The change in U.S stock market during past viral outbreaks

By: Ha My Nguyen

1. Objectives

The objective of this report is to depict how U.S. equity markets recently have experienced turbulent trade due to deadly corona virus and also look back to compare with the past condition when the world have faced another storm of epidemic diseases. The report is based on the information that being scrapped from the website https://www.marketwatch.com/story/heres-how-the-stock-market-has-performed-during-past-viral-outbreaks-as-chinas-coronavirus-spreads-2020-01-22

Firstly, this report will explore the data by printing it to get the idea of how the data look like. Further exploration will be investigated by getting all the table tag. Next, data cleaning will be performed to clean raw data and generate useful data into data frame. Afterwards, this report will contain different types of charts to compare the change in US stock market by month, year and by disease phenomenon.

Finally, conclusion will be made with several recommendations.

2. Data mining

```
In [1]: from bs4 import BeautifulSoup
```

```
In [2]: import requests
        url = 'https://www.marketwatch.com/story/heres-how-the-stock-market-has-performed-during-past-viral-outbreaks-as-china
        s-coronavirus-spreads-2020-01-22'
        response = requests.get(url)
        cnn = BeautifulSoup(response.text)
        print(cnn.prettify())
```

```
<!DOCTYPE html>
<html lang="en-US">
 <head>
  <title>
   How the stock market has performed during past viral outbreaks, as coronavirus infects 31,000 - MarketWatch
  </title>
  <link href="https://www.marketwatch.com/story/heres-how-the-stock-market-has-performed-during-past-viral-outbreaks-
as-chinas-coronavirus-spreads-2020-01-22" id="canonical-link" rel="canonical"/>
   <link href="https://www.marketwatch.com/amp/story/guid/6A955A3E-3D43-11EA-A3B8-BC9F4A046959" rel="amphtml"/>
   <link href="mwatch://article?id=6A955A3E-3D43-11EA-A3B8-BC9F4A046959" rel="alternate"/>
   <meta content="app-id=336693422,app-argument=mwatch://article?id=6A955A3E-3D43-11EA-A3B8-BC9F4A046959" name="apple-
itunes-app"/>
   <link href="https://mw4.wsj.net/mw5/content/images/favicons/apple-touch-icon.png" rel="apple-touch-icon"/>
    <link href="https://mw4.wsj.net/mw5/content/images/favicons/apple-touch-icon-152x152.png" rel="apple-touch-icon" si
zes="152x152"/>
```

```
In [3]: #get the table tag
        table = cnn.select('#article-body table')[0]
        table
```

```
Out[3]: <table> <tbody> <tr class="data"> <td align="left" colspan="" id="" valign="top"> <strong>Epidemic</strong> </td> <td
        align="left" colspan="" id="" valign="top"> <strong>Month end</strong> </td> <td align="left" colspan="" id="" valign
        ="top"> <strong>6-month % change of S&amp;P</strong> </td> <td align="left" colspan="" id="" valign="top"> <strong>12
        -month % change of S&amp;P</strong> </td> </tr> <tr class="data even"> <td align="left" colspan="" id="" valign="to
        p">HIV/AIDS</td> <td align="left" colspan="" id="" valign="top">June 1981</td> <td align="left" colspan="" id="" vali
        gn="top">-0.3</td> <td align="left" colspan="" id="" valign="top">-16.5</td> </tr> <tr class="data"> <td align="left"
        colspan="" id="" valign="top">Pneumonic plague</td> <td align="left" colspan="" id="" valign="top">September 1994</td
        > <td align="left" colspan="" id="" valign="top">8.2</td> <td align="left" colspan="" id="" valign="top">26.3</td> </
        tr> <tr class="data even"> <td align="left" colspan="" id="" valign="top">SARS</td> <td align="left" colspan="" id=""
```

Then, drop the unnecessary data and generate pandas data frame

```
In [4]: import pandas as pd
        data = table.findAll('tr')
        print("Data is a {} and {} items long".format(type(data), len(data)))
        data_without_header = data[1:-2]
        headers = data[0]
        col_headers = [th.getText().strip() for th in headers.findAll('td')]
        df_data = [[td.getText() for td in tr.findAll('td')] for tr in data_without_header] # nested list-comp for 2d struct
        df = pd.DataFrame(df_data, columns=col_headers)
        df

        Data is a <class 'bs4.element.ResultSet'> and 14 items long
```

Out[4]:

	Epidemic	Month end	6-month % change of S&P	12-month % change of S&P
0	HIV/AIDS	June 1981	-0.3	-16.5
1	Pneumonic plague	September 1994	8.2	26.3
2	SARS	April 2003	14.59	20.76
3	Avian flu	June 2006	11.66	18.36
4	Dengue Fever	September 2006	6.36	14.29
5	Swine flu	April 2009	18.72	35.96
6	Cholera	November 2010	13.95	5.63
7	MERS	May 2013	10.74	17.96
8	Ebola	March 2014	5.34	10.44
9	Measles/Rubeola	December 2014	0.20	-0.73
10	Zika	January 2016	12.03	17.45

Next, convert the object type into float

```
In [5]: import matplotlib.pyplot as plt
        df['6-month % change of S&P'] = df['6-month % change of S&P'].astype(float)
        df['12-month % change of S&P'] = df['12-month % change of S&P'].astype(float)
        df.dtypes
```
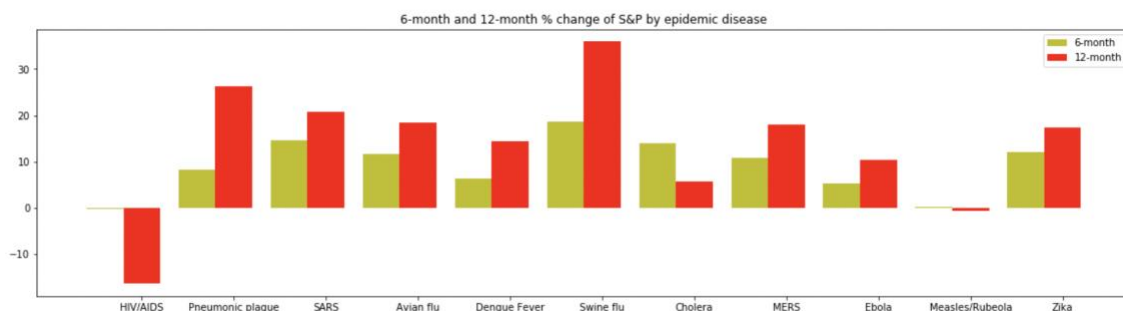
```
Out[5]: Epidemic                      object
        Month end                     object
        6-month % change of S&P       float64
        12-month % change of S&P      float64
        dtype: object
```

3. Plot

3.1   Bar chart

```
In [6]: plt.figure(figsize = (20,5))
        x = list(range(len(df['Epidemic'])))
        total_width, n = 0.8, 2
        width = total_width / n
        plt.bar(x, df['6-month % change of S&P'],width=width,label='6-month',fc = 'y')
        for i in range(len(x)):
            x[i] = x[i] + width
        plt.bar(x, df['12-month % change of S&P'],width=width,label='12-month',tick_label=df['Epidemic'],fc = 'r')
        plt.title('6-month and 12-month % change of S&P by epidemic disease')
        plt.legend()
```

```
Out[6]: <matplotlib.legend.Legend at 0x1148fa978>
```
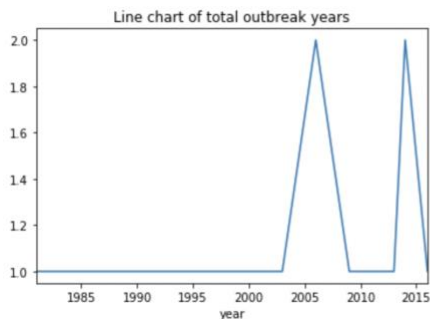
According to the chart, Swine flu seems likely made highest influence in percentage change (raise up) of the US stock market during both timing period of 6-month and 12-month, however HIV/AIDS was showing negative growth.

3.2   Line chart

```
In [8]: dic = {'January': '1',
        'February': '2',
        'March': '3',
        'April': '4',
        'May': '5',
        'June': '6',
        'July': '7',
        'August': '8',
        'September': '9',
        'October': '10',
        'November': '11',
        'December': '12'}
        def func(x):
            month = dic[x.split(' ')[0]]
            year = x.split(' ')[1]
            return year + '-' + month
        df['format-year-month'] = df['Month end'].apply(func)
        df['format-year-month'] = pd.to_datetime(df['format-year-month'], format = '%Y-%m')
        df['year'] = df['format-year-month'].dt.year
        df['month'] = df['format-year-month'].dt.month
        yearEpidemicGroup = df.groupby('year')['Epidemic'].count()
        yearEpidemicGroup.plot()
        plt.title('Line chart of total outbreak years')
```

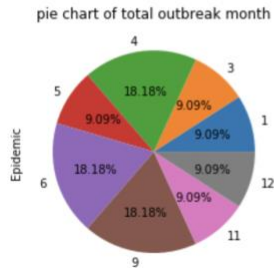Out[8]: Text(0.5, 1.0, 'Line chart of total outbreak years')

The chart indicates that 2002-2003 and 2006, the US equity market reached benchmark. Particularly, the S&P 500 posted a gain of 14.59% after the first occurrence of SARS back in 2002-03. After 12 months, broad-market benchmark was up 20.76%. After that the stock marker sharply decreased until the year 2006, the S&P 500 once again raised 11.66% in the roughly six months following reports for H5N1 pandemic. However, it hugely declined the period after.

3.3   Pie chart

```
In [9]: monthEpidemicGroup = yearEpidemicGroup = df.groupby('month')['Epidemic'].count()
        monthEpidemicGroup.plot(kind='pie', autopct='%.2f%%')
        plt.title('pie chart of total outbreak month')
```

```
Out[9]: Text(0.5, 1.0, 'pie chart of total outbreak month')
```
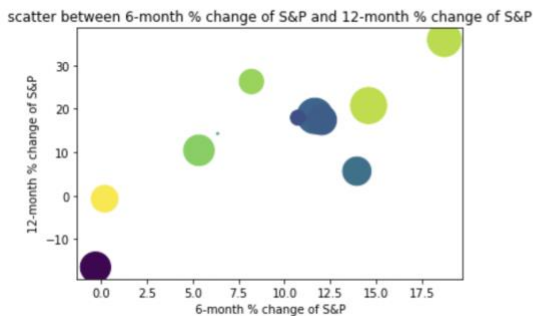
It is likely that April, June, and September are the epidemic break out seasons which contribute to increase total stock market (18.18%)

3.4   Scatter plot

```
In [11]: import numpy as np
         l = len(df['6-month % change of S&P'])
         z = np.random.rand(l)
         colors = np.random.rand(l)
         plt.scatter(x=df['6-month % change of S&P'], y=df['12-month % change of S&P'], s=z*1000,c=colors)
         plt.xlabel('6-month % change of S&P')
         plt.ylabel('12-month % change of S&P')
         plt.title('scatter between 6-month % change of S&P and 12-month % change of S&P')
```

```
Out[11]: Text(0.5, 1.0, 'scatter between 6-month % change of S&P and 12-month % change of S&P')
```

It can be seen that the larger the value of 6-month% change of S & P, the larger the trend of 12-month% change of S & P.

4. Findings

This report has investigated several facts that:

- Swine flu seems likely made highest influence in percentage change (raise up) of the US stock market
- April, June, and September are likely become frequent months of virus outbreaks
- Back in 2002-2003, SARS has been recognized as dangerous epidemic, which resulted in a total of about 8,100 people being sickened during the 2003 outbreak, with 774 people dying. However, the S&P 500 posted a gain of 14.59% after the first occurrence.
- In the year 2006, the S&P 500 once again raised 11.66% in the roughly six months following reports for H5N1 pandemic.
- Indeed, the past epidemic and current corona virus caused potential disruptions and economic decline in China

5. Recommendation

There are always potentials dangerous epidemic out there. We need to pay more attention to hygiene, usually strengthen exercise and increase autoimmunity, especially stay away from bat. It is estimated that the period of time is intensive and the virus is easily transmitted during the school year: February, July, August, and October. Swine flu can be seen as causing the strongest change in S&P, while HIV caused the lowest change in S&P, but the highest mortality rate caused by HIV.