

## AIM & MOTIVATION



- Aim:
- To explore whether the speed of covid-19 infection will affect people's emotions in social networks.
- Motivation:
- Improve the research of epidemiological mental problems.
- Avoid the impact of disease transmission on people's mental health.

## BACKGROUND INFORMATION



- Previous Study:
  - Be able to analyze the characteristics of public opinion and the differences of concerned topics in different regions.
  - Taking COVID-19 as a case study.
- 
- Our study:
  - We analyze the impact of daily new infections and deaths of epidemic on people's emotions on social networks.
  - Taking COVID-19 as a case study.
- 
- Reference : Han K K, Xing Z Y, Liu Z, et al. Research on public opinion analysis methods in major public health events: Take COVID-19 Epidemic as an Example[J]. Journal of Geo-information Science, 2021, 23(2).]  
DOI:10.12082/dqxxkx.2021.200226

## DATA SET

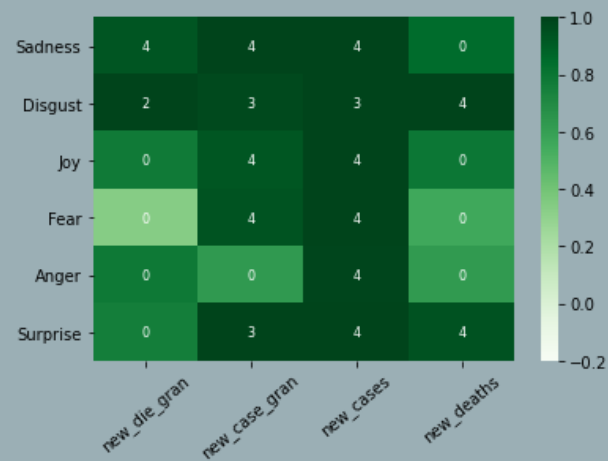
COVID-19 growth data set

Emotion data on Twitter

- COVID-19 growth data:
- We use this data set to calculate the gradient of daily new infections and death.
- Key words : daily new case, daily new death
- Emotion data on Twitter:
- We use this data set to get the proportion of each emotion.
- Key words : send time,
- 140,000 tweets about COVID-19
- References:
- tweet: <https://www.kaggle.com/gpreda/covid19-tweets>
- daily cases: <https://data.world/markmarkoh/coronavirus-data>

## METHOD

There are three main steps in our study.



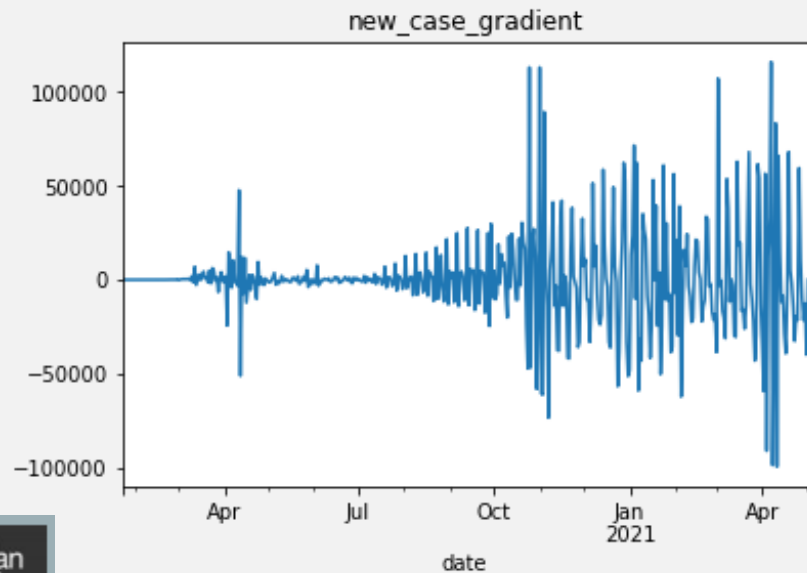
First step:

Use natural language processing to get emotions from Twitter data set.

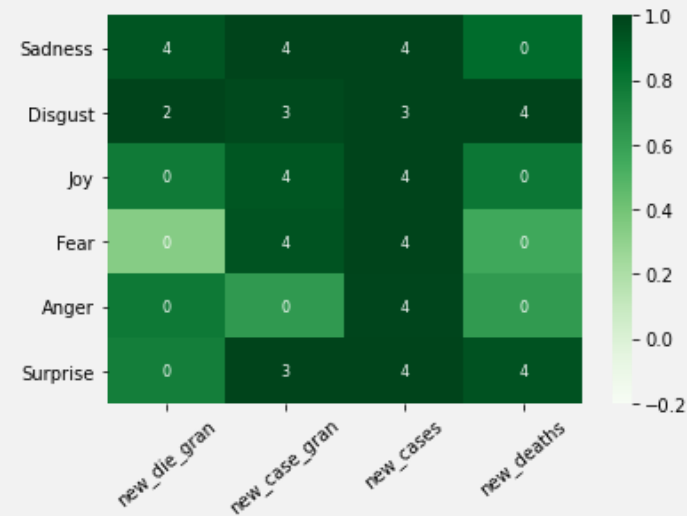
	date	Emotion	count	percent
0	2020-07-24	Anger	6	0.020339
1	2020-07-24	Disgust	3	0.010169
2	2020-07-24	Fear	99	0.335593
3	2020-07-24	Joy	116	0.393220
4	2020-07-24	Sadness	17	0.057627
...	...	...	...	...
121	2020-08-16	Disgust	49	0.006533
122	2020-08-16	Fear	2614	0.348533
123	2020-08-16	Joy	2992	0.398933
124	2020-08-16	Sadness	676	0.090133
125	2020-08-16	Surprise	1042	0.138933

# METHOD

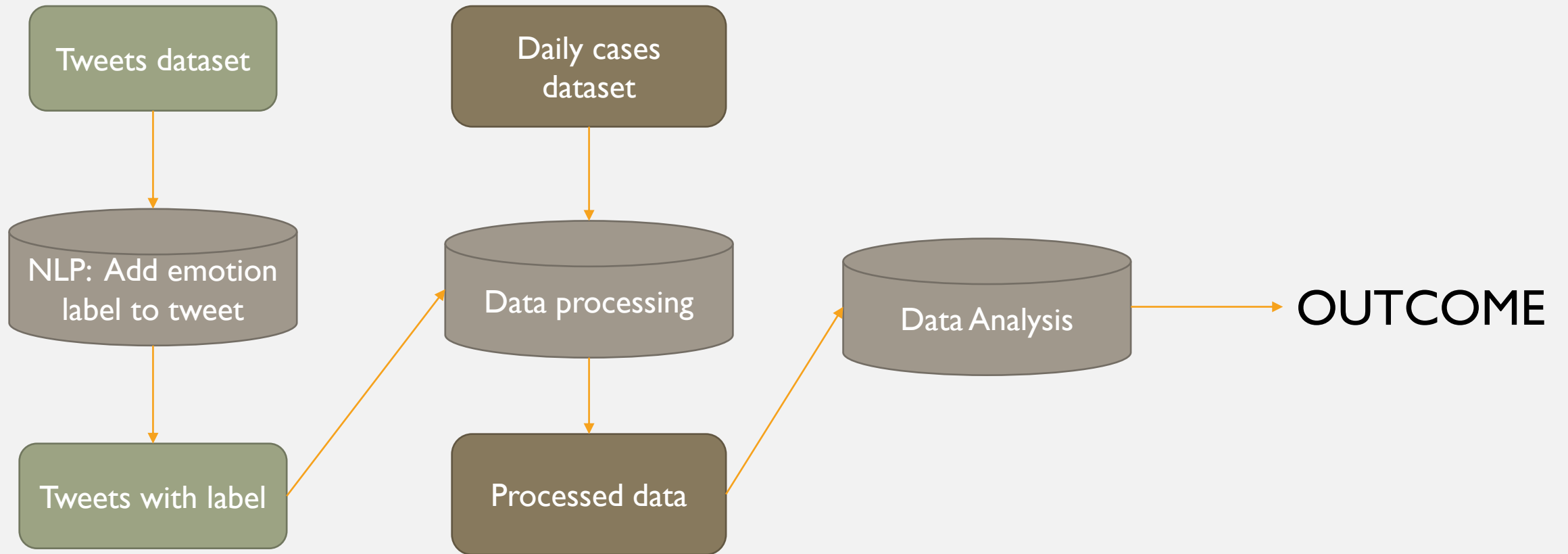
- Second step:
- We analyze the data and calculate the gradient of infections and deaths.



- Third step:
- Use statistical methods(Granger Causality) to find the causal relationship.



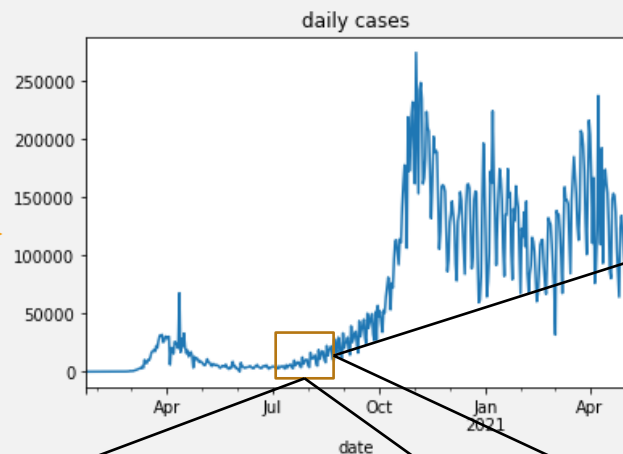
# DETAILS OF IMPLEMENTATION



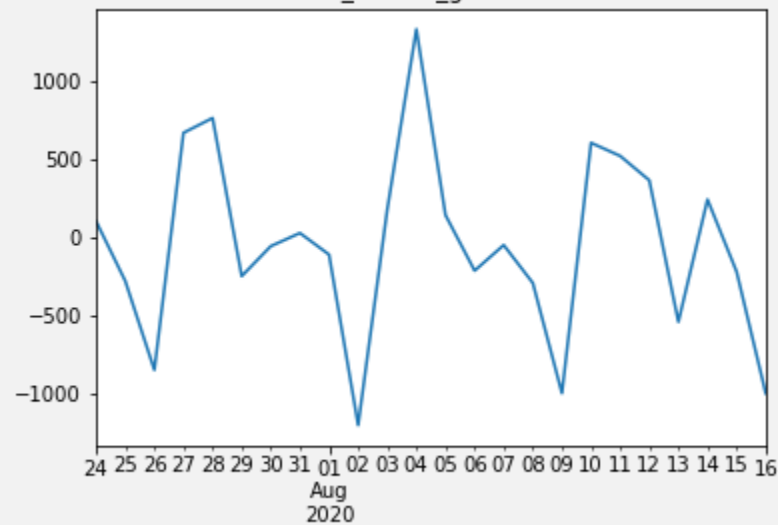
# DATA PREPROCESS

We took a short clip from July and August

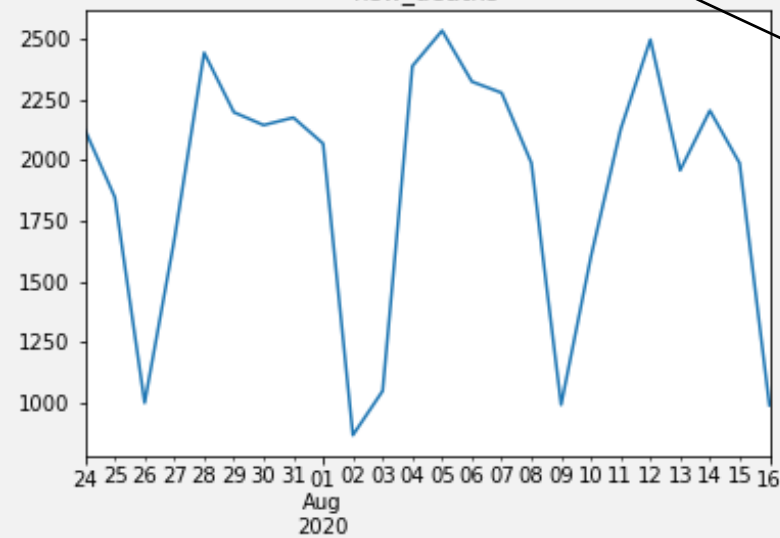
Daily cases  
dataset



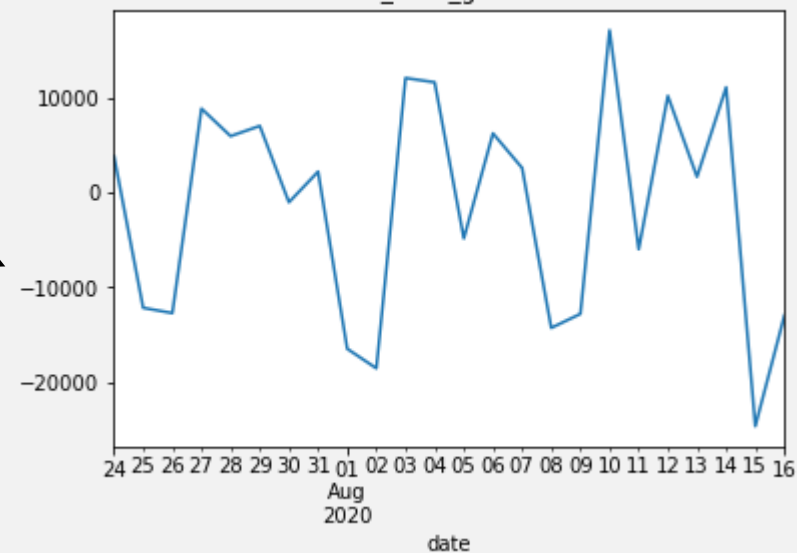
new\_deaths\_grandiant



new\_deaths

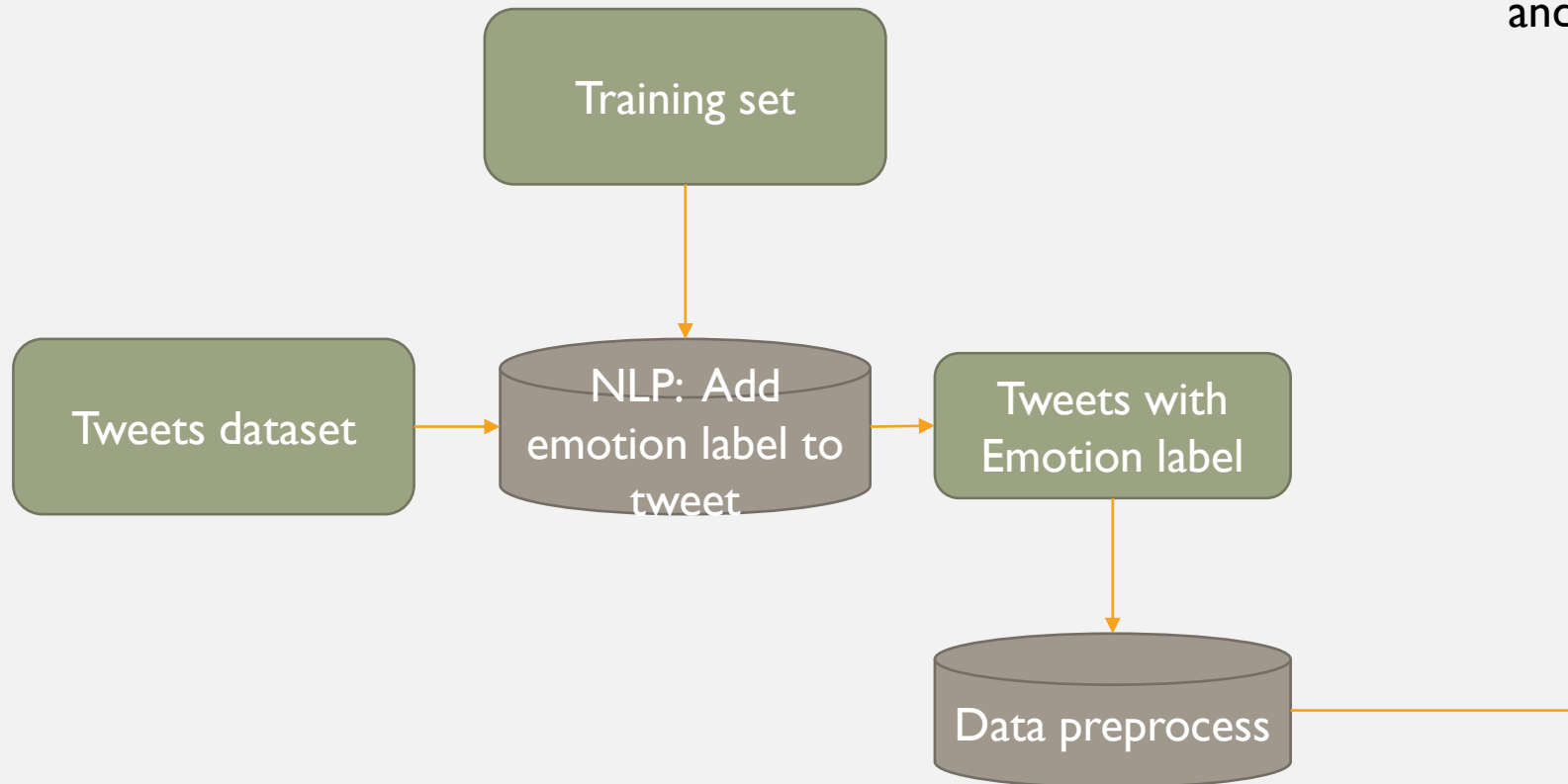


new\_case\_gradient





# DATA PREPROCESS

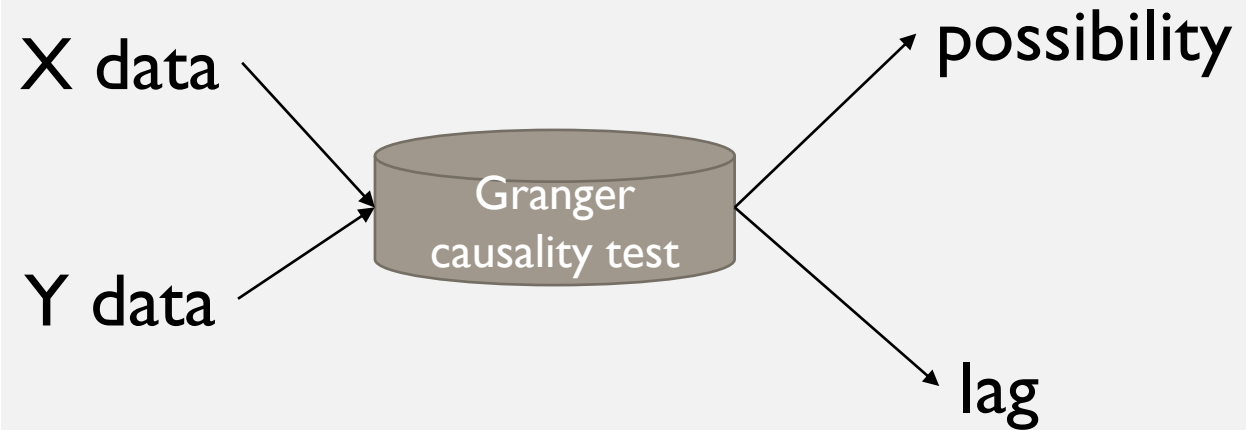
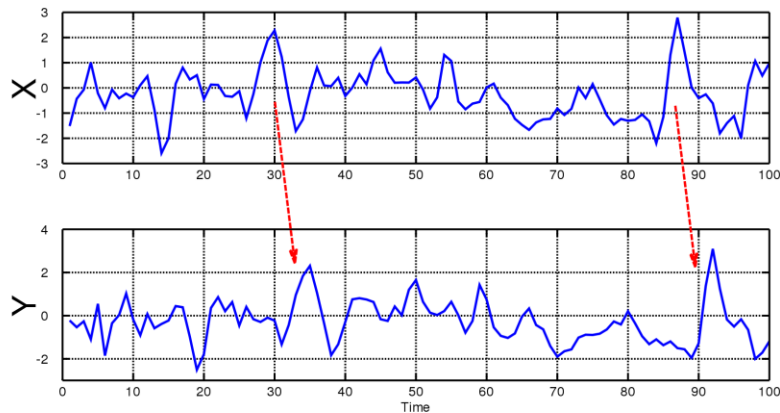


The count is the number of each mood, and the percent is proportion of each mood.

	date	Emotion	count	percent
0	2020-07-24	Anger	6	0.020339
1	2020-07-24	Disgust	3	0.010169
2	2020-07-24	Fear	99	0.335593
3	2020-07-24	Joy	116	0.393220
4	2020-07-24	Sadness	17	0.057627
...	...	...	...	...
121	2020-08-16	Disgust	49	0.006533
122	2020-08-16	Fear	2614	0.348533
123	2020-08-16	Joy	2992	0.398933
124	2020-08-16	Sadness	676	0.090133
125	2020-08-16	Surprise	1042	0.138933

# DATA ANALYSIS

## Method: Granger causality



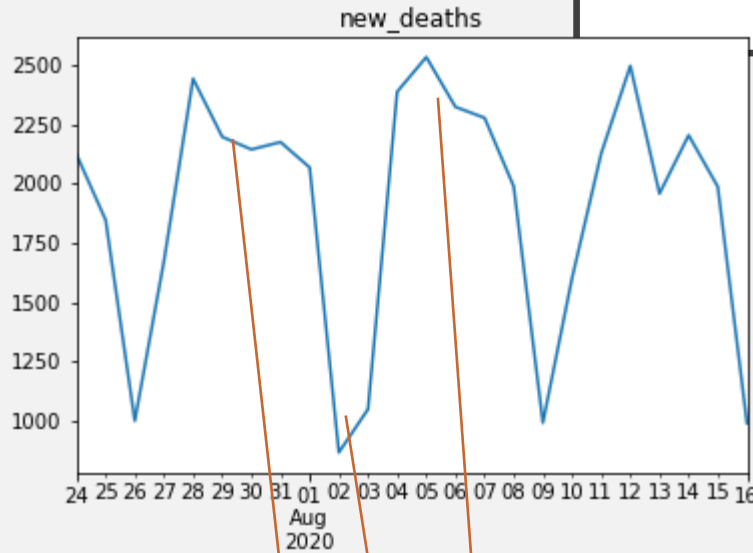
X and Y may have relationship,  
But not all at the same time. This delay is called lag



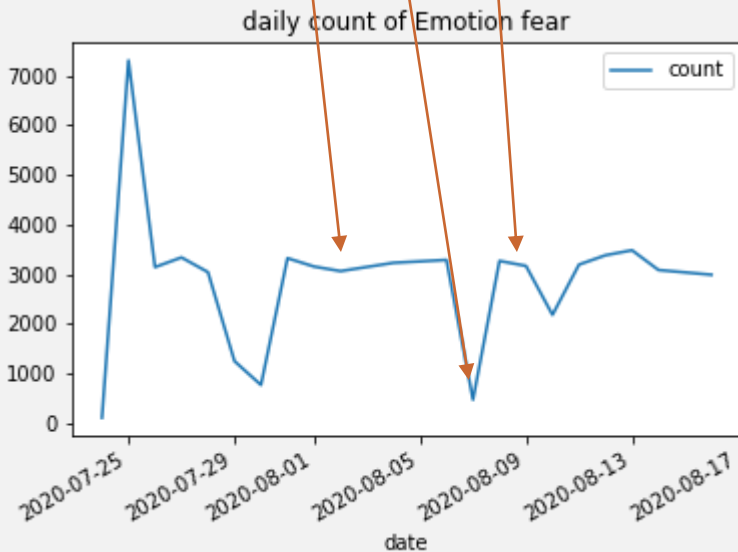
Reference: [https://en.wikipedia.org/wiki/Granger\\_causality](https://en.wikipedia.org/wiki/Granger_causality)

# EXAMPLE OUTPUT

P: the possibility of no relationship



Looks like has relationship



Granger  
causality test

```
Granger Causality
number of lags (no zero) 1
ssr based F test:      F=0.3131 , p=0.5831 , df_denom=17, df_num=1
ssr based chi2 test:   chi2=0.3684 , p=0.5439 , df=1
likelihood ratio test: chi2=0.3650 , p=0.5457 , df=1
parameter F test:      F=0.3131 , p=0.5831 , df_denom=17, df_num=1
```

```
Granger Causality
number of lags (no zero) 2
ssr based F test:      F=1.1941 , p=0.3320 , df_denom=14, df_num=2
ssr based chi2 test:   chi2=3.2412 , p=0.1978 , df=2
likelihood ratio test: chi2=2.9927 , p=0.2239 , df=2
parameter F test:      F=1.1941 , p=0.3320 , df_denom=14, df_num=2
```

```
Granger Causality
number of lags (no zero) 3
ssr based F test:      F=4.0738 , p=0.0358 , df_denom=11, df_num=3
ssr based chi2 test:   chi2=19.9985 , p=0.0002 , df=3
likelihood ratio test: chi2=13.4492 , p=0.0038 , df=3
parameter F test:      F=4.0738 , p=0.0358 , df_denom=11, df_num=3
```

```
Granger Causality
number of lags (no zero) 4
ssr based F test:      F=4.0055 , p=0.0451 , df_denom=8, df_num=4
ssr based chi2 test:   chi2=34.0465 , p=0.0000 , df=4
likelihood ratio test: chi2=18.6919 , p=0.0009 , df=4
parameter F test:      F=4.0055 , p=0.0451 , df_denom=8, df_num=4
```

```
Granger Causality
number of lags (no zero) 5
ssr based F test:      F=7.7387 , p=0.0212 , df_denom=5, df_num=5
ssr based chi2 test:   chi2=123.8192 , p=0.0000 , df=5
likelihood ratio test: chi2=34.6842 , p=0.0000 , df=5
parameter F test:      F=7.7387 , p=0.0212 , df_denom=5, df_num=5
```

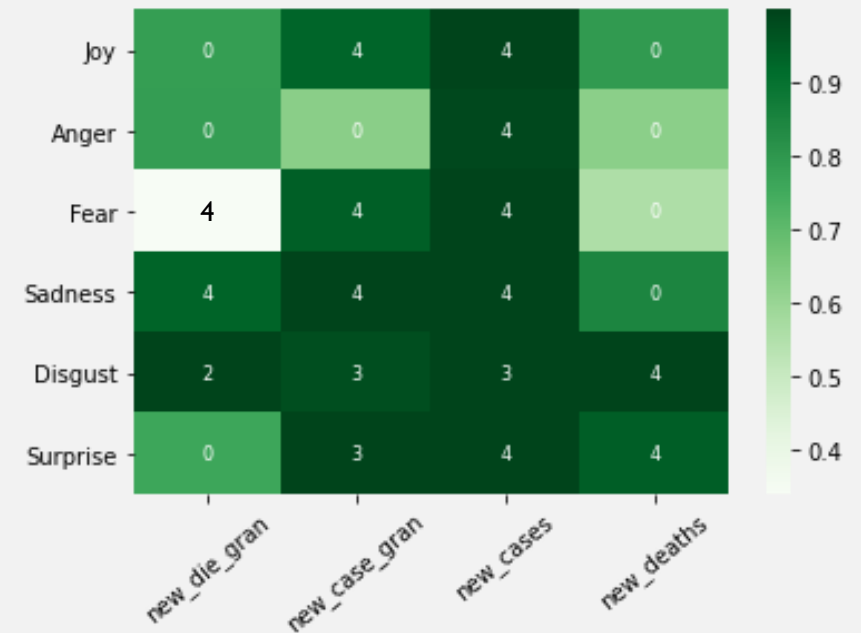
Can confirm there is relationship and lag is 5

## OUR OUTCOME

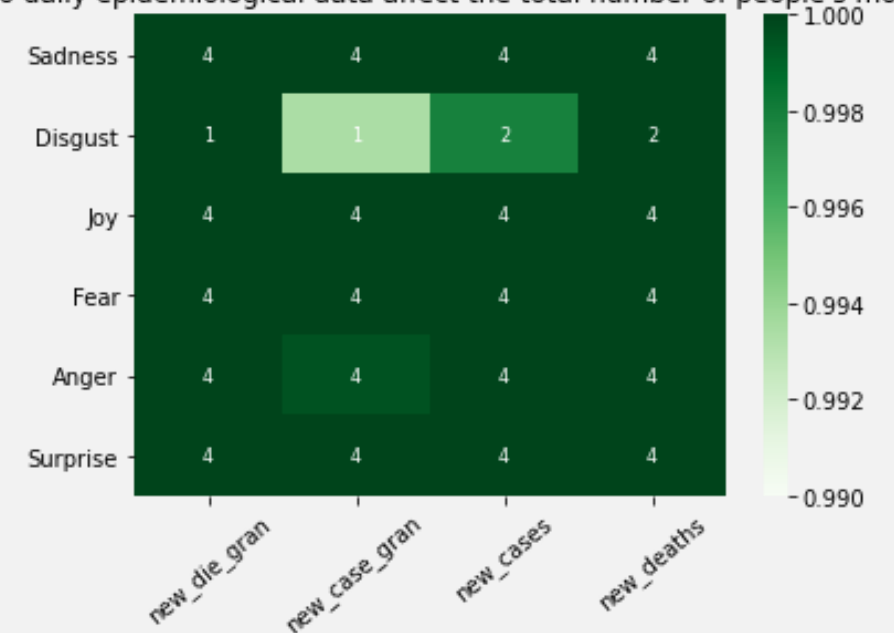
### Method: Granger causality: outcome (two image has different scaler)

- In general, when the probability is greater than 0.9, we believe that there is a causal relationship between the two variables
- The proportion of each mood does not correlate with all the data
- The total number of tweets for each mood was significantly influenced by daily cases.

How do daily epidemiological data affect the proportions of people's moods?  
The blue numbers means the time when the impact reaches its maximum



How do daily epidemiological data affect the total number of people's moods



## CONCLUSIONS AND GUESSES

### **Why don't all data cases data affect the proportion of people in each mood ?**

- Search engines often show a limited number of daily cases, making people who don't get sick pay more attention to new cases than to new deaths
- The probability of COVID-causing death is not high, so the number of deaths per day is usually much smaller than the number of new deaths per day, so the impact of the data on the number of deaths per day is overridden by the impact of other data

### **Why would all the epidemic data affect the total number of people in each mood ?**

- Daily data can affect people's moods, so of course, there is a correlation with the total number of each emotion

### **Why the difference exist?**

- Familiar people's illness can affect people's emotions more, so the total number of each mood is affected when the daily case changes, but other people don't care about strangers, so there is no impact on the percentage

THANKS FOR LISTENING