China_US_tweets_infor

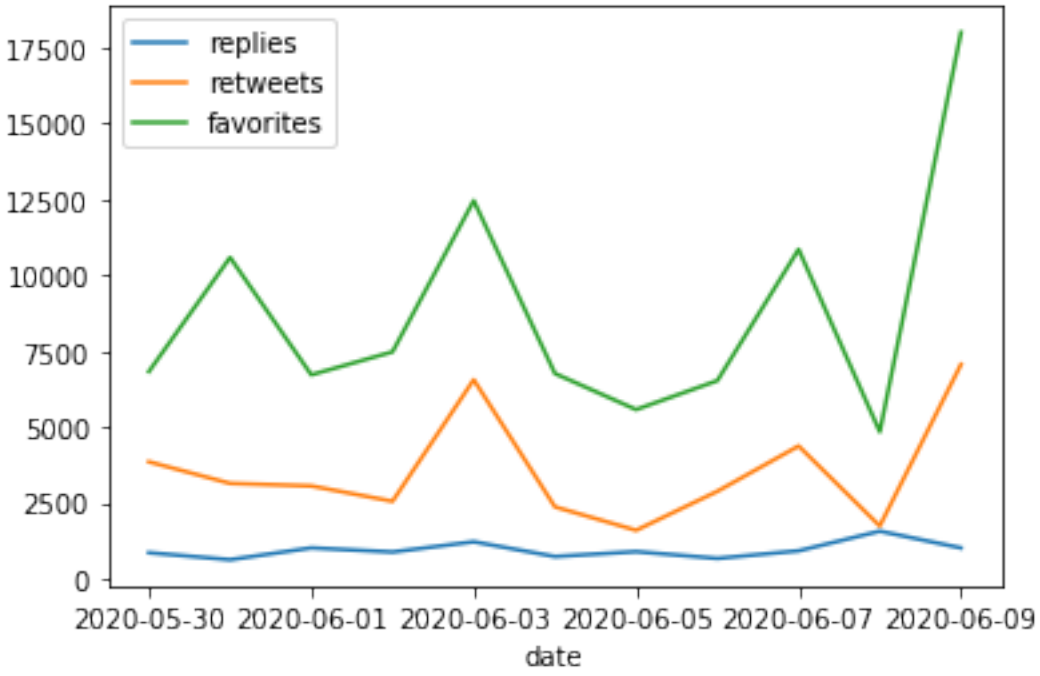
June 10, 2020

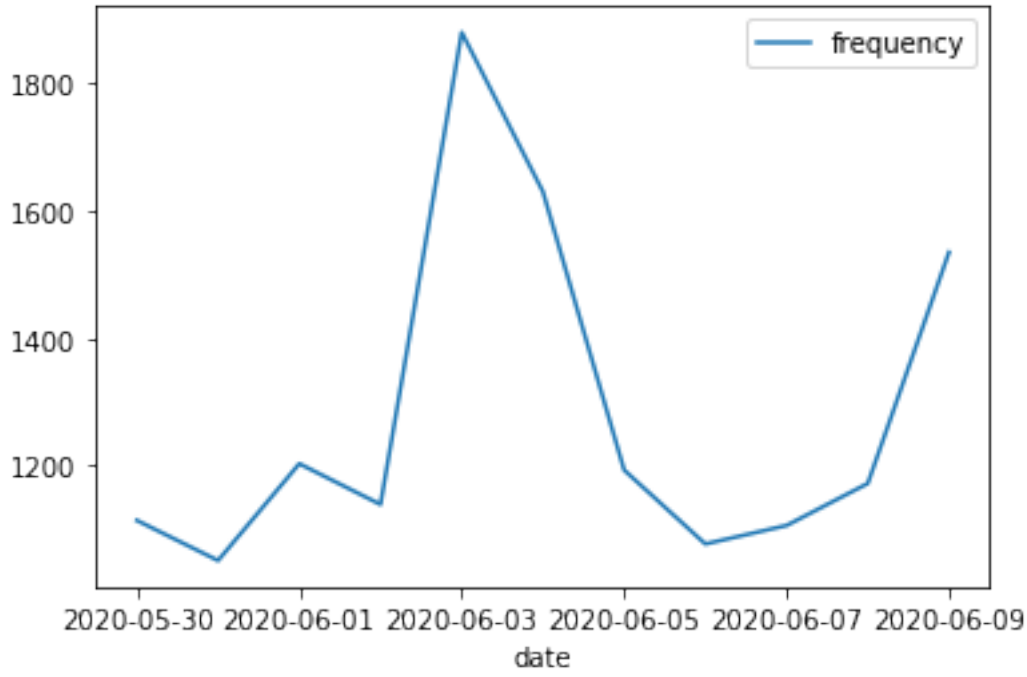
[2]:
```
import pandas

url = '/Users/min/OneDrive - The University of Texas at Dallas/UTD/Courses/EPPS␣
 ↪7V81 Advanced Data Programming/Exercise 1/ChinaUS.csv'

df = pandas.read_csv(url)

df['date'] = pandas.to_datetime(df['date'])

dfm1 = df.groupby(df['date'].dt.strftime('%Y-%m-%d')).sum()

dfm2 = df.groupby(df['date'].dt.strftime('%Y-%m-%d')).count()

dfm1 = dfm1.drop(['geo','id'], axis=1)

dfm2 = dfm2.
 ↪drop(['username','to','replies','retweets','favorites','text','geo','mentions','hashtags','
 ↪axis=1)

dfm2.columns = ['frequency']

dfm3 = pandas.concat([dfm1, dfm2], axis=1)

#dfm1['counts'] = pandas.Series(data=dfm1.groupby(dfm1['date']).count())

#counts = df.groupby(df['date'].dt.strftime('%B')).['date'].count()

print(dfm3)

dfm1.plot()
dfm2.plot()
dfm3.plot()
```

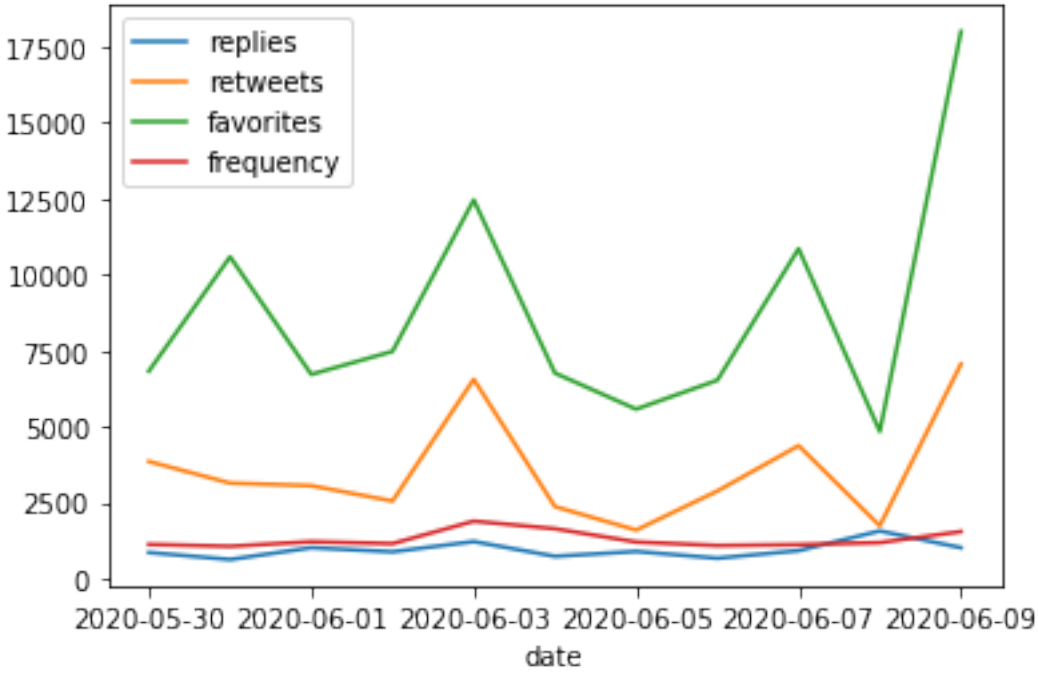
```
            replies  retweets  favorites  frequency
date
2020-05-30    847.0    3846.0     6821.0       1114
2020-05-31    616.0    3133.0    10586.0       1051
2020-06-01   1006.0    3049.0     6714.0       1203
```

2020-06-02	870.0	2543.0	7472.0	1139
2020-06-03	1208.0	6553.0	12460.0	1880
2020-06-04	721.0	2359.0	6757.0	1630
2020-06-05	882.0	1586.0	5568.0	1193
2020-06-06	659.0	2872.0	6522.0	1077
2020-06-07	909.0	4367.0	10853.0	1106
2020-06-08	1557.0	1729.0	4836.0	1172
2020-06-09	1005.0	7059.0	18011.0	1535

[2]: <matplotlib.axes._subplots.AxesSubplot at 0x7ff46941fa90>




frequency
1800
1600
1400
1200
2020-05-30
2020-06-01
2020-06-03
2020-06-05
2020-06-07
2020-06-09
date


replies
retweets
favorites
frequency
17500
15000
12500
10000
7500
5000
2500
0
2020-05-30
2020-06-01
2020-06-03
2020-06-05
2020-06-07
2020-06-09
date