Finance Project

June 10, 2023

1 Group Information:

	Full Name	Student ID
1	Nguyen Dang Khoa	2115113125
...	...	...

2 Project Information

In this data project we will focus on exploratory data analysis of stock prices.

We'll focus on bank stocks and see how they progressed throughout the financial crisis all the way to early 2016.

2.1 The Imports

[1]:
```
import os
import pandas_datareader as pdr
import pandas as pd
import numpy as np
import datetime as dt
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
sns.set_style('whitegrid')
import cufflinks
import chart_studio
import chart_studio.tools as tls
chart_studio.tools.set_credentials_file(username='khoadangnguyen1810',
                                        api_key='djHh5Rpz9wC2YuyyZhWD')
import warnings
warnings.filterwarnings('ignore')
```

2.2 Data

1. Getting stock data for the following banks, using pandas datareader:

- Bank of America (BAC)
- CitiGroup (C)
- Goldman Sachs (GS)
- JPMorgan Chase (JPM)
- Morgan Stanley (MS)

- Wells Fargo (WFC)

[3]:
```
#Bank of America
BAC = pdr.get_data_tiingo('BAC', start='2006-01-01', end='2016-01-01',
 ↪api_key='55aa5d43f48dfc9f241463896b8dc6d96b68284f')
BAC.drop(['adjClose','adjHigh','adjLow','adjOpen',
          'adjVolume','divCash','splitFactor'], axis =1, inplace = True)
BAC.reset_index(inplace=True)
BAC.drop('symbol', axis=1, inplace = True)
BAC.set_index('date', inplace = True)
```

[4]:
```
#CitiGroup
C = pdr.get_data_tiingo('C', start='2006-01-01', end='2016-01-01',
                        api_key='55aa5d43f48dfc9f241463896b8dc6d96b68284f')
C.drop(['adjClose','adjHigh','adjLow','adjOpen',
        'adjVolume','divCash','splitFactor'], axis =1, inplace = True)
C.reset_index(inplace=True)
C.drop('symbol', axis=1, inplace = True)
C.set_index('date', inplace = True)
```

[5]:
```
#Goldman Sachs
GS = pdr.get_data_tiingo('GS', start='2006-01-01', end='2016-01-01',
                         api_key='55aa5d43f48dfc9f241463896b8dc6d96b68284f')
GS.drop(['adjClose','adjHigh','adjLow','adjOpen',
         'adjVolume','divCash','splitFactor'], axis =1, inplace = True)
GS.reset_index(inplace=True)
GS.drop('symbol', axis=1, inplace = True)
GS.set_index('date', inplace = True)
```

[6]:
```
#JPMorgan Chase
JPM = pdr.get_data_tiingo('JPM', start='2006-01-01', end='2016-01-01',
 ↪api_key='55aa5d43f48dfc9f241463896b8dc6d96b68284f')
JPM.drop(['adjClose','adjHigh','adjLow','adjOpen',
          'adjVolume','divCash','splitFactor'], axis =1, inplace = True)
JPM.reset_index(inplace=True)
JPM.drop('symbol', axis=1, inplace = True)
JPM.set_index('date', inplace=True)
```

[7]:
```
#Morgan Stanley
MS = pdr.get_data_tiingo('MS', start='2006-01-01', end='2016-01-01',
                         api_key='55aa5d43f48dfc9f241463896b8dc6d96b68284f')
MS.drop(['adjClose','adjHigh','adjLow','adjOpen',
         'adjVolume','divCash','splitFactor'], axis =1, inplace = True)
MS.reset_index(inplace=True)
MS.drop('symbol', axis=1, inplace = True)
MS.set_index('date', inplace = True)
```

[8]:
```
#Wells Fargo
WFC = pdr.get_data_tiingo('WFC', start='2006-01-01', end='2016-01-01',
 ↪api_key='55aa5d43f48dfc9f241463896b8dc6d96b68284f')
```

```
WFC.drop(['adjClose','adjHigh','adjLow','adjOpen',
          'adjVolume','divCash','splitFactor'], axis =1, inplace = True)
WFC.reset_index(inplace=True)
WFC.drop('symbol', axis=1, inplace = True)
WFC.set_index('date', inplace = True)
```

2. List of the bank ticker.

[2]:
```
bank_ticker = ['BAC', 'C', 'GS', 'JPM', 'MS', 'WFC']
```

3. Concatenate the bank dataframes together to a single data frame called bank_stocks.

[9]:
```
bank_stocks = pd.concat([BAC, C, GS, JPM, MS, WFC],axis=1,keys=bank_ticker)
```

[10]:
```
bank_stocks.columns.names = ['bank ticker','stock info']
```

[11]:
```
bank_stocks.head()
```

[11]:
```
bank ticker                 BAC                                     C       ␣
↪  \
stock info                close   high    low   open    volume  close   high
date
2006-01-03 00:00:00+00:00  47.08  47.18  46.15  46.92  16296700  49.29  49.38
2006-01-04 00:00:00+00:00  46.58  47.24  46.45  47.00  17757900  48.38  49.10
2006-01-05 00:00:00+00:00  46.64  46.83  46.32  46.58  14970700  48.62  48.78
2006-01-06 00:00:00+00:00  46.57  46.91  46.35  46.80  12599800  48.62  48.90
2006-01-09 00:00:00+00:00  46.60  46.97  46.36  46.72  15619400  48.39  48.74

bank ticker                                      ...     MS                ␣
↪  \
stock info                 low   open    volume  ...  close   high      low
date                                             ...
2006-01-03 00:00:00+00:00  48.11  49.00  15376000  ...  58.31  58.49  56.7400
2006-01-04 00:00:00+00:00  48.35  48.86  18709000  ...  58.35  59.28  58.3500
2006-01-05 00:00:00+00:00  48.40  48.44  11431000  ...  58.51  58.59  58.0200
2006-01-06 00:00:00+00:00  48.20  48.88  13702000  ...  58.57  58.85  58.0500
2006-01-09 00:00:00+00:00  48.30  48.60  16807000  ...  59.19  59.29  58.6244

bank ticker                                  WFC
stock info                 open   volume  close   high    low   open  ␣
↪volume
date
2006-01-03 00:00:00+00:00  57.17  5377000  63.80  63.95  62.39  63.20  ␣
↪5508200
2006-01-04 00:00:00+00:00  58.70  7977800  63.06  63.64  62.73  63.60  ␣
↪5435000
2006-01-05 00:00:00+00:00  58.55  5778000  62.99  63.11  62.62  63.00  ␣
↪5079000
2006-01-06 00:00:00+00:00  58.77  6889800  63.36  63.55  62.77  63.16  ␣
↪4201900
2006-01-09 00:00:00+00:00  58.63  4144500  63.35  63.65  63.11  63.35  ␣
↪2809800

[5 rows x 30 columns]
```

3 EDA

Let's explore the data a bit!

1 .Max Close price for each bank's stock throughout the time period

[12]:
```
bank_stocks.xs(key='close',axis=1,level='stock info').max()
```

[12]:
```
bank ticker
BAC     54.90
C       60.34
GS     247.92
JPM     70.08
MS      89.30
WFC     73.00
dtype: float64
```

2. Create a new DataFrame called returns. This dataframe will contain the returns for each bank's stock. Returns are typically defined by:

$$r_t = \frac{p_t - p_{t-1}}{p_{t-1}} = \frac{p_t}{p_{t-1}} - 1$$

[13]:
```
returns = pd.DataFrame()
for x in bank_ticker:
    returns[x+' return']=bank_stocks[x]['close'].pct_change()
returns.head()
```

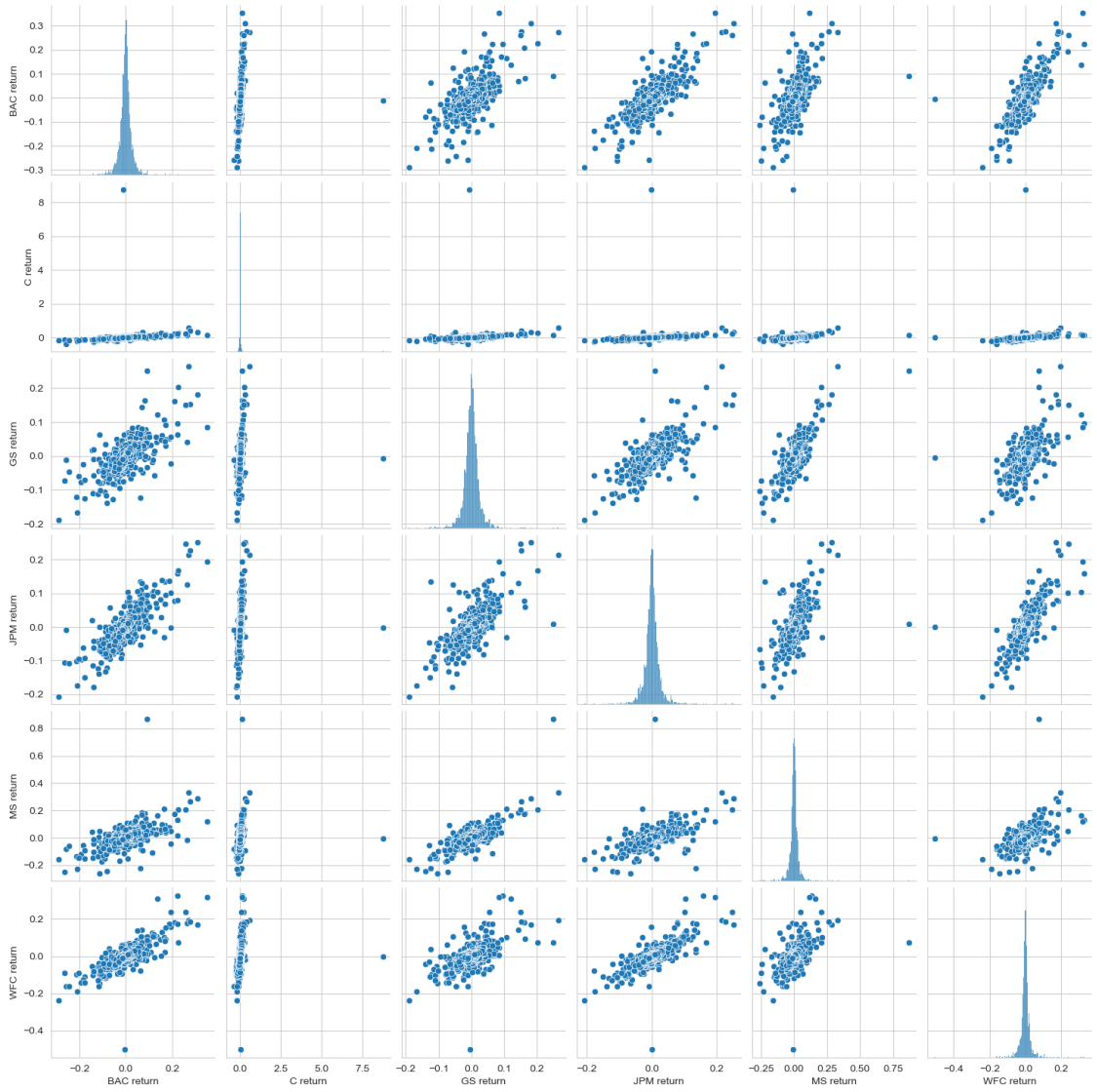
[13]:

date	BAC return	C return	GS return	JPM return	MS return	WFC return
2006-01-03 00:00:00+00:00	NaN	NaN	NaN	NaN	NaN	NaN
2006-01-04 00:00:00+00:00	-0.010620	-0.018462	-0.013812	-0.014183	0.000686	-0.011599
2006-01-05 00:00:00+00:00	0.001288	0.004961	-0.000393	0.003029	0.002742	-0.001110
2006-01-06 00:00:00+00:00	-0.001501	0.000000	0.014169	0.007046	0.001025	0.005874
2006-01-09 00:00:00+00:00	0.000644	-0.004731	0.012030	0.016242	0.010586	-0.000158

3 .Create a pairplot of the returns dataframe to see what stock stands out.

[14]:
```
sns.pairplot(returns[1:])
```

[14]:
```
<seaborn.axisgrid.PairGrid at 0x241a18e87f0>
```



We can see that stock C seems to stand out the most. Because its return pattern is flat and not similar to others'.

3. Check on what dates each bank stock had the best and worst single day returns.

[15]:
```
returns.idxmax()
```

[15]:
```
BAC return   2009-04-09 00:00:00+00:00
C return     2011-05-09 00:00:00+00:00
GS return    2008-11-24 00:00:00+00:00
JPM return   2009-01-21 00:00:00+00:00
MS return    2008-10-13 00:00:00+00:00
WFC return   2008-07-16 00:00:00+00:00
dtype: datetime64[ns, UTC]
```

[16]:
```
returns.idxmin()
```

```
[16]: BAC return   2009-01-20 00:00:00+00:00
      C return     2011-05-06 00:00:00+00:00
      GS return    2009-01-20 00:00:00+00:00
      JPM return   2009-01-20 00:00:00+00:00
      MS return    2008-10-09 00:00:00+00:00
      WFC return   2006-08-14 00:00:00+00:00
      dtype: datetime64[ns, UTC]
```

BAC, GS and JPM share the same day, 2009-01-20, for the worst drop. On Jan 20 2009, there are some events:

- Barack Obama, inaugurated as the 44th President of the United States of America, becomes the United States' first African-American president. Where Was the Dow Jones When Obama Took Office?

  Before Obama took office, the stock market went through a severe recession due to the global financial crisis in 2008. During that time, the Dow Jones Industrial Average (DJIA) reached its lowest point at 7,949 points on March 9, 2009. The article emphasizes that the significant decline of the Dow Jones before Obama took office was not due to his policies or decisions but rather the economic issues he inherited when assuming the presidency. Investors and consumers worldwide were affected by the ongoing financial crisis and economic recession.

- The subprime mortgage crisis also had a mayor part in the decline of prices.

  The subprime mortgage crisis was a global financial crisis that occurred in the late 2000s and early 2010s. It originated from the US housing market, specifically the high-risk subprime mortgages. Financial institutions created complex financial products to distribute risk and attract investors. However, when the economic downturn hit, many borrowers were unable to repay their loans, leading to a collapse in the housing market and widespread financial instability. Governments and central banks had to intervene to stabilize the financial system, as banks and major institutions faced the risk of bankruptcy.

4. Citigroup's largest drop and biggest gain seems to be very close to one another, let see if there were any events happened in that time frame.

```
[17]: bank_stocks['C'].loc['2007-01-01':'2015-01-01'].iplot(kind='candle')
```

```
[17]: <IPython.lib.display.IFrame at 0x241a710b8e0>
```

[ ]:
```
bank_stocks['C'].loc['2011-05-06':'2011-05-09']
```

[ ]:
```
stock info                  close   high    low   open     volume
date
2011-05-06 00:00:00+00:00    4.52   4.58   4.50   4.55   51316000
2011-05-09 00:00:00+00:00   44.16  45.12  43.85  44.89  491681000
```

For a brief period, the financial giant saw its stock trade below $1 per share, and even after many of its peers had fully recovered from the crisis, Citigroup did **a 10-for-1 reverse split in 2011** to get its stock price back into double digits.

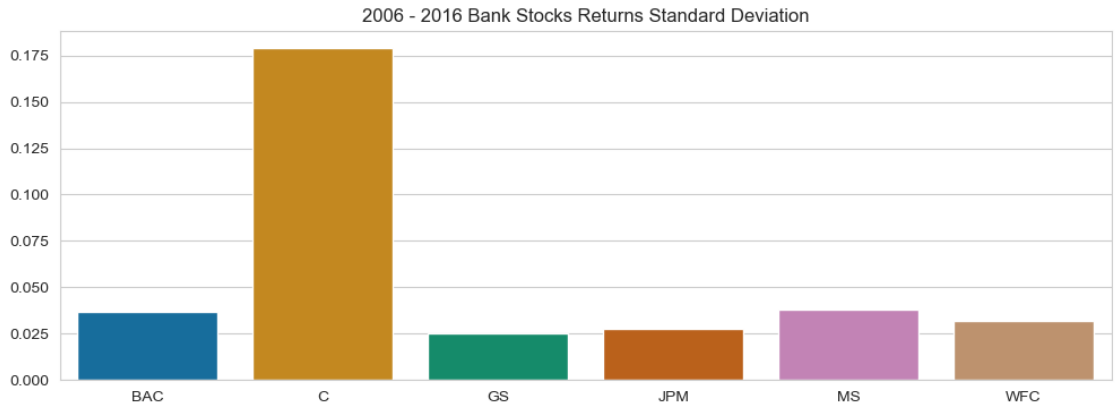
> **A reverse stock split** is a measure taken by a public company to reduce its number of outstanding shares in the market. Existing shares are consolidated into fewer shares. This results in a higher stock price for the stock shares but has no immediate effect on the total value of the stock to the investor or the market capitalization of the stock. **For example**: if a stock is trading at 50 cents on the market, and the company declares a two-for-one reverse stock split, an investor who owned 100 shares worth 50 cents would own 50 shares worth $1 each.

5. Take a look at the standard deviation of the returns to see which stocks are the riskiest over the entire time period and the riskiest for the year 2015

[18]:
```
print(returns.std())
plt.figure(figsize=(12,4))
sns.barplot(x=bank_ticker,y=returns.std(), palette='colorblind')
plt.title('2006 - 2016 Bank Stocks Returns Standard Deviation')
```

```
BAC return    0.036628
C return      0.179066
GS return     0.025358
JPM return    0.027651
MS return     0.037821
WFC return    0.031838
dtype: float64
```

[18]:
```
Text(0.5, 1.0, '2006 - 2016 Bank Stocks Returns Standard Deviation')
```



Through out the time period from 2006 to 2016, CitiGroup stock seems to be the riskiest to invest in (high standard deviation), because the stock price experienced significant volatility during the years 2008-2012.

- **Citigroup's exposure to the subprime mortgage market and significant losses during the subprime mortgage crisis** made its stock risky to invest in.

  The company held a large number of mortgage-backed securities and suffered substantial losses as the subprime mortgage crisis unfolded. This negatively impacted the company's financial position and investor confidence.

- **Concerns about Citigroup's financial stability arose during the global financial crisis**, as the company heavily relied on short-term funding and had exposure to risky assets.

  The company was heavily reliant on short-term funding and had significant exposure to risky assets. These factors contributed to doubts about the company's ability to weather the financial storm and raised concerns about its solvency.

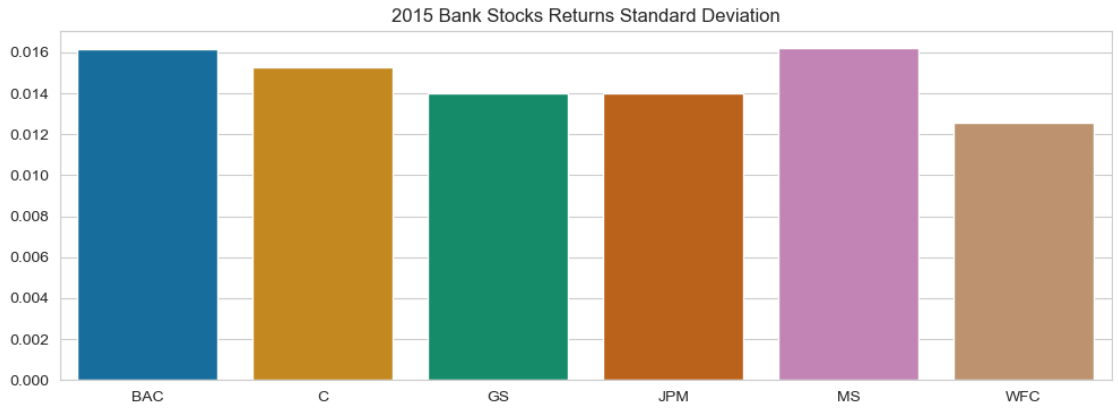
- **Regulatory scrutiny and legal challenges** related to issues such as mortgage fraud and inadequate risk management practices increased the perceived riskiness of investing in Citigroup stock.

- **The overall market volatility and uncertainty during the subprime mortgage crisis and global financial crisis** created a turbulent environment for financial institutions, including Citigroup, making investors cautious and risk-averse.

[19]:
```
print(returns.loc['2014-12-31':'2016-01-01'].std())
plt.figure(figsize=(12,4))
sns.barplot(x=bank_ticker,y=returns.loc['2014-12-31':'2016-01-01'].std(),
            palette='colorblind')
plt.title('2015 Bank Stocks Returns Standard Deviation')
```

```
BAC return    0.016152
C return      0.015282
GS return     0.014031
JPM return    0.014001
MS return     0.016219
WFC return    0.012585
dtype: float64
```

[19]: `Text(0.5, 1.0, '2015 Bank Stocks Returns Standard Deviation')`



MS and BAC stock seem to be the riskiest for the year 2015. Because In 2015, the Federal Reserve (Fed) faced a crucial decision on **whether to increase interest rates. After the 2008 financial crisis, the Fed had maintained near-zero interest rates to stimulate the economy.** However, with the US economy recovering, there was a debate

about tightening monetary policy. **Eventually, the Fed decided to raise the key interest rate, marking the first increase in almost a decade.** which affected not only to MS and BAC but also to all of other stocks.

When banks raise their lending interest rates, there are several reasons why their stocks become less attractive to investors:

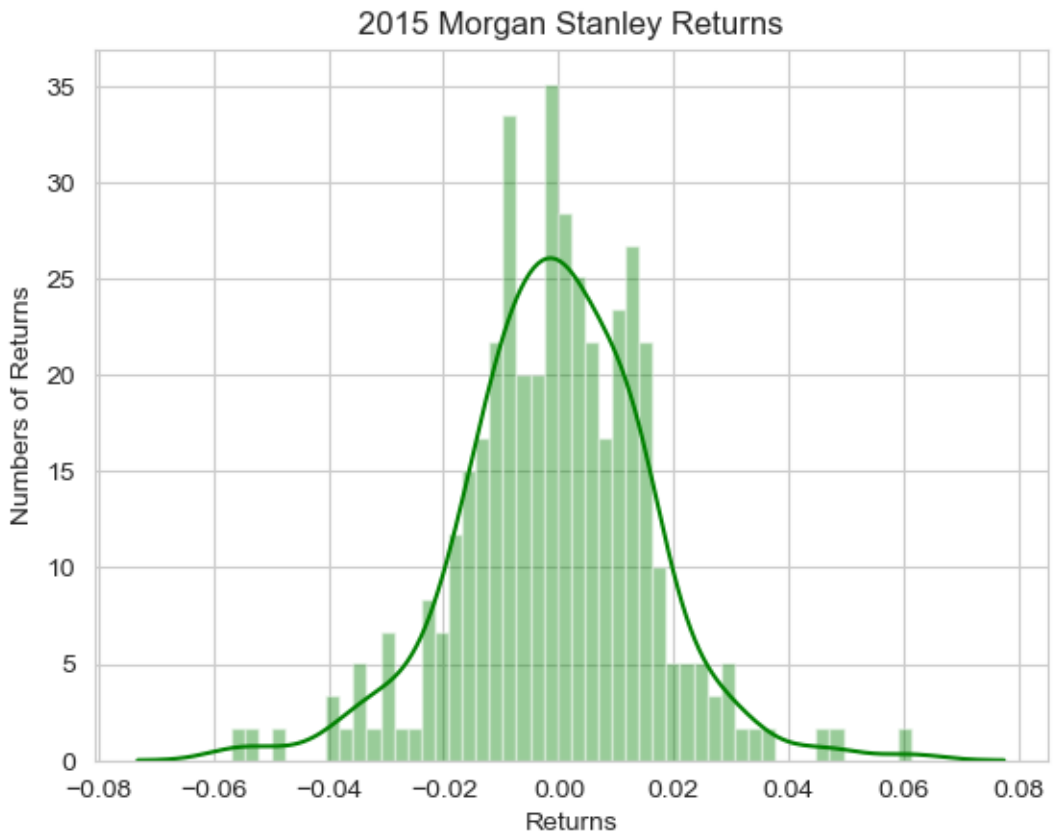
- **Increased borrowing costs**: Higher interest rates result in increased borrowing costs for banks. This can impact the banks' profitability, reduce investment returns, and diminish the potential for shareholder profits.
- **Prolonged capital recovery time**: When interest rates rise, it takes longer to recover capital from loans. Long-term loans may become less appealing, which can affect the ability to generate profits for investors.
- **Competition with alternative investments**: When interest rates increase, alternative investments such as government bonds or the stock market may become more appealing to investors. Bank stocks may be perceived as less attractive compared to other investment options, reducing their appeal to investors.

NOTE: Although MS and BAC stocks were considered the riskiest in 2015, the standard deviations of all the given stocks were not significantly different. This implies that both MS and BAC stocks did not exhibit strong volatility.

6 . Distplot of the 2015 returns for Morgan Stanley

[20]:
```
MS_return_2015 = returns[['MS return']].loc['2014-12-31':'2016-01-01']
sns.distplot(MS_return_2015, bins = 50, color ='g')
plt.title('2015 Morgan Stanley Returns')
plt.xlabel('Returns')
plt.ylabel('Numbers of Returns')
```

[20]:
```
Text(0, 0.5, 'Numbers of Returns')
```



Based on the observed this distribution plot, it is evident that MS stock remains relatively stable with minimal volatility centered around the range -0.02 to 0.02.

7. Distplot of the 2008 returns for CitiGroup

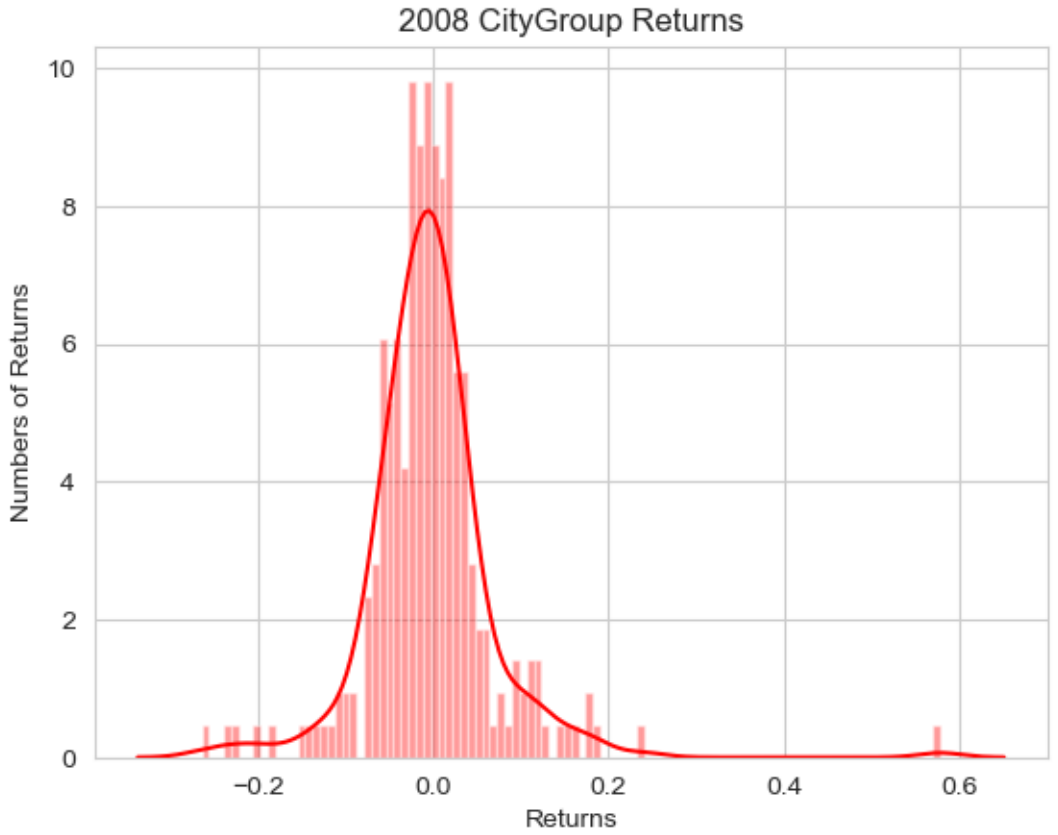
[21]:
```
print('2015 CityGroup Stock Standard Deviation: ' + str(round(returns['C␣
 ↪return'].loc['2014-12-31':'2016-01-01'].std(),4)))
print('')
print('2008 CityGroup Stock Standard Deviation: ' + str(round(returns['C␣
 ↪return'].loc['2006-01-01':'2016-01-01'].std(),4)))

C_return_2008 = returns[['C return']].loc['2007-12-31':'2009-01-01']
sns.distplot(C_return_2008, bins = 100, color ='r')
plt.title('2008 CityGroup Returns')
plt.xlabel('Returns')
plt.ylabel('Numbers of Returns')
```

```
2015 CityGroup Stock Standard Deviation: 0.0153

2008 CityGroup Stock Standard Deviation: 0.1791
```

[21]:
```
Text(0, 0.5, 'Numbers of Returns')
```



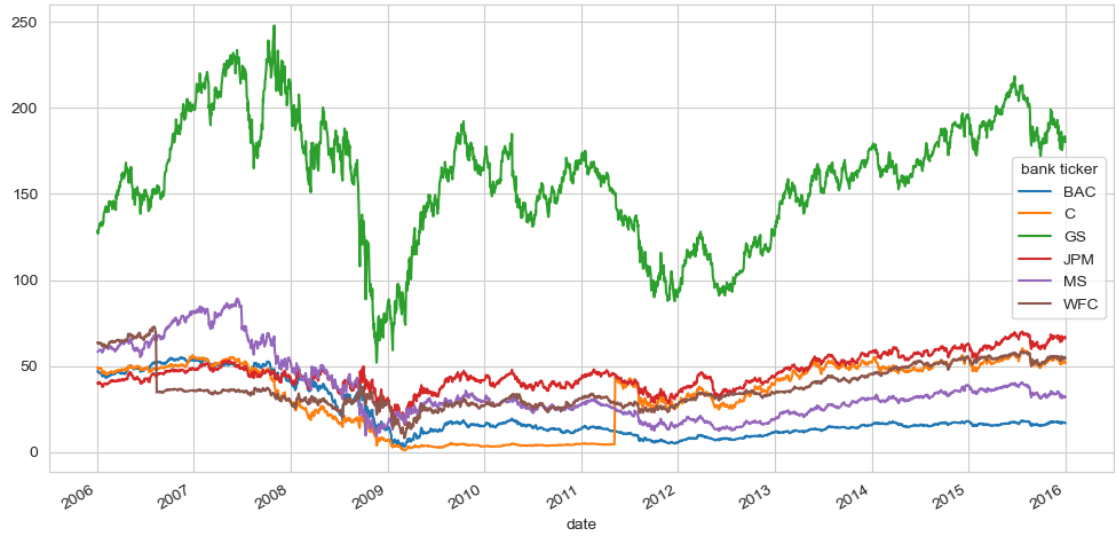
Based on the observed distribution plot, it is evident that **the standard deviation of C stock in 2008 is significantly larger and more spread out compared to 2015**. In 2008, the stock exhibits much higher volatility, with the standard deviation being **approximately 10 times larger** than in 2015.

4 Additional Information around the Financial Crisis 2008

1 . Line plot shows Close price for each bank for the entire index of time.

[22]:
```
bank_stocks.xs(key='close', level = 'stock info', axis =1).
↪plot(figsize=(12,6), kind='line')
```

[22]: <Axes: xlabel='date'>



By observing this line plot, it is evident that **GS (Goldman Sachs) stock stands out from the other stocks**. **Firstly, the close price of GS stock consistently remains higher than the other stocks** throughout the period, indicating its strong performance and market position. **Secondly, GS stock demonstrates a rapid recovery from the financial crisis in 2008**, suggesting the bank's resilience and ability to navigate challenging economic conditions.

1. Out-standing from the other stocks: Goldman Sachs (GS) is distinguished from other banks and financial institutions for the following reasons:

- **History and reputation**: GS has a long-standing history and a strong reputation in the financial industry, making it a trusted and reputable institution globally.
- **Extensive network**: GS has a wide-ranging network of relationships across industries and countries, providing access to diverse business opportunities and collaborations.
- **Talented workforce**: GS attracts highly skilled and experienced professionals in the financial sector, enabling the company to leverage their expertise in innovative and analytical problem-solving.
- **Diversified business model**: GS's business model extends beyond traditional investment banking, encompassing asset management, advisory services, core trading, and more. This diversification enhances revenue streams and mitigates risks.

2. Rapid recovery from the financial crisis in 2008: GS demonstrated a relatively swift recovery and outperformed many competitors. Key factors contributing to GS's recovery include:

- **Government bailout**: GS received a 10 billion Dollars bailout loan through the TARP program, aiding their financial position and restructuring efforts.
- **Transition to a commercial bank**: GS shifted from an investment bank model to a commercial bank, accessing interbank funding and government support programs.
- **Business focus**: GS prioritized core activities like proprietary trading, asset management, and financial advisory services, while strengthening brokerage operations and distributing valuable securities.
- **Risk management and structural adjustments**: GS improved risk management, increased equity capital levels, and reduced financial leverage to enhance stability and resilience.

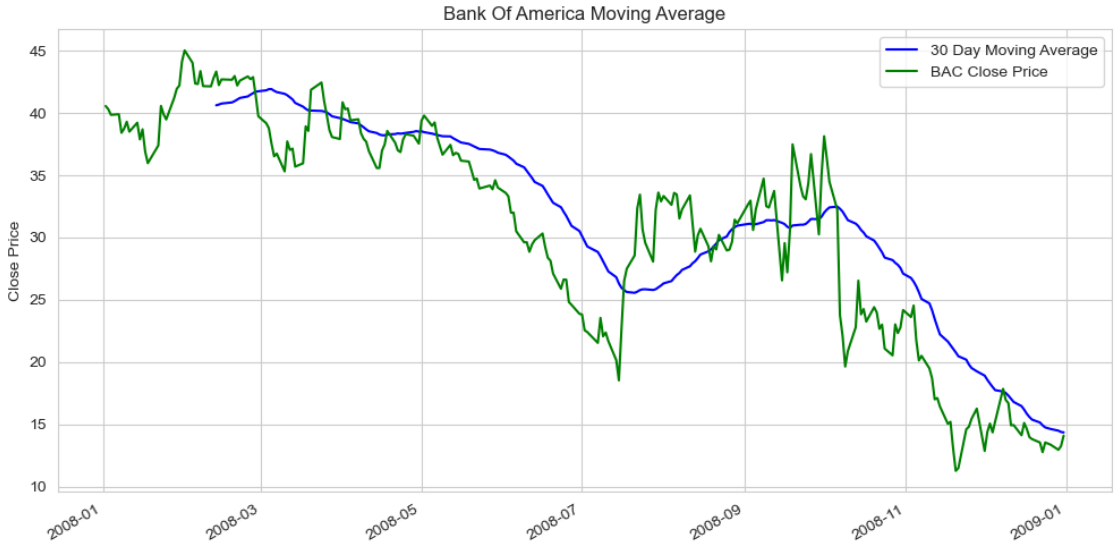
4.1 Moving Averages

Let's analyze the moving averages for these stocks in the year 2008.

The rolling 30 day average against the Close Price for Bank Of America's stock for the year 2008

[23]:
```
plt.figure(figsize=(12,6))
BAC['close'].loc['2008-01-01':'2008-12-31'].rolling(window=30).mean().
 ↪plot(color='blue', label='30 Day Moving Average')
BAC['close'].loc['2008-01-01':'2008-12-31'].plot(color='green', label='BAC␣
 ↪Close Price')
plt.ylabel('Close Price')
plt.xlabel('')
plt.title('Bank Of America Moving Average')
plt.legend()
```

[23]:
```
<matplotlib.legend.Legend at 0x241a8a31840>
```



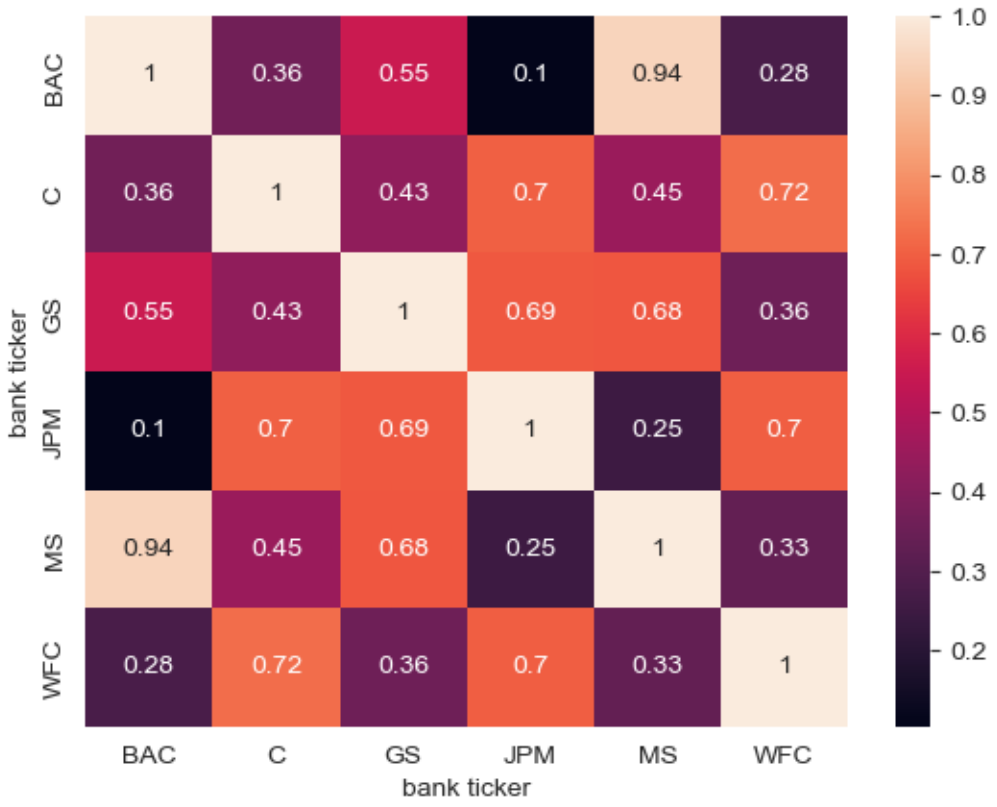
Based on the observation of this plot, it can be concluded that the rolling 30-day average line plot of the close price is **a useful tool for predicting and analyzing the close price trend. This plot presents a smooth line that is easier to interpret and provides a quick overview of the close price trend.**.

4.2 Heatmap

Create a heatmap of the correlation between the stocks Close Price.

[24]:
```
sns.heatmap(bank_stocks.xs(key='close', level='stock info', axis = 1).
↪corr(), annot=True)
```

[24]: <Axes: xlabel='bank ticker', ylabel='bank ticker'>

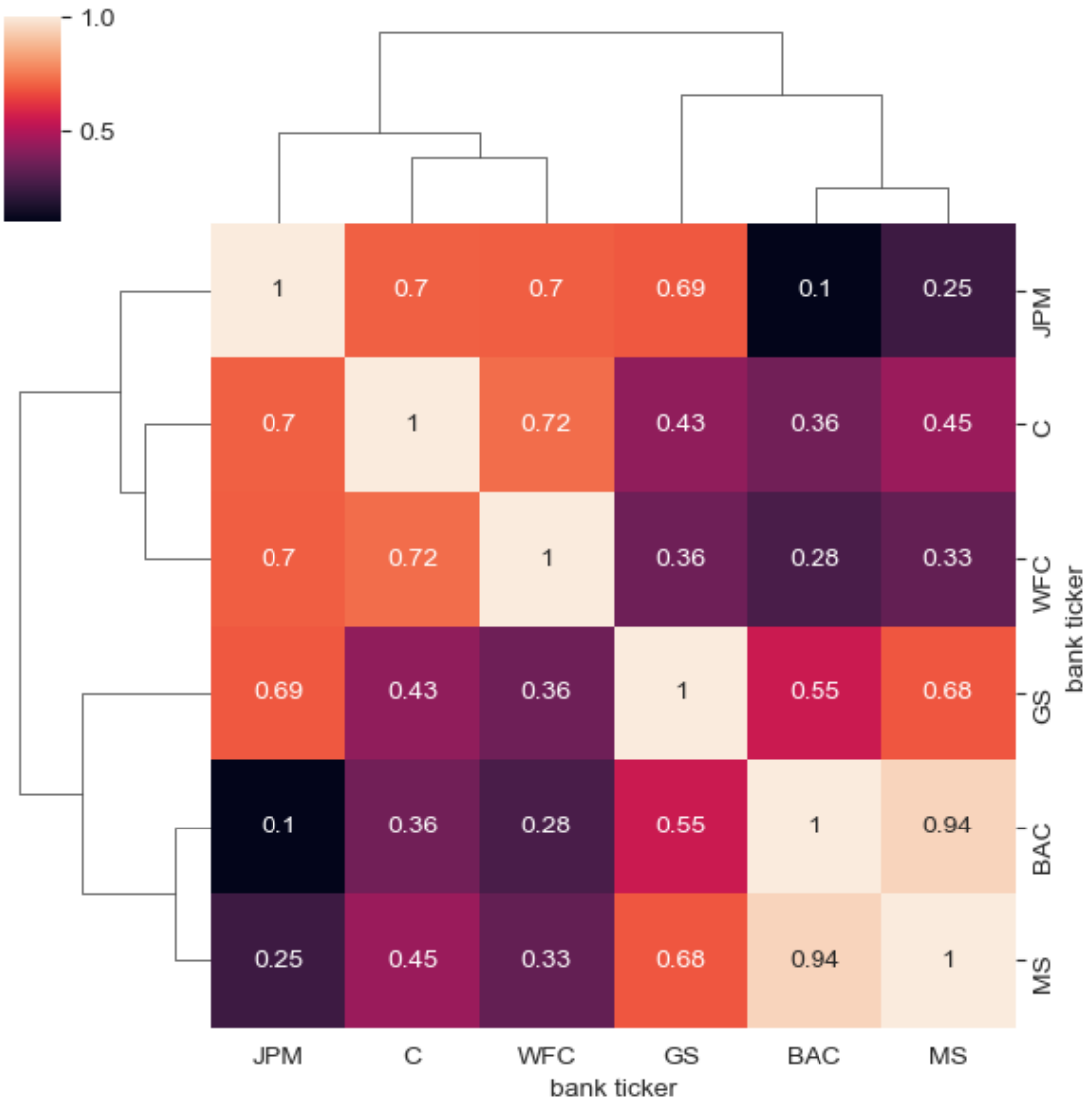


The heatmap alone may not provide a clear explanation of the correlation between bank stocks' close prices. In the real world, the close prices of two individual bank stocks may not have a direct correlation with each other but are often influenced by various factors, including the performance of other stocks and external market conditions.

Use seaborn's clustermap to cluster the correlations together:

[25]:
```
sns.clustermap(bank_stocks.xs(key='close', level='stock info', axis = 1).
↪corr(), annot=True, figsize=(6,6))
```

[25]: <seaborn.matrix.ClusterGrid at 0x241a8bce110>



By observing the clustermap chart, several conclusions can be drawn as follows:

- The close prices of C and WFC stocks show a strong correlation. These two stocks form a group that also correlates with the close price of JPM stock, forming Group 1.
- Similarly, the close prices of BAC and MS stocks exhibit a strong correlation. These two stocks form a group that also correlates with the close price of GS stock, forming Group 2.
- Group 1 and Group 2 show a correlation with each other.

5 Thank you for your reading!