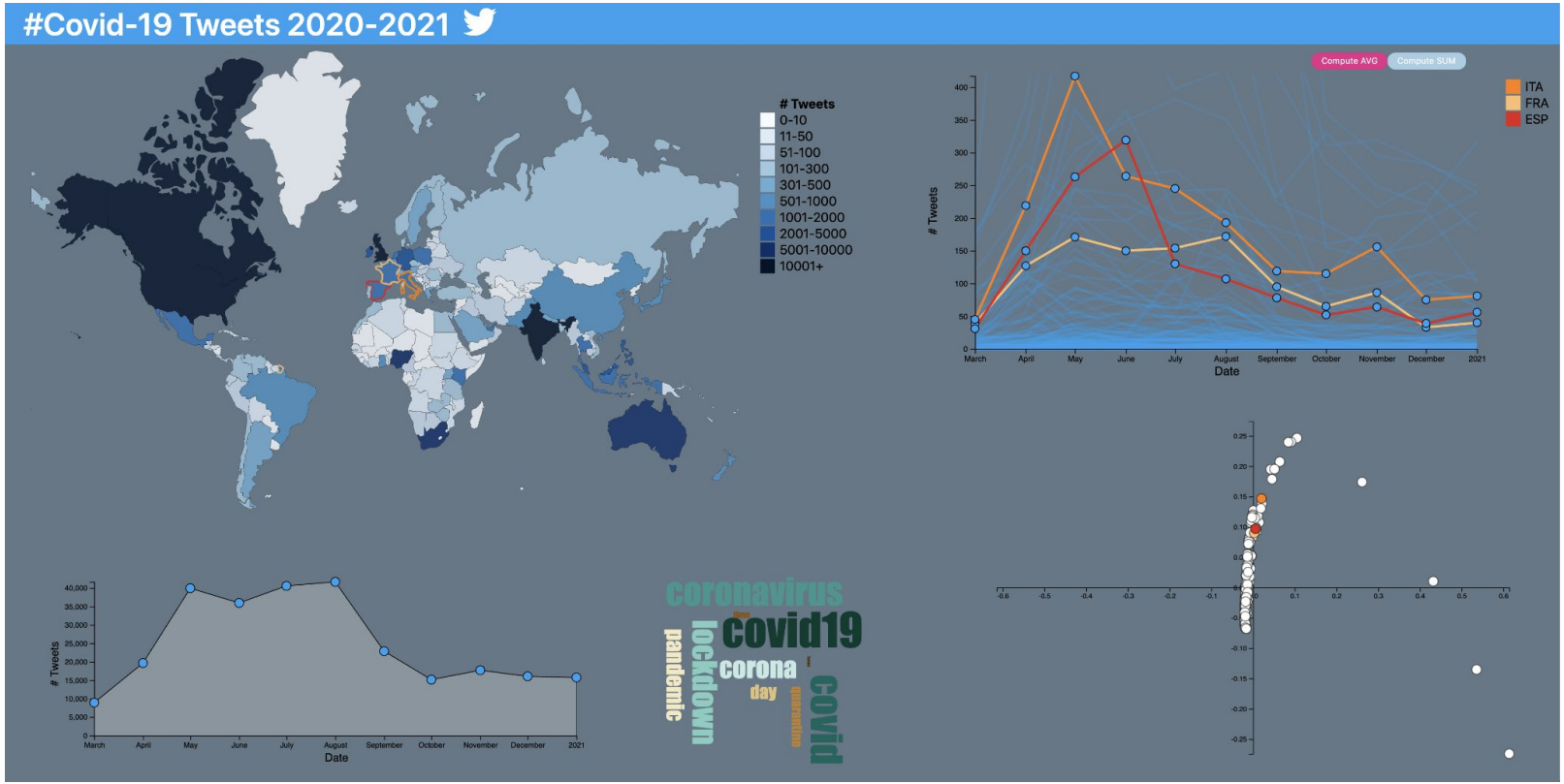


```
id  name    population
CHN China   "1330141295"
IND India   "1173108018"
USA United States   "310232863"
IDN Indonesia   "242968342"
BRA Brazil  "201103330"
PAK Pakistan    "177276594"
BGD Bangladesh  "158065841"
NGA Nigeria "152217341"
RUS Russia  "139390205"
JPN Japan   "126804433"
MEX Mexico  "112468855"
PHL Philippines "99900177"
VNM Vietnam "89571130"
ETH Ethiopia    "88013491"
DEU Germany "82282988"
EGY Egypt   "80471869"
TUR Turkey  "77804122"
```

	id	city
0	JPN	681
1	IDN	132
2	IND	433
3	PHL	1532
4	CHN	1497
5	BRA	3370
6	KOR	69
7	MEX	1025
8	EGY	108
9	USA	7823
10	RUS	1486
11	THA	392
12	ARG	384
13	BGD	18
14	NGA	77
15	TUR	195
16	PAK	108
17	IRN	222
18	COD	88
19	VNM	79
20	FRA	2016

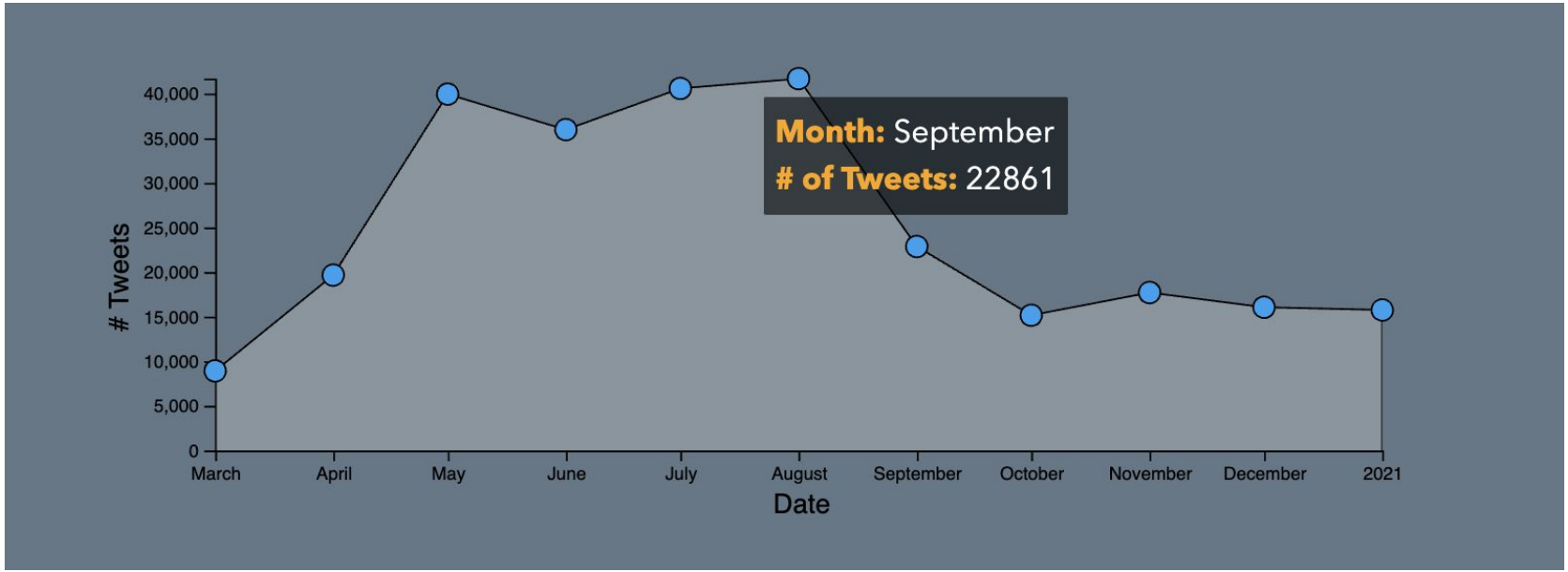
Visualization Techniques

Overview


#Covid-19 Tweets 2020-2021
Tweets
0-10
11-50
51-100
101-300
301-500
501-1000
1001-2000
2001-5000
5001-10000
10001+
Compute AVG
Compute SUM
ITA
FRA
ESP
Tweets
Date
March
April
May
June
July
August
September
October
November
December
2021
coronavirus
covid19
lockdown
corona
pandemic
day
covid
quarantine

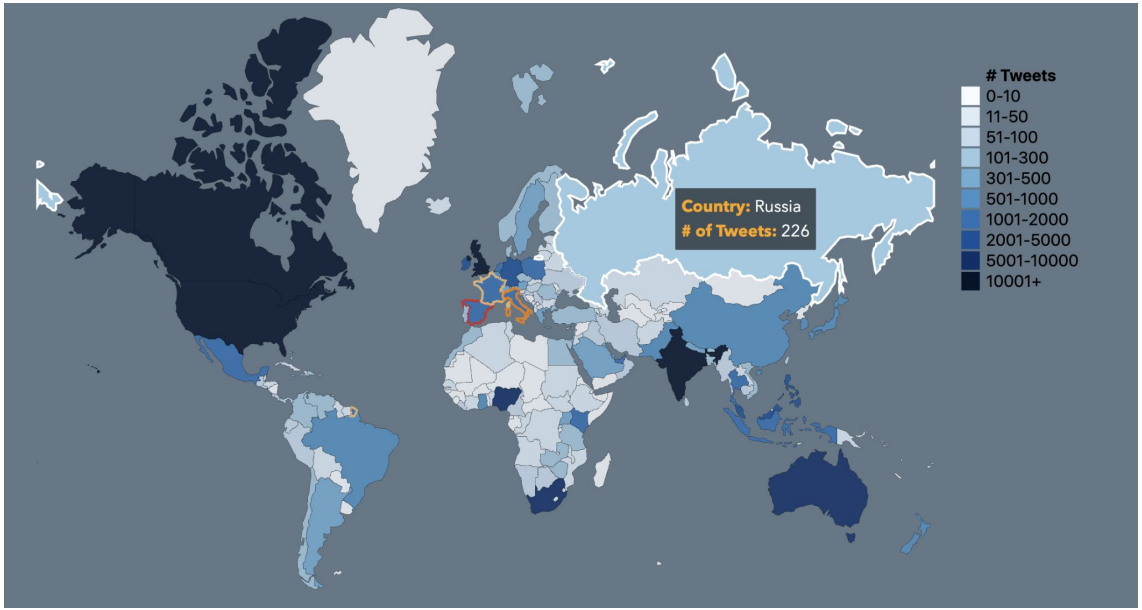
Number of tweets trend plot

- Shows the trend of number of tweets
- Brushing on the horizontal axis to select the period of interest for all the visualizations
- Tooltip on each point with the exact number of tweets
- Monthly aggregated data



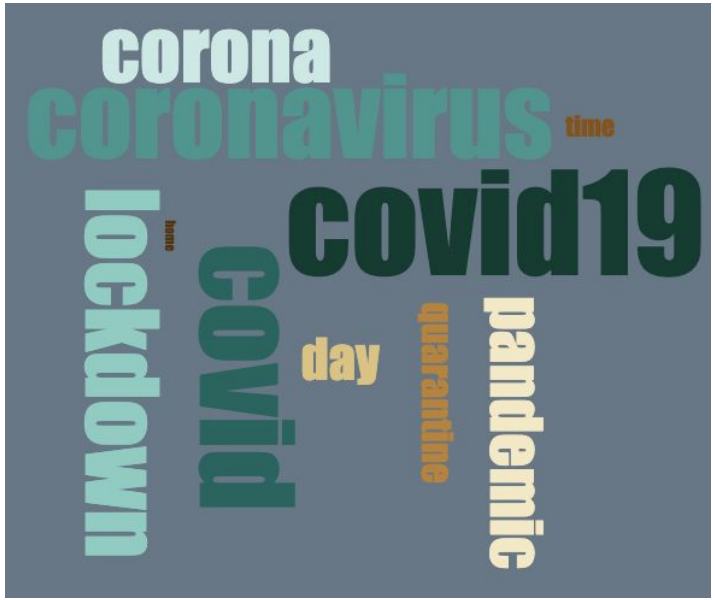
World Heatmap

- Number of tweets grouped by nation
- All nations are selectable, bidirectional interaction with MDS and nations trend plot
- Tooltip on each nation to show exact number of tweets
- Color scale shows the approximate number of tweets



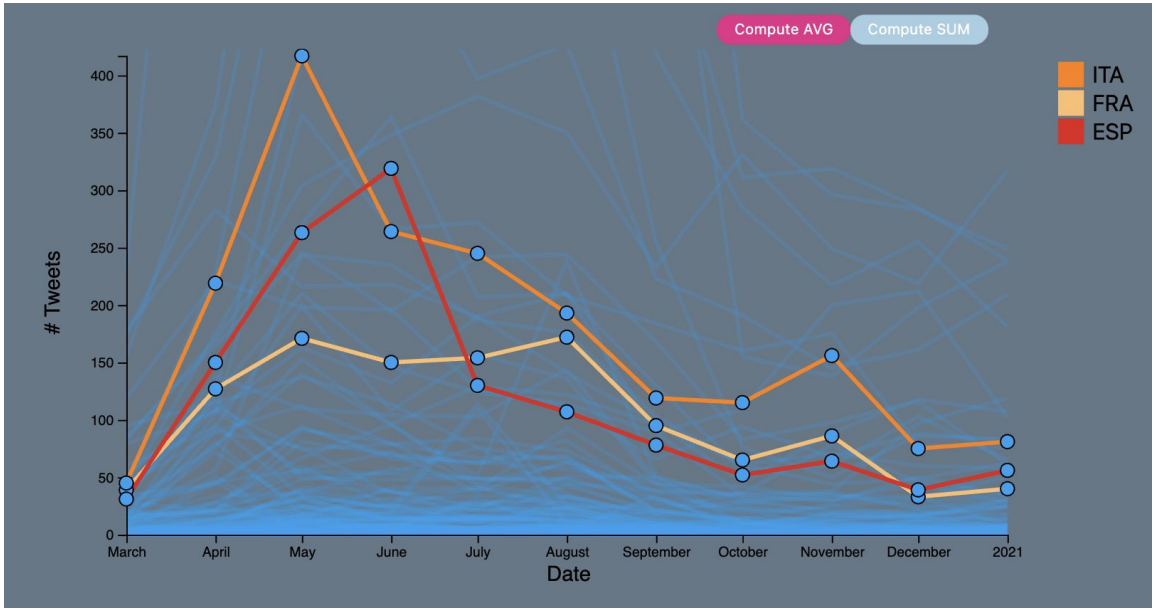
Wordcloud

- Top 10 used words in all tweets
- Bigger font size stands for more frequent word
- Color scale that represents the frequency of a word
- Interesting features coming up from the less frequent words of the word cloud



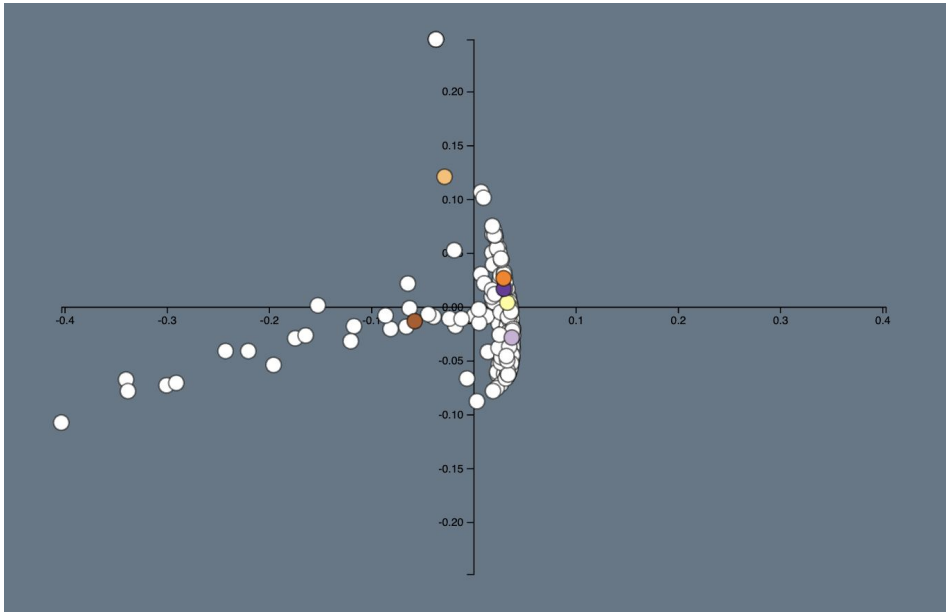
Nations trend plot

- One trend for each nation
- Selected nations are colored coherently with the other visualizations
- Number of tweets for each month of the selected interval with a tooltip that shows the exact number of tweets
- Compute average and sum of selected nations



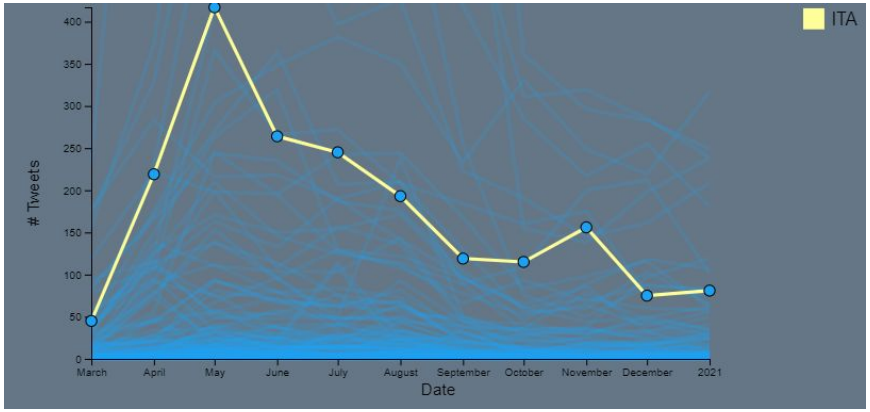
Multidimensional scaling

- Dimensionality reduction scatterplot
- Each point represents a nation, and points that are near to each other are considered similar
- Bidirectional interaction with the world map and the nations trend plot through nation selection
- Possibility of zooming by brushing

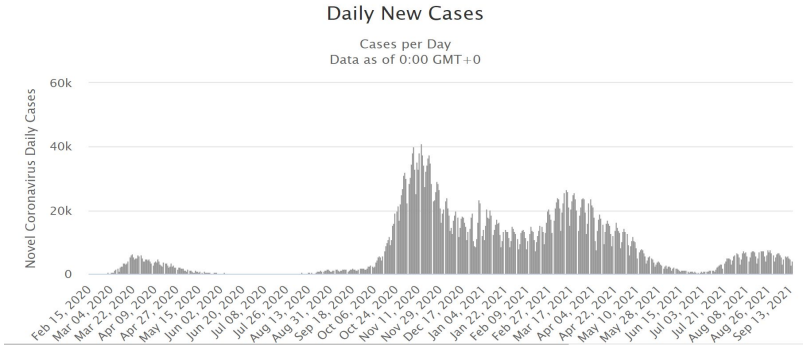


Results: Tweets Trend vs Covid new cases Curve

Italy



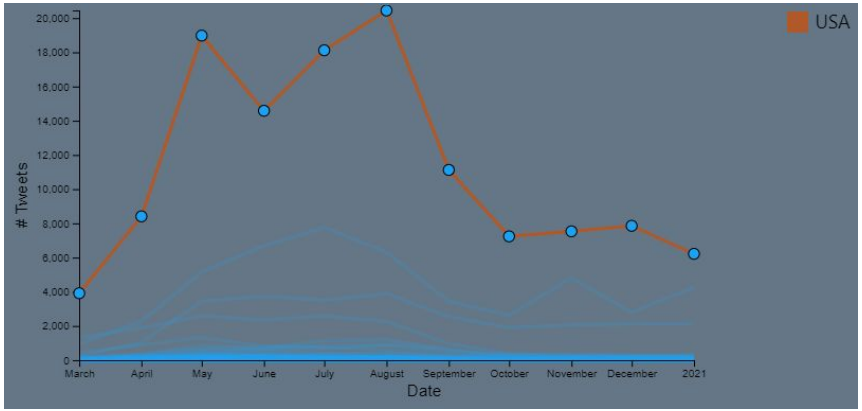
Daily New Cases in Italy

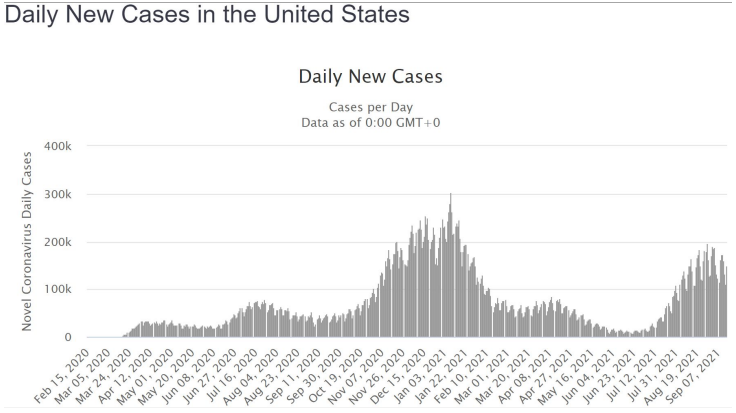


Fonte: https://www.worldometers.info/coronavirus/country/italy/

Results: Tweets Trend vs Covid new cases Curve

USA





Fonte: https://www.worldometers.info/coronavirus/country/us/

Future work

- Animated temporal axis chart that when activated, triggers the the other plots and shows sequentially data month by month, in a way in which user can better appreciate all changes.
- Include a dataset composed of tweets published in 2021 or live twitter data streaming.

Conclusion

- This tool is useful for showing the spatial and temporal trends of tweets about Covid-19
- Our period of study goes from March 2020 to January 2021
- A potential user is capable of investigating on the number of tweets on any time within our period of study discovering some insights and maybe compare his result with Covid-19 trends of infection