Election Data - Polls and Donor - US 2012

1.) Who was being polled and what was their party affiliation?

2.) Did the poll results favor Romney or Obama?

3.) How do undecided voters effect the poll?

4.) Can we account for the undecided voters?

5.) How did voter sentiment change over time?

6.) Can we see an effect in the polls from the debates?

In [1]:
```
import pandas as pd
from pandas import Series, DataFrame
import numpy as np
```

In [2]:
```
import matplotlib.pyplot as plt
import seaborn as sns
sns.set_style('whitegrid')
%matplotlib inline
```

In [3]:
```
import requests
```

In [4]:
```
from io import StringIO
```

In [5]:
```
#grabs election poll data from url

url = 'http://elections.huffingtonpost.com/pollster/2012-general-election-romney-vs-obama.csv'

source = requests.get(url).text

poll_data = StringIO(source)
```

In [6]:
```
poll_df = pd.read_csv(poll_data)
```

In [7]:
```
poll_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 586 entries, 0 to 585
Data columns (total 17 columns):
 #   Column                  Non-Null Count  Dtype  
---  ------                  --------------  -----  
 0   Pollster                586 non-null    object 
 1   Start Date              586 non-null    object 
 2   End Date                586 non-null    object 
 3   Entry Date/Time (ET)    586 non-null    object 
 4   Number of Observations  564 non-null    float64
 5   Population              586 non-null    object 
 6   Mode                    586 non-null    object 
 7   Obama                   586 non-null    float64
 8   Romney                  586 non-null    float64
 9   Undecided               423 non-null    float64
 10  Other                   202 non-null    float64
 11  Pollster URL            586 non-null    object 
 12  Source URL              584 non-null    object 
 13  Partisan                586 non-null    object 
 14  Affiliation             586 non-null    object 
 15  Question Text           0 non-null      float64
 16  Question Iteration      586 non-null    int64  
dtypes: float64(6), int64(1), object(10)
memory usage: 78.0+ KB
```

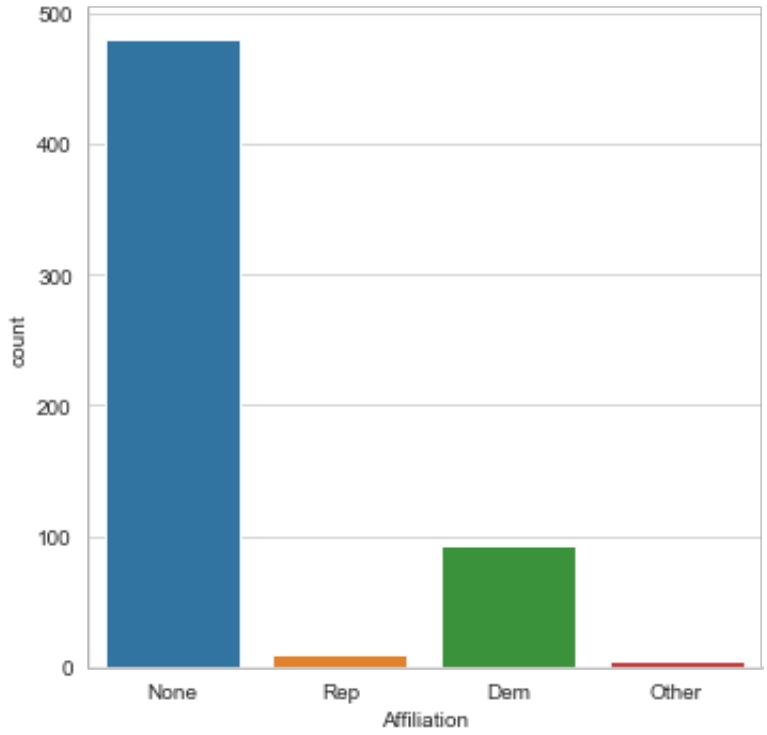
In [8]:
```
poll_df.head()
```

Out[8]:

	Pollster	Start Date	End Date	Entry Date/Time (ET)	Number of Observations	Population	Mo
0	Politico/GWU/Battleground	2012-11-04	2012-11-05	2012-11-06T08:40:26Z	1000.0	Likely Voters	Li Pho
1	YouGov/Economist	2012-11-03	2012-11-05	2012-11-26T15:31:23Z	740.0	Likely Voters	Interr
2	Gravis Marketing	2012-11-03	2012-11-05	2012-11-06T09:22:02Z	872.0	Likely Voters	Automat Pho
3	IBD/TIPP	2012-11-03	2012-11-05	2012-11-06T08:51:48Z	712.0	Likely Voters	Li Pho
4	Rasmussen	2012-11-03	2012-11-05	2012-11-06T08:47:50Z	1500.0	Likely Voters	Automat Pho

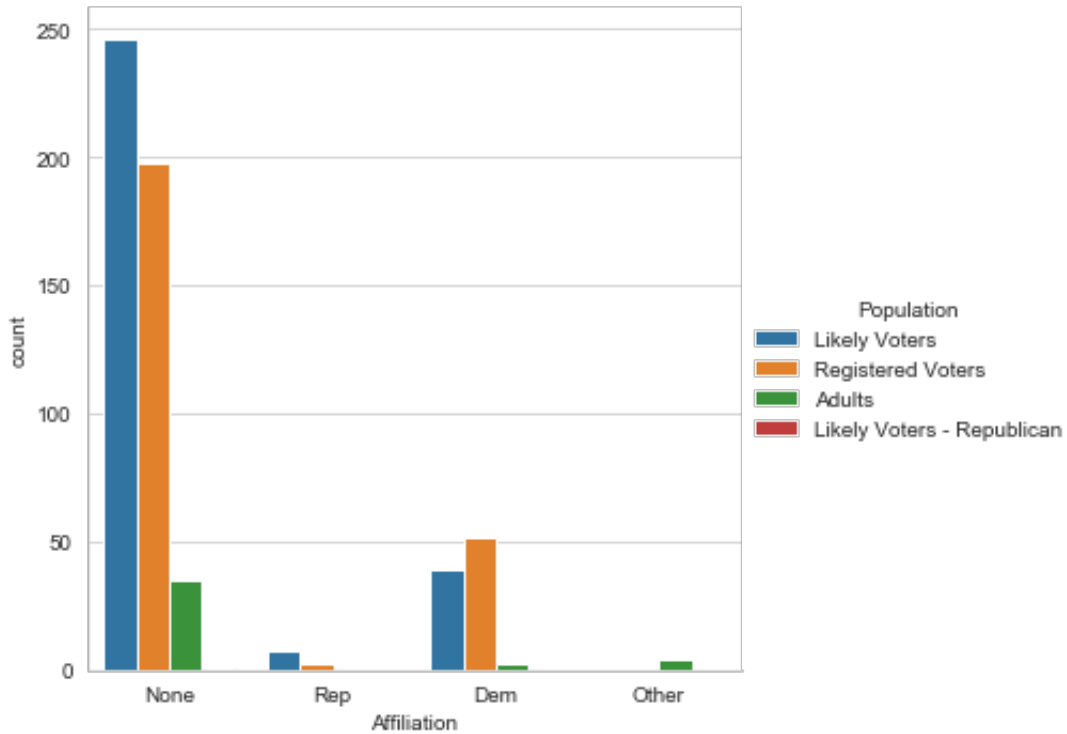
In [9]:
```
sns.catplot('Affiliation', data= poll_df, kind= 'count')
```

Out[9]: <seaborn.axisgrid.FacetGrid at 0x1a258cfb10>



In [10]:
```
sns.catplot('Affiliation', data= poll_df, hue= 'Population', kind='
count')
```

Out[10]: <seaborn.axisgrid.FacetGrid at 0x1a2607ddd0>



In [11]:
```
poll_df.head()
```

Out[11]:

	Pollster	Start Date	End Date	Entry Date/Time (ET)	Number of Observations	Population	Mo
0	Politico/GWU/Battleground	2012-11-04	2012-11-05	2012-11-06T08:40:26Z	1000.0	Likely Voters	Li Pho
1	YouGov/Economist	2012-11-03	2012-11-05	2012-11-26T15:31:23Z	740.0	Likely Voters	Interr
2	Gravis Marketing	2012-11-03	2012-11-05	2012-11-06T09:22:02Z	872.0	Likely Voters	Automat Pho
3	IBD/TIPP	2012-11-03	2012-11-05	2012-11-06T08:51:48Z	712.0	Likely Voters	Li Pho
4	Rasmussen	2012-11-03	2012-11-05	2012-11-06T08:47:50Z	1500.0	Likely Voters	Automat Pho

In [12]:
```
avg = pd.DataFrame(poll_df.mean())

avg.drop(['Number of Observations', 'Question Text', 'Question Iteration'], axis=0 , inplace=True)
```

In [13]:
```
avg.head()
```

Out[13]:

	0
Obama	46.805461
Romney	44.614334
Undecided	6.550827
Other	3.376238

In [14]:
```
std = pd.DataFrame(poll_df.std())

std.drop(['Number of Observations', 'Question Text', 'Question Iteration'], axis=0 , inplace=True)
```

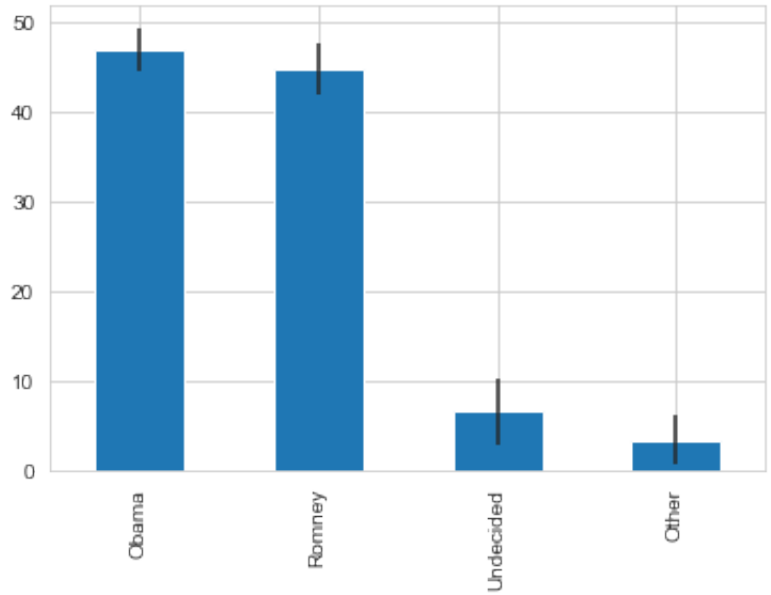
In [15]:
```
std.head()
```

Out[15]:

	0
Obama	2.422058
Romney	2.906180
Undecided	3.701754
Other	2.692726

In [16]:
```
avg.plot(yerr=std, kind='bar', legend= False,)
```

Out[16]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1a2620d850>
```



In [17]:
```
poll_avg = pd.concat([avg, std], axis=1)
```

In [18]:
```
poll_avg.columns = ['Average', 'STD']
```

In [19]:
```
poll_avg
```

Out[19]:

	Average	STD
Obama	46.805461	2.422058
Romney	44.614334	2.906180
Undecided	6.550827	3.701754
Other	3.376238	2.692726

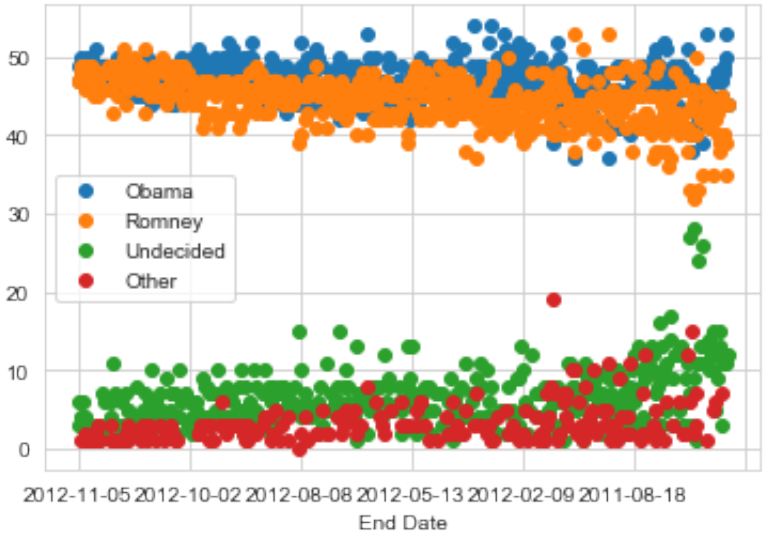
In [20]:
```
poll_df.head()
```

Out[20]:

	Pollster	Start Date	End Date	Entry Date/Time (ET)	Number of Observations	Population	Mo
0	Politico/GWU/Battleground	2012-11-04	2012-11-05	2012-11-06T08:40:26Z	1000.0	Likely Voters	Li Pho
1	YouGov/Economist	2012-11-03	2012-11-05	2012-11-26T15:31:23Z	740.0	Likely Voters	Interr
2	Gravis Marketing	2012-11-03	2012-11-05	2012-11-06T09:22:02Z	872.0	Likely Voters	Automat Pho
3	IBD/TIPP	2012-11-03	2012-11-05	2012-11-06T08:51:48Z	712.0	Likely Voters	Li Pho
4	Rasmussen	2012-11-03	2012-11-05	2012-11-06T08:47:50Z	1500.0	Likely Voters	Automat Pho

In [22]:
```
poll_df.plot(x='End Date', y=['Obama', 'Romney', 'Undecided', 'Other'], linestyle='', marker = 'o')
```

Out[22]: `<matplotlib.axes._subplots.AxesSubplot at 0x1a26428910>`



In [23]:
```
from datetime import datetime
```

In [24]:
```
poll_df['Difference'] = (poll_df.Obama - poll_df.Romney)/100
```

In [25]:
```
poll_df.head()
```

Out[25]:

	Pollster	Start Date	End Date	Entry Date/Time (ET)	Number of Observations	Population	Mo
0	Politico/GWU/Battleground	2012-11-04	2012-11-05	2012-11-06T08:40:26Z	1000.0	Likely Voters	Li Pho
1	YouGov/Economist	2012-11-03	2012-11-05	2012-11-26T15:31:23Z	740.0	Likely Voters	Interr
2	Gravis Marketing	2012-11-03	2012-11-05	2012-11-06T09:22:02Z	872.0	Likely Voters	Automat Pho
3	IBD/TIPP	2012-11-03	2012-11-05	2012-11-06T08:51:48Z	712.0	Likely Voters	Li Pho
4	Rasmussen	2012-11-03	2012-11-05	2012-11-06T08:47:50Z	1500.0	Likely Voters	Automat Pho

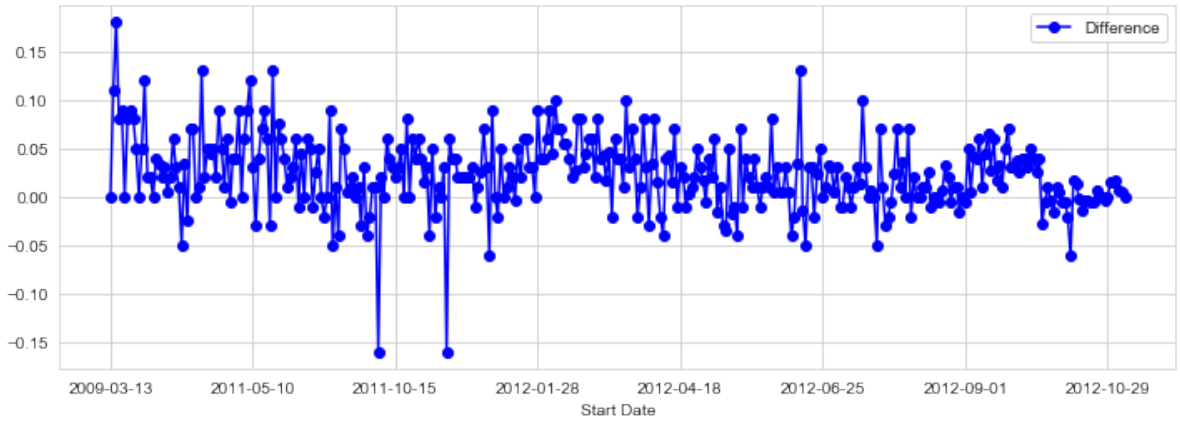
In [35]:
```
poll_df = poll_df.groupby(['Start Date'], as_index = False).mean()

poll_df.drop(['Other'], axis=1 , inplace=True)

poll_df.head()
```

Out[35]:

	Start Date	Obama	Romney	Undecided	Difference
0	2009-03-13	44.0	44.0	12.0	0.00
1	2009-04-17	50.0	39.0	11.0	0.11
2	2009-05-14	53.0	35.0	12.0	0.18
3	2009-06-12	48.0	40.0	12.0	0.08
4	2009-07-15	49.0	40.0	11.0	0.09

In [36]:
```
poll_df.plot('Start Date', 'Difference', figsize=(12,4), marker='o'
, linestyle='-', color='blue' )
```

Out[36]: <matplotlib.axes._subplots.AxesSubplot at 0x1a26994f90>

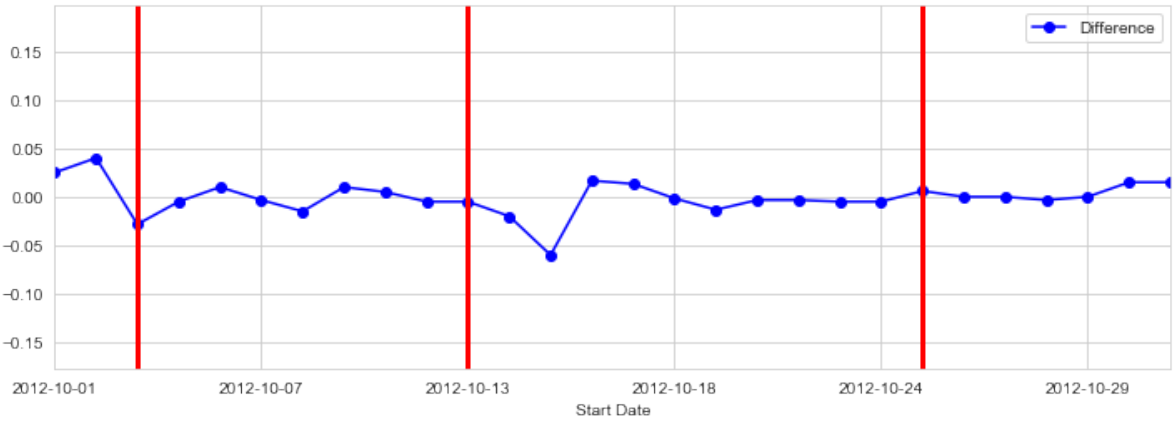


In [41]:
```
row_in = 0
xlimit = []

for date in poll_df['Start Date']:
    if date[0:7] == '2012-10':
        xlimit.append(row_in)
        row_in += 1
    else:
        row_in += 1

print (min(xlimit))
print (max(xlimit))
```

```
325
352
```

In [48]:
```
poll_df.plot('Start Date', 'Difference', figsize=(12,4), marker='o'
, linestyle='-', color='blue', xlim=(325,352) )

#Oct 03
plt.axvline(x=325+2, linewidth = 3, color='red')
#Oct 11
plt.axvline(x=325+10, linewidth = 3, color='red')
#Oct 22
plt.axvline(x=325+21, linewidth = 3, color='red')
```

Out[48]:
```
<matplotlib.lines.Line2D at 0x1a26eac790>
```



In [49]:
```
donor_df = pd.read_csv('Election_Donor_Data.csv')
```

```
/Users/Martin_Hopkins/opt/anaconda3/lib/python3.7/site-packages/IP
ython/core/interactiveshell.py:3063: DtypeWarning: Columns (6) hav
e mixed types.Specify dtype option on import or set low_memory=Fal
se.
  interactivity=interactivity, compiler=compiler, result=result)
```

In [50]:
```
donor_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1001731 entries, 0 to 1001730
Data columns (total 16 columns):
 #   Column             Non-Null Count    Dtype  
---  ------             --------------    -----  
 0   cmte_id            1001731 non-null  object 
 1   cand_id            1001731 non-null  object 
 2   cand_nm            1001731 non-null  object 
 3   contbr_nm          1001731 non-null  object 
 4   contbr_city        1001712 non-null  object 
 5   contbr_st          1001727 non-null  object 
 6   contbr_zip         1001620 non-null  object 
 7   contbr_employer    988002 non-null   object 
 8   contbr_occupation  993301 non-null   object 
 9   contb_receipt_amt  1001731 non-null  float64
 10  contb_receipt_dt   1001731 non-null  object 
 11  receipt_desc       14166 non-null    object 
 12  memo_cd            92482 non-null    object 
 13  memo_text          97770 non-null    object 
 14  form_tp            1001731 non-null  object 
 15  file_num           1001731 non-null  int64  
dtypes: float64(1), int64(1), object(14)
memory usage: 122.3+ MB
```

In [51]:
```
donor_df.head()
```

Out[51]:

	cmte_id	cand_id	cand_nm	contbr_nm	contbr_city	contbr_st	contbr_zip
0	C00410118	P20002978	Bachmann, Michelle	HARVEY, WILLIAM	MOBILE	AL	3.6601e+08
1	C00410118	P20002978	Bachmann, Michelle	HARVEY, WILLIAM	MOBILE	AL	3.6601e+08
2	C00410118	P20002978	Bachmann, Michelle	SMITH, LANIER	LANETT	AL	3.68633e+08
3	C00410118	P20002978	Bachmann, Michelle	BLEVINS, DARONDA	PIGGOTT	AR	7.24548e+08
4	C00410118	P20002978	Bachmann, Michelle	WARDENBURG, HAROLD	HOT SPRINGS NATION	AR	7.19016e+08

In [53]:
```
donor_df['contb_receipt_amt'].value_counts()
```

Out[53]:
```
100.00    178188
50.00     137584
25.00     110345
250.00     91182
500.00     57984
           ...
97.15          1
122.32         1
188.65         1
122.40         1
132.12         1
Name: contb_receipt_amt, Length: 8079, dtype: int64
```

In [56]:
```
don_mean = donor_df['contb_receipt_amt'].mean()

don_std = donor_df['contb_receipt_amt'].std()
```

In [59]:
```
print (f'The average donation was {round(don_mean, 2)} with a std {
round(don_std,2)}')
```

```
The average donation was 298.24 with a std 3749.67
```

In [108]:
```
top_donor = donor_df['contb_receipt_amt'].copy()

top_donor.sort_values(ascending=False, inplace=True)

top_donor
```

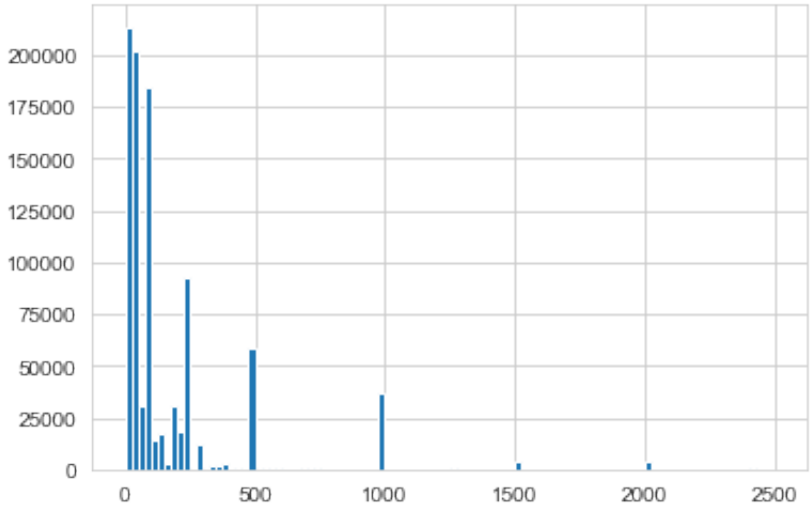
Out[108]:
```
325136    2014490.51
326651    1944042.43
344539    1679114.65
344419    1511192.17
319478     526246.17
              ...
250737      -5455.00
398429      -5500.00
101356      -7500.00
226986     -25800.00
114604     -30800.00
Name: contb_receipt_amt, Length: 1001731, dtype: float64
```

In [109]:
```
top_donor = top_donor[top_donor > 0]

top_donor.sort_values(ascending=False, inplace=True)

top_donor.value_counts().head(10)
```

Out[109]:
```
100.0     178188
50.0      137584
25.0      110345
250.0      91182
500.0      57984
2500.0     49005
35.0       37237
1000.0     36494
10.0       33986
200.0      27813
Name: contb_receipt_amt, dtype: int64
```

In [110]:
```
com_don = top_donor[top_donor < 2500]

com_don.hist(bins=100)
```

Out[110]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1a287d3490>
```



In [111]:
```
candidates = donor_df.cand_nm.unique()

candidates
```

Out[111]:
```
array(['Bachmann, Michelle', 'Romney, Mitt', 'Obama, Barack',
       "Roemer, Charles E. 'Buddy' III", 'Pawlenty, Timothy',
       'Johnson, Gary Earl', 'Paul, Ron', 'Santorum, Rick',
       'Cain, Herman', 'Gingrich, Newt', 'McCotter, Thaddeus G',
       'Huntsman, Jon', 'Perry, Rick'], dtype=object)
```

In [112]:
```
party_map = {'Bachmann, Michelle': 'Republican',
             'Cain, Herman': 'Republican',
             'Gingrich, Newt': 'Republican',
             'Huntsman, Jon': 'Republican',
             'Johnson, Gary Earl': 'Republican',
             'McCotter, Thaddeus G': 'Republican',
             'Obama, Barack': 'Democrat',
             'Paul, Ron': 'Republican',
             'Pawlenty, Timothy': 'Republican',
             'Perry, Rick': 'Republican',
             "Roemer, Charles E. 'Buddy' III": 'Republican',
             'Romney, Mitt': 'Republican',
             'Santorum, Rick': 'Republican'}

donor_df['Party'] = donor_df.cand_nm.map(party_map)
```

In [113]:
```
donor_df = donor_df[donor_df.contb_receipt_amt > 0]
```

In [114]:
```
donor_df
```

Out[114]:

	cmte_id	cand_id	cand_nm	contbr_nm	contbr_city	contbr_st	cor
0	C00410118	P20002978	Bachmann, Michelle	HARVEY, WILLIAM	MOBILE	AL	3.66
1	C00410118	P20002978	Bachmann, Michelle	HARVEY, WILLIAM	MOBILE	AL	3.66
2	C00410118	P20002978	Bachmann, Michelle	SMITH, LANIER	LANETT	AL	3.686
3	C00410118	P20002978	Bachmann, Michelle	BLEVINS, DARONDA	PIGGOTT	AR	7.245
4	C00410118	P20002978	Bachmann, Michelle	WARDENBURG, HAROLD	HOT SPRINGS NATION	AR	7.190
...	...	...	...	...	...	...	
1001726	C00500587	P20003281	Perry, Rick	GORMAN, CHRIS D. MR.	INFO REQUESTED	XX	
1001727	C00500587	P20003281	Perry, Rick	DUFFY, DAVID A. MR.	INFO REQUESTED	XX	
1001728	C00500587	P20003281	Perry, Rick	GRANE, BRYAN F. MR.	INFO REQUESTED	XX	
1001729	C00500587	P20003281	Perry, Rick	TOLBERT, DARYL MR.	INFO REQUESTED	XX	
1001730	C00500587	P20003281	Perry, Rick	ANDERSON, MARILEE MRS.	INFO REQUESTED	XX	

991475 rows × 17 columns

In [115]:
```
donor_df.groupby('cand_nm')['contb_receipt_amt'].count()
```

Out[115]:
```
cand_nm
Bachmann, Michelle                13082
Cain, Herman                      20052
Gingrich, Newt                    46883
Huntsman, Jon                      4066
Johnson, Gary Earl                 1234
McCotter, Thaddeus G                 73
Obama, Barack                     589127
Paul, Ron                         143161
Pawlenty, Timothy                  3844
Perry, Rick                       12709
Roemer, Charles E. 'Buddy' III     5844
Romney, Mitt                      105155
Santorum, Rick                     46245
Name: contb_receipt_amt, dtype: int64
```

In [116]:
```
donor_df.groupby('cand_nm')['contb_receipt_amt'].sum()
```

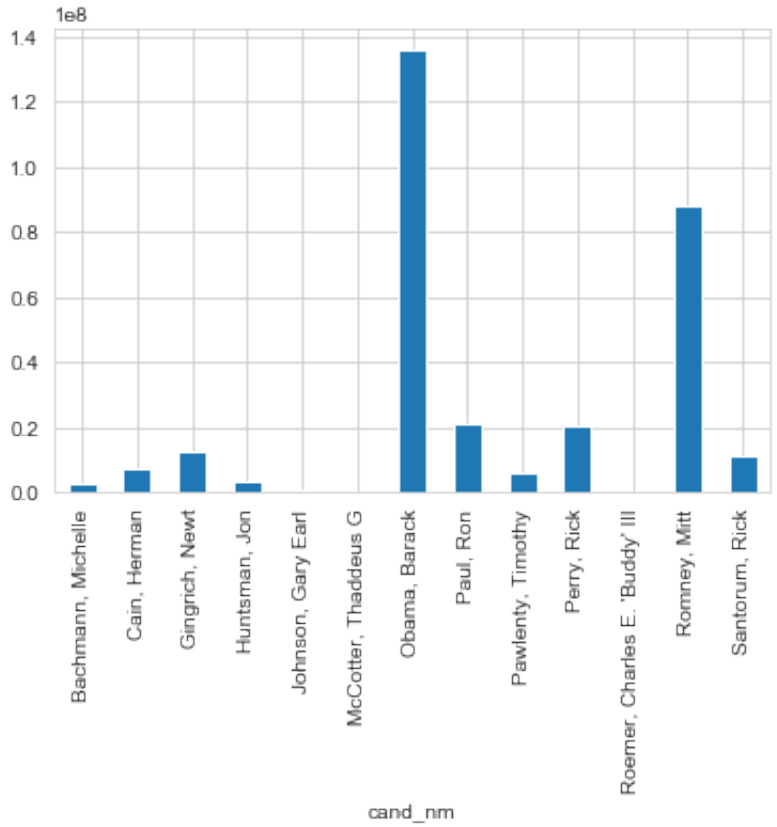
Out[116]:
```
cand_nm
Bachmann, Michelle                2.711439e+06
Cain, Herman                      7.101082e+06
Gingrich, Newt                    1.283277e+07
Huntsman, Jon                     3.330373e+06
Johnson, Gary Earl                5.669616e+05
McCotter, Thaddeus G              3.903000e+04
Obama, Barack                     1.358774e+08
Paul, Ron                         2.100962e+07
Pawlenty, Timothy                 6.004819e+06
Perry, Rick                       2.030575e+07
Roemer, Charles E. 'Buddy' III    3.730099e+05
Romney, Mitt                      8.833591e+07
Santorum, Rick                    1.104316e+07
Name: contb_receipt_amt, dtype: float64
```

In [125]:
```
cand_amount = donor_df.groupby('cand_nm')['contb_receipt_amt'].sum(
)

i =0

for don in cand_amount:
    print(f'The candidtate {cand_amount.index[i]} raised {round(don
,0)} dollars \n')
    i += 1
```

```
The candidtate Bachmann, Michelle raised 2711439.0 dollars

The candidtate Cain, Herman raised 7101082.0 dollars

The candidtate Gingrich, Newt raised 12832770.0 dollars

The candidtate Huntsman, Jon raised 3330373.0 dollars

The candidtate Johnson, Gary Earl raised 566962.0 dollars

The candidtate McCotter, Thaddeus G raised 39030.0 dollars

The candidtate Obama, Barack raised 135877427.0 dollars

The candidtate Paul, Ron raised 21009620.0 dollars

The candidtate Pawlenty, Timothy raised 6004819.0 dollars

The candidtate Perry, Rick raised 20305754.0 dollars

The candidtate Roemer, Charles E. 'Buddy' III raised 373010.0 doll
ars

The candidtate Romney, Mitt raised 88335908.0 dollars

The candidtate Santorum, Rick raised 11043159.0 dollars
```

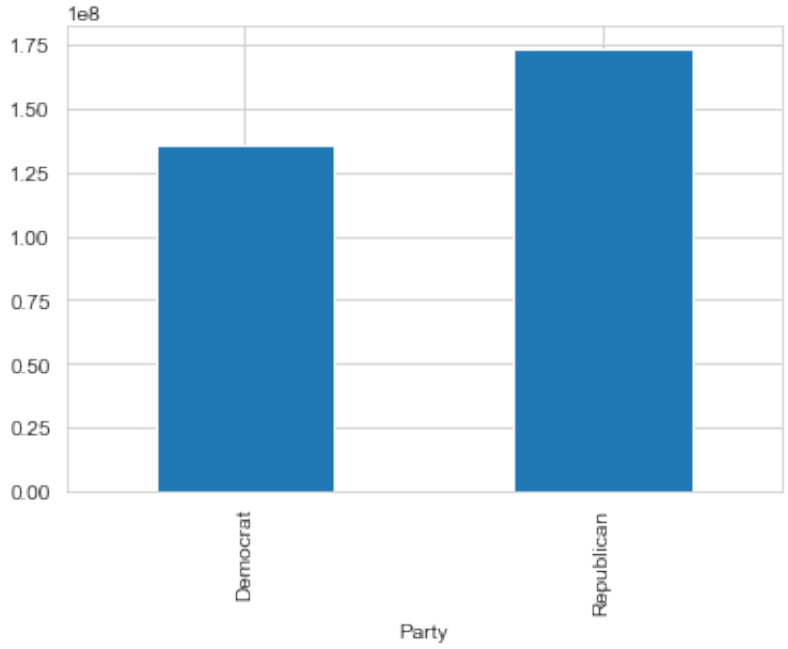
In [129]:
```
cand_amount.plot(kind='bar')
```

Out[129]: <matplotlib.axes._subplots.AxesSubplot at 0x1a26d76390>



In [130]:
```
donor_df.groupby('Party')['contb_receipt_amt'].sum().plot(kind='bar
')
```

Out[130]: <matplotlib.axes._subplots.AxesSubplot at 0x1a26c651d0>



In [132]:
```
occupation_df = donor_df.pivot_table('contb_receipt_amt', index = '
contbr_occupation', columns='Party', aggfunc ='sum')

occupation_df.head()
```

Out[132]:

Party	Democrat	Republican
contbr_occupation		
MIXED-MEDIA ARTIST / STORYTELLER	100.0	NaN
AREA VICE PRESIDENT	250.0	NaN
RESEARCH ASSOCIATE	100.0	NaN
TEACHER	500.0	NaN
THERAPIST	3900.0	NaN

In [134]:
```
occupation_df.shape
```

Out[134]:
```
(45067, 2)
```

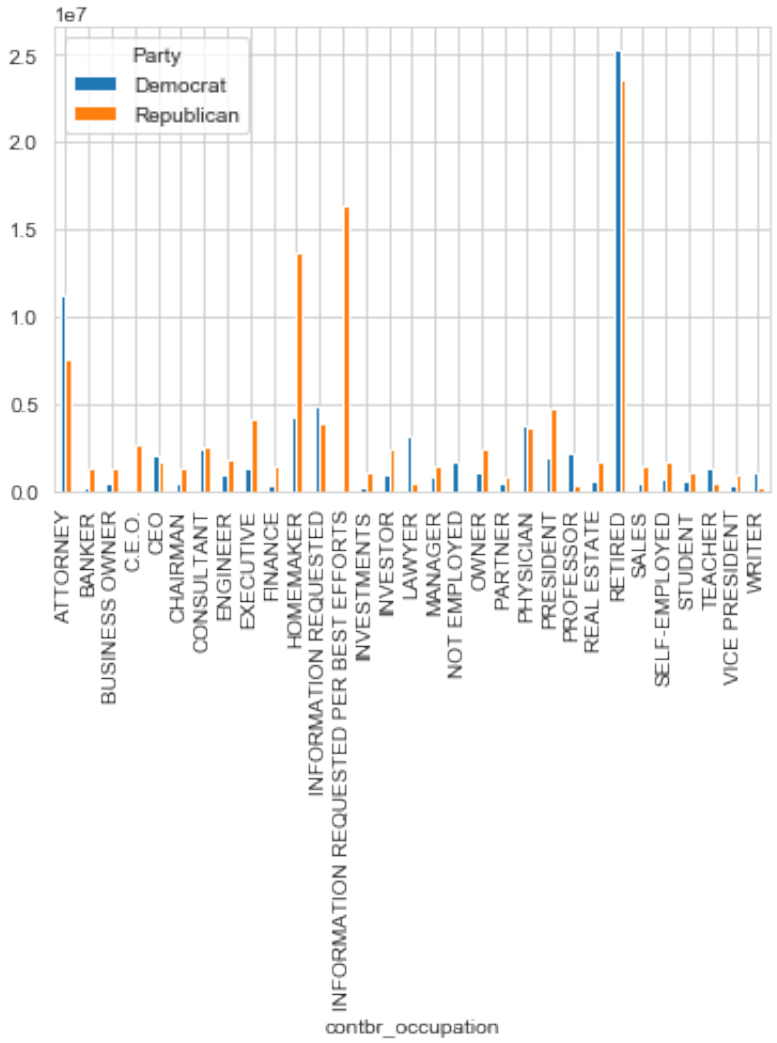
In [135]:
```
occupation_df = occupation_df[occupation_df.sum(1) > 1000000]
```

In [136]:
```
occupation_df.shape
```

Out[136]:
```
(31, 2)
```

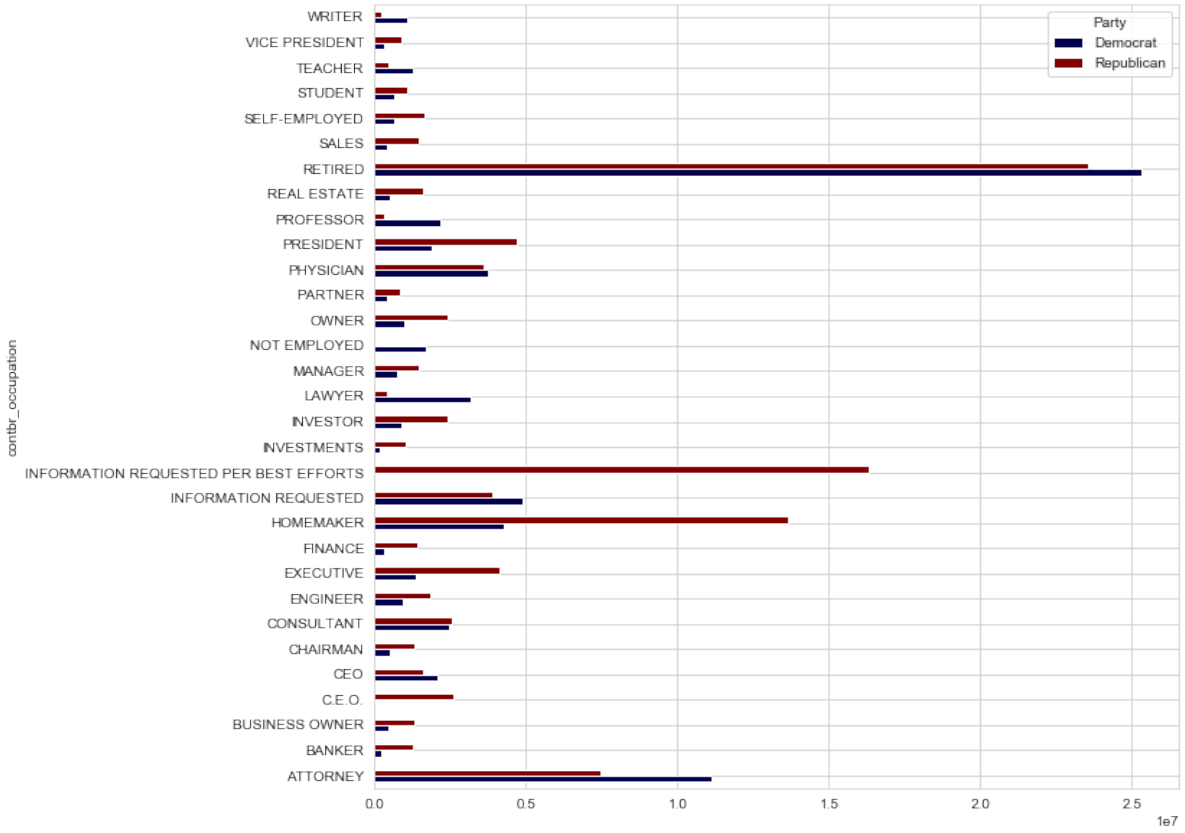
In [139]:
```
occupation_df.plot(kind='bar')
```

Out[139]: <matplotlib.axes._subplots.AxesSubplot at 0x1a2547af90>



In [150]:
```
occupation_df.plot(kind='barh', figsize=(10,10), cmap='seismic')
```

Out[150]: <matplotlib.axes._subplots.AxesSubplot at 0x1a2c712790>

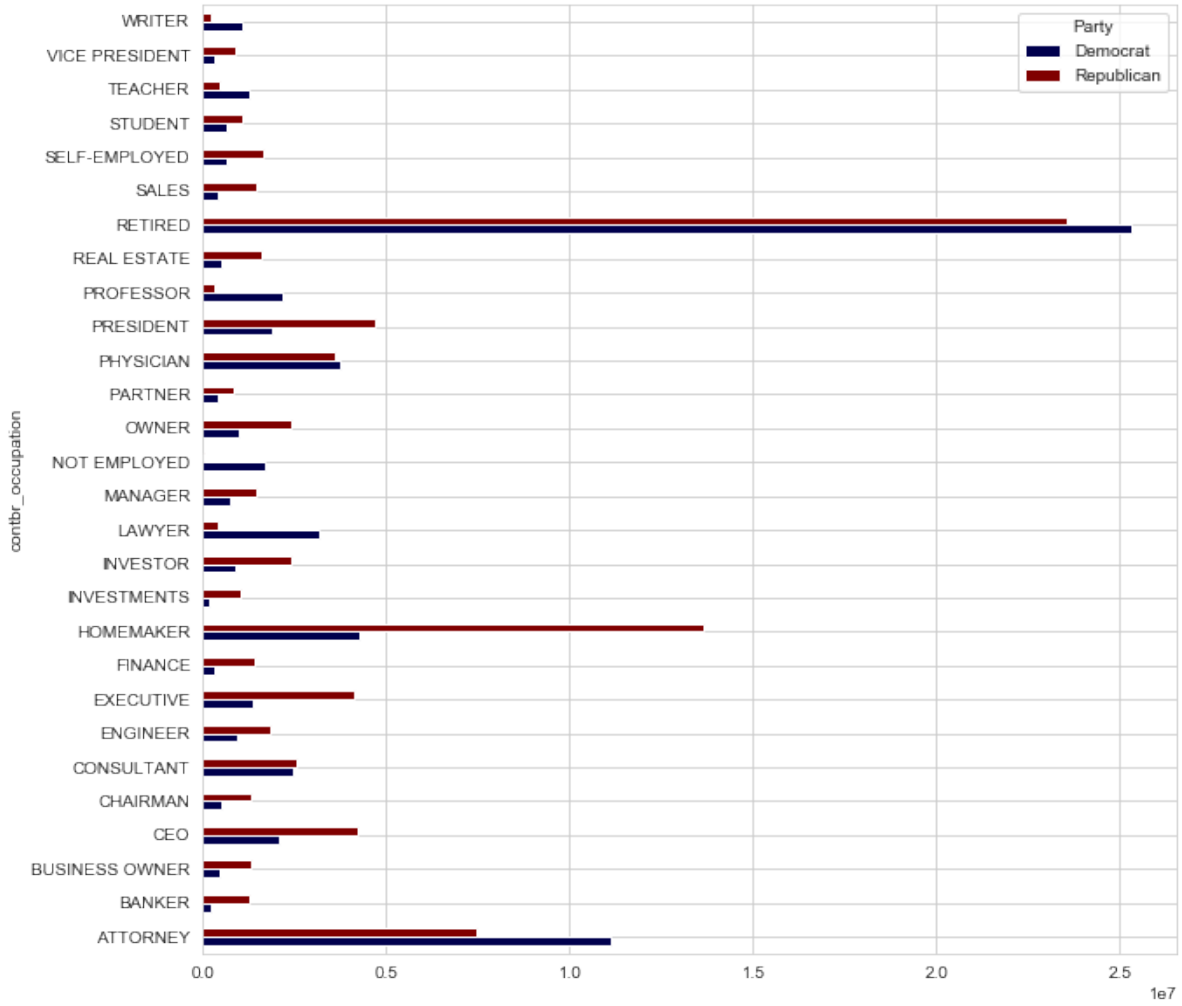


In [151]:
```
occupation_df.drop(['INFORMATION REQUESTED PER BEST EFFORTS', 'INFO
RMATION REQUESTED'], axis=0, inplace=True)
```

In [152]:
```
occupation_df.loc['CEO'] = occupation_df.loc['CEO'] + occupation_df
.loc['C.E.O.']

occupation_df.drop('C.E.O.', inplace = True)
```

In [153]:
```
occupation_df.plot(kind='barh', figsize=(10,10), cmap='seismic')
```

Out[153]: <matplotlib.axes._subplots.AxesSubplot at 0x1a2c712b90>



In [ ]:
```

```