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]:
```
from __future__ import division
```

In [4]:
```
import requests
```

In [5]:
```
from StringIO import StringIO
```

In [6]:
```
url = 'http://elections.huffingtonpost.com/pollster/2016-general-electio
n-trump-vs-clinton.csv'

source = requests.get(url).text

poll_data = StringIO(source)
```

In [7]:
```
poll_df = pd.read_csv(poll_data)
```

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1377 entries, 0 to 1376
Data columns (total 17 columns):
Pollster                  1377 non-null object
Start Date                1377 non-null object
End Date                  1377 non-null object
Entry Date/Time (ET)      1377 non-null object
Number of Observations    906 non-null float64
Population                1377 non-null object
Mode                      1377 non-null object
Trump                     1377 non-null float64
Clinton                   1377 non-null float64
Other                     947 non-null float64
Undecided                 1322 non-null float64
Pollster URL              1377 non-null object
Source URL                1377 non-null object
Partisan                  1377 non-null object
Affiliation               1377 non-null object
Question Text             602 non-null object
Question Iteration        1377 non-null int64
dtypes: float64(5), int64(1), object(11)
memory usage: 183.0+ KB
```

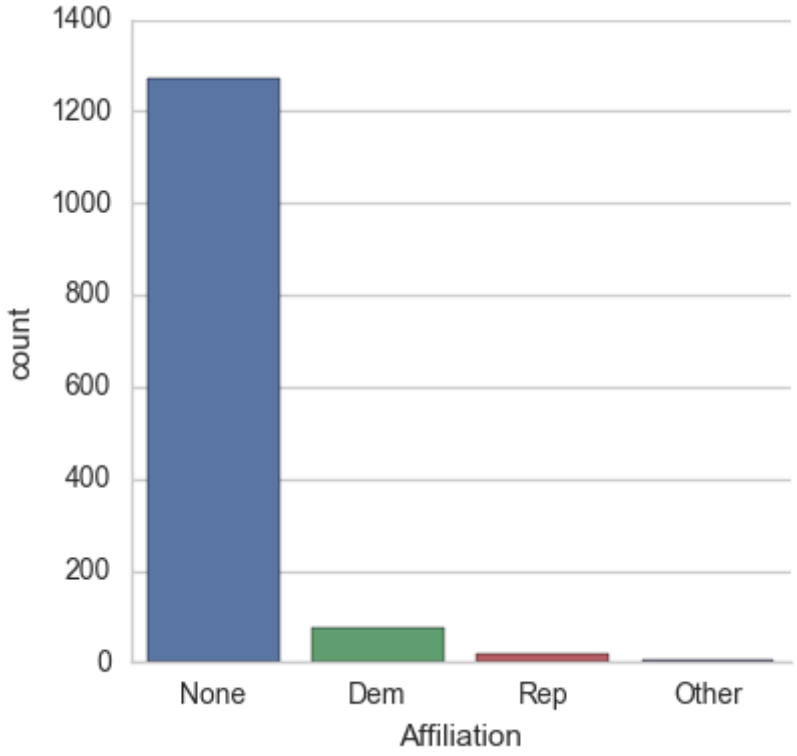
In [9]:
```
poll_df.head()
```

Out[9]:

| | Pollster | Date Start Date | Date End Date | Date/Time (ET) Entry Date/Time (ET) | Number of Observations Observations | Population | Mode | Trump |
|---|---|---|---|---|---|---|---|---|
| **0** | Insights West | 2016-10-10 | 2016-10-11 | 2016-10-12T14:34:44Z | 953.0 | Likely Voters | Internet | 37.0 |
| **1** | Insights West | 2016-10-10 | 2016-10-11 | 2016-10-12T14:34:44Z | NaN | Likely Voters - Democrat | Internet | 5.0 |
| **2** | Insights West | 2016-10-10 | 2016-10-11 | 2016-10-12T14:34:44Z | NaN | Likely Voters - Republican | Internet | 74.0 |
| **3** | Insights West | 2016-10-10 | 2016-10-11 | 2016-10-12T14:34:44Z | NaN | Likely Voters - independent | Internet | 31.0 |
| **4** | Rasmussen | 2016-10-07 | 2016-10-11 | 2016-10-12T12:39:57Z | 1500.0 | Likely Voters | IVR/Online | 39.0 |

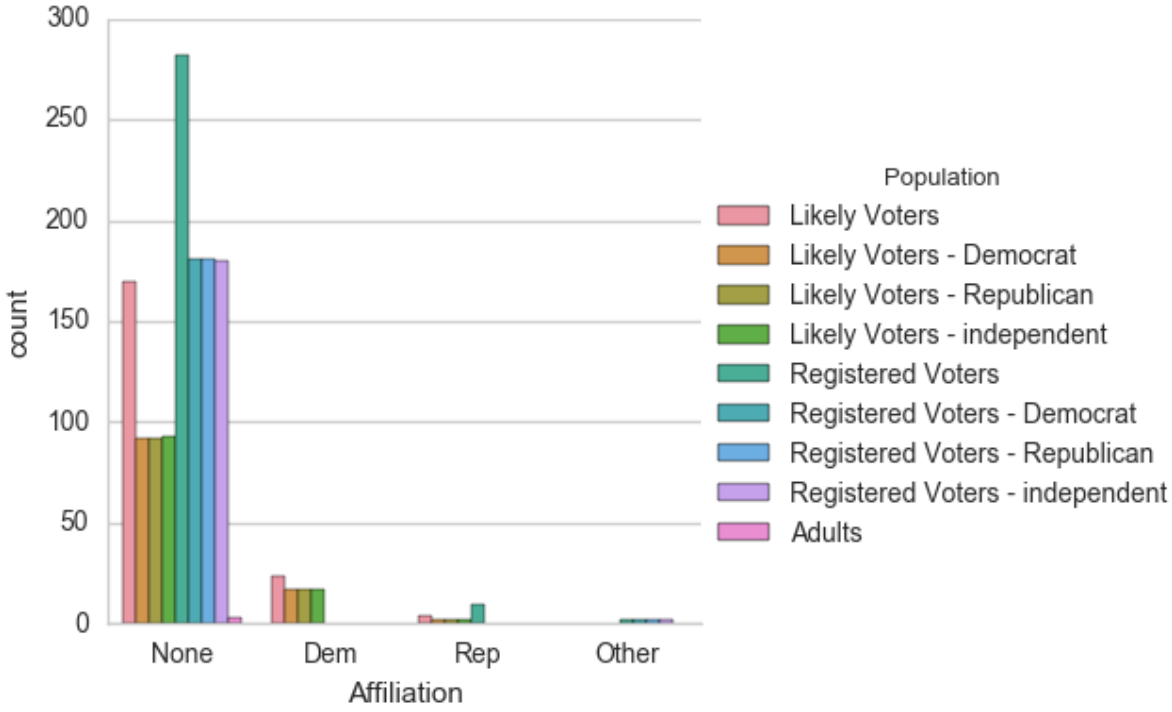
In [10]:
```
sns.factorplot('Affiliation',data=poll_df, kind = 'count')
```

Out[10]: <seaborn.axisgrid.FacetGrid at 0x1148ffe10>




In [11]:
```
sns.factorplot('Affiliation', data = poll_df, hue = 'Population', kind =
 'count')
```

Out[11]: <seaborn.axisgrid.FacetGrid at 0x116ccc2d0>



In [12]:
```
poll_df.head()
```

Out[12]:

| | Pollster | Start Date | End Date | Date/Time (ET) | Number of Observations | Population | Mode | Trump |
|---|---|---|---|---|---|---|---|---|
| | **Pollster** | **Start Date** | **End Date** | **Date/Time (ET)** | **Number of Observations** | **Population** | **Mode** | **Trump** |
| **0** | Insights West | 2016-10-10 | 2016-10-11 | 2016-10-12T14:34:44Z | 953.0 | Likely Voters | Internet | 37.0 |
| **1** | Insights West | 2016-10-10 | 2016-10-11 | 2016-10-12T14:34:44Z | NaN | Likely Voters - Democrat | Internet | 5.0 |
| **2** | Insights West | 2016-10-10 | 2016-10-11 | 2016-10-12T14:34:44Z | NaN | Likely Voters - Republican | Internet | 74.0 |
| **3** | Insights West | 2016-10-10 | 2016-10-11 | 2016-10-12T14:34:44Z | NaN | Likely Voters - independent | Internet | 31.0 |
| **4** | Rasmussen | 2016-10-07 | 2016-10-11 | 2016-10-12T12:39:57Z | 1500.0 | Likely Voters | IVR/Online | 39.0 |

In [13]:
```
avg = pd.DataFrame(poll_df.mean())

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

In [14]:
```
avg
```

| | 0 |
|---|---|
| **avg** | **0** |

Out[14]:

| | 0 |
|---|---|
| **Trump** | 40.727669 |
| **Clinton** | 42.792302 |
| **Other** | 5.921859 |
| **Undecided** | 9.660363 |

In [15]:
```
std = pd.DataFrame (poll_df.std())
std.drop('Number of Observations', axis = 0, inplace = True)
std.drop('Question Iteration', axis = 0, inplace = True)
```

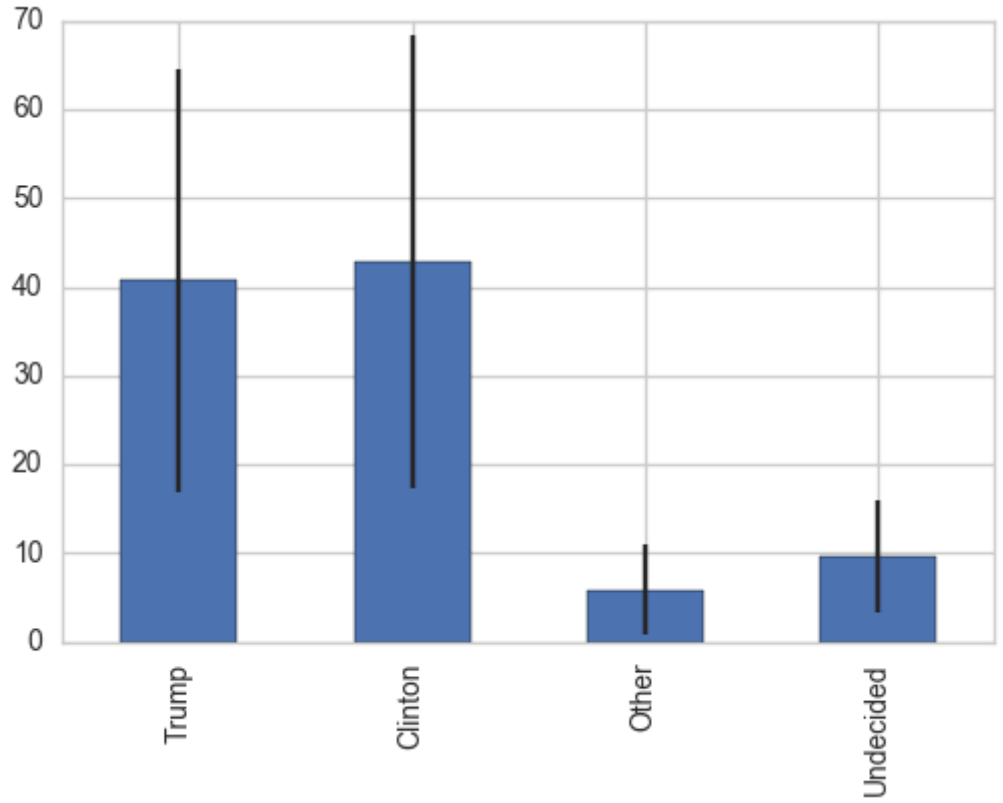
In [16]:
```
std.head()
```

Out[16]:

| | 0 |
|---|---|
| **Trump** | 23.735443 |
| **Clinton** | 25.477141 |
| **Other** | 5.118286 |
| **Undecided** | 6.383558 |

In [17]:
```
avg.plot(yerr = std, kind = 'bar', legend = False)
```

Out[17]: <matplotlib.axes._subplots.AxesSubplot at 0x1178fd550>



| | Average | STD |
|---|---|---|

Since Trump and Clinton is very closed to each other, the 'Undecided' can make a big difference in who will win the race

In [18]:
```
poll_avg = pd.concat([avg, std], axis = 1)
```

In [19]:
```
poll_avg.columns = ['Average', 'STD']
poll_avg
```

Out[19]:

| | Average | STD |
|---|---|---|
| **Trump** | 40.727669 | 23.735443 |
| **Clinton** | 42.792302 | 25.477141 |
| **Other** | 5.921859 | 5.118286 |
| **Undecided** | 9.660363 | 6.383558 |

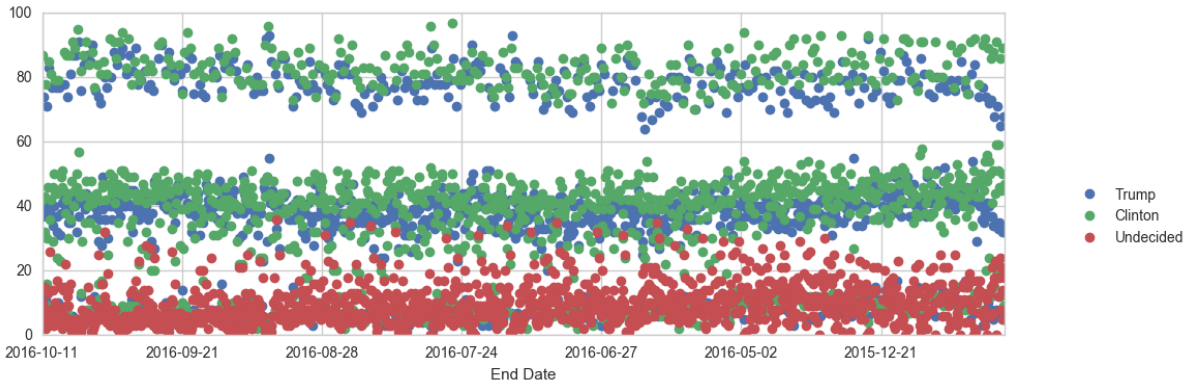
In [20]:
```
poll_df.head()
```

Out[20]:

| | Pollster | Start Date | End Date | Entry Date/Time (ET) | Number of Observations | Population | Mode | Trump |
|---|---|---|---|---|---|---|---|---|
| | **Pollster** | **Start Date** | **End Date** | **Date/Time (ET)** | **Observations** | **Population** | **Mode** | **Trump** |
| **0** | Insights West | 2016-10-10 | 2016-10-11 | 2016-10-12T14:34:44Z | 953.0 | Likely Voters | Internet | 37.0 |
| **1** | Insights West | 2016-10-10 | 2016-10-11 | 2016-10-12T14:34:44Z | NaN | Likely Voters - Democrat | Internet | 5.0 |
| **2** | Insights West | 2016-10-10 | 2016-10-11 | 2016-10-12T14:34:44Z | NaN | Likely Voters - Republican | Internet | 74.0 |
| **3** | Insights West | 2016-10-10 | 2016-10-11 | 2016-10-12T14:34:44Z | NaN | Likely Voters - independent | Internet | 31.0 |
| **4** | Rasmussen | 2016-10-07 | 2016-10-11 | 2016-10-12T12:39:57Z | 1500.0 | Likely Voters | IVR/Online | 39.0 |

In [21]:
```
poll_df.plot(x = 'End Date', y = ['Trump', 'Clinton', 'Undecided'], figsize = (12,4), linestyle = '', marker = 'o').legend(bbox_to_anchor=(1.2, 0
```

Out[21]: <matplotlib.legend.Legend at 0x1179cd650>




The graph scatter pretty evenly thoroughout the time line. Clinton seems to always on top of Trump

In [22]:
```
from datetime import datetime
```

In [23]:
```
poll_df['Difference'] = (poll_df.Clinton - poll_df.Trump)/100
```

In [24]:
```
poll_df.head()
```

Out[24]:

| | Pollster | Start Date | End Date | Entry Date/Time (ET) | Number of Observations | Population | Mode | Trump |
|---|---|---|---|---|---|---|---|---|
| | **Pollster** | **Start Date** | **End Date** | **Date/Time (ET)** | **Observations** | **Population** | **Mode** | **Trump** |
| **0** | Insights West | 2016-10-10 | 2016-10-11 | 2016-10-12T14:34:44Z | 953.0 | Likely Voters | Internet | 37.0 |
| **1** | Insights West | 2016-10-10 | 2016-10-11 | 2016-10-12T14:34:44Z | NaN | Likely Voters - Democrat | Internet | 5.0 |
| **2** | Insights West | 2016-10-10 | 2016-10-11 | 2016-10-12T14:34:44Z | NaN | Likely Voters - Republican | Internet | 74.0 |
| **3** | Insights West | 2016-10-10 | 2016-10-11 | 2016-10-12T14:34:44Z | NaN | Likely Voters - independent | Internet | 31.0 |
| **4** | Rasmussen | 2016-10-07 | 2016-10-11 | 2016-10-12T12:39:57Z | 1500.0 | Likely Voters | IVR/Online | 39.0 |

Positive number in the Difference column means Clinton is leading Negative number in the Difference column means Trump is leading

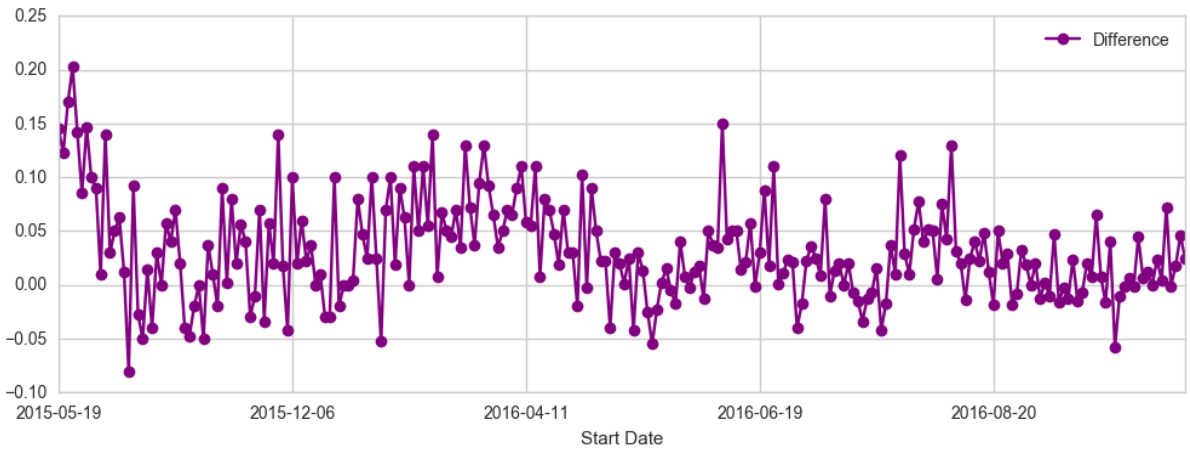
In [25]:
```
poll_df = poll_df.groupby(['Start Date'], as_index = False).mean()
poll_df.head()
```

Out[25]:

| | Start Date | Number of Observations | Trump | Clinton | Other | Undecided | Question Iteration | Difference |
|---|---|---|---|---|---|---|---|---|
| **0** | 2015-05-19 | 1046.00 | 34.25 | 48.75 | 2.5 | 14.00 | 1.0 | 0.1450 |
| **1** | 2015-06-20 | 420.75 | 35.00 | 47.25 | NaN | 17.75 | 1.0 | 0.1225 |
| **2** | 2015-06-21 | 1005.00 | 34.00 | 51.00 | 3.0 | 12.00 | 1.0 | 0.1700 |
| **3** | 2015-06-26 | 890.00 | 36.75 | 57.00 | 6.0 | 0.00 | 1.0 | 0.2025 |
| **4** | 2015-07-09 | 499.25 | 35.25 | 49.50 | NaN | 16.00 | 1.0 | 0.1425 |

In [26]:
```
poll_df.plot('Start Date', 'Difference', figsize = (12,4), marker = 'o',
 linestyle = '-', color = 'purple')
```

Out[26]: `<matplotlib.axes._subplots.AxesSubplot at 0x117ae0cd0>`



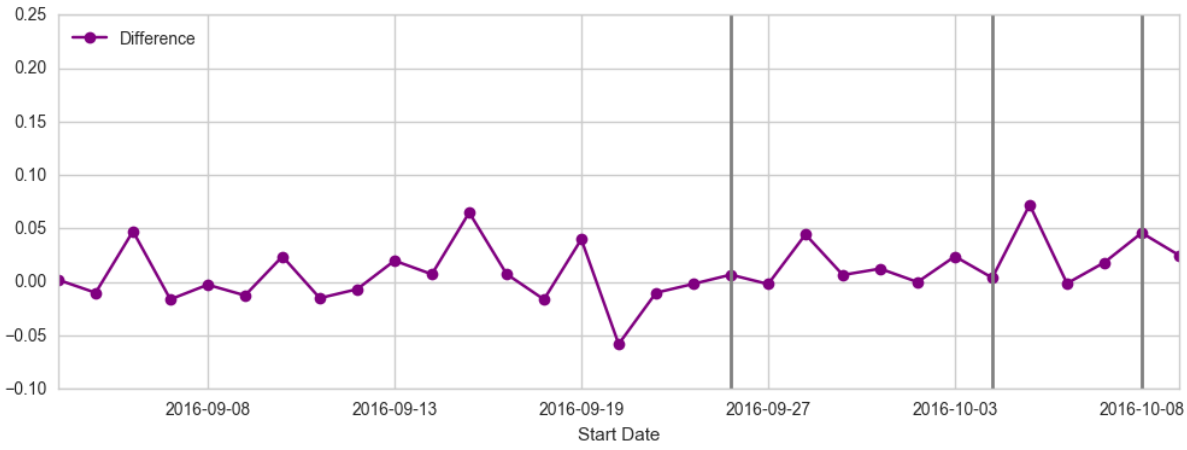

Up means people favor Clintin more, Down means people favor Trump more

In [27]:
```
row_in = 0
xlimit = []

for date in poll_df['Start Date']:
    if date[0:7] == '2016-09':
        xlimit.append(row_in)
        row_in += 1
    else:
        row_in += 1

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

```
211
233
```

In [28]:
```
poll_df.plot('Start Date', 'Difference', figsize = (12,4), marker = 'o',
 linestyle = '-', color = 'purple', xlim = (211,241))

#Sep 26 First Debate
plt.axvline(x = 211 + 18, linewidth = 2, color = 'grey')
#Tues 4 Vice Pre Debate
plt.axvline(x = 211 + 25, linewidth = 2, color = 'grey')
#Oct 8 Trump Scandal Tape
plt.axvline(x = 211+ 29, linewidth = 2, color = 'grey')
```

Out[28]: <matplotlib.lines.Line2D at 0x117ffc890>




## CAMPAIGN DONATIONS FOR TRUMP AND CLINTON

In [29]:
```
Trump_donor = pd.read_csv('TrumpDonation.csv', index_col = False, dtype={
x. CPC': pd.np.float64})
```

```
/Users/tedlam/anaconda/envs/py27/lib/python2.7/site-packages/IPython/co
re/interactiveshell.py:2717: DtypeWarning: Columns (11) have mixed type
s. Specify dtype option on import or set low_memory=False.
  interactivity=interactivity, compiler=compiler, result=result)
```

In [30]:
```
Trump_donor
```

Out[30]:

| | cmte_id | cand_id | cand_nm | contbr_nm | contbr_city | contbr_ |
|---|---|---|---|---|---|---|
| **0** | C00580100 | P80001571 | Trump, Donald J. | ECKERSALL, LAWRENCE | DPO | AA |
| **1** | C00580100 | P80001571 | Trump, Donald J. | VANBEUGE, ROBERT | APO | AA |
| **2** | C00580100 | P80001571 | Trump, Donald J. | VANBEUGE, ROBERT | APO | AA |
| **3** | C00580100 | P80001571 | Trump, Donald J. | VANBEUGE, ROBERT | APO | AA |
| **4** | C00580100 | P80001571 | Trump, Donald J. | RIVERA, JOSEPH | APO | AE |
| **5** | C00580100 | P80001571 | Trump, Donald J. | RIVERA, JOSEPH | APO | AE |
| **6** | C00580100 | P80001571 | Trump, Donald J. | LINDELL, TIMOTHY | FPO | AE |
| **7** | C00580100 | P80001571 | Trump, Donald J. | LINDELL, TIMOTHY | FPO | AE |
| **8** | C00580100 | P80001571 | Trump, Donald J. | LECOMTE, BEATRIZ | APO | AE |
| **9** | C00580100 | P80001571 | Trump, Donald J. | NASIR, SHAHRIYAR | TORONTO | AE |
| **10** | C00580100 | P80001571 | Trump, Donald J. | LINDELL, TIMOTHY | FPO | AE |
| **11** | C00580100 | P80001571 | Trump, Donald J. | COLLINS, DAVE | FPO | AE |
| **12** | C00580100 | P80001571 | Trump, Donald J. | COLLINS, DAVE | FPO | AE |
| **13** | C00580100 | P80001571 | Trump, Donald J. | COLLINS, DAVE | FPO | AE |
| **14** | C00580100 | P80001571 | Trump, Donald J. | LINDELL, TIMOTHY | FPO | AE |
| **15** | C00580100 | P80001571 | Trump, Donald J. | COLLINS, DAVE | FPO | AE |
| **16** | C00580100 | P80001571 | Trump, Donald J. | LASAGE-BISHOP, JEAN-PIERRE | FAIRBANKS | AK |

| | cmte_id | cand_id | cand_nm | contbr_nm | contbr_city | contbr_ |
|---|---|---|---|---|---|---|
| **17** | C00580100 | P80001571 | Trump, Donald J. | EGGER, PATRICK | WASILLA | AK |
| **18** | C00580100 | P80001571 | Trump, Donald J. | MECHON, LINDA | ANCHORAGE | AK |
| **19** | C00580100 | P80001571 | Trump, Donald J. | COWDERY, PAMELA | ANCHORAGE | AK |
| **20** | C00580100 | P80001571 | Trump, Donald J. | CUELLAR, BEN | ANCHORAGE | AK |
| **21** | C00580100 | P80001571 | Trump, Donald J. | JACKSON, SHERRI | ANCHORAGE | AK |
| **22** | C00580100 | P80001571 | Trump, Donald J. | JANELLI, REBECCA T | ANCHORAGE | AK |
| **23** | C00580100 | P80001571 | Trump, Donald J. | JANELLI, REBECCA T | ANCHORAGE | AK |
| **24** | C00580100 | P80001571 | Trump, Donald J. | RICE, JIMMY | SOLDOTNA | AK |
| **25** | C00580100 | P80001571 | Trump, Donald J. | RICE, JIMMY | SOLDOTNA | AK |
| **26** | C00580100 | P80001571 | Trump, Donald J. | BARNES, WILLIAM | PALMER | AK |
| **27** | C00580100 | P80001571 | Trump, Donald J. | HORJES, JOHN | ANCHORAGE | AK |
| **28** | C00580100 | P80001571 | Trump, Donald J. | MANNING, SLADE | ANCHORAGE | AK |
| **29** | C00580100 | P80001571 | Trump, Donald J. | MINOGUE, JOHN | NORTH POLE | AK |
| **...** | ... | ... | ... | ... | ... | ... |
| **394578** | C00580100 | P80001571 | Trump, Donald J. | FANTASKEY, MATTHEW | WORLAND | WY |
| **394579** | C00580100 | P80001571 | Trump, Donald J. | CUBIN, FREDERICK | CASPER | WY |
| **394580** | C00580100 | P80001571 | Trump, Donald J. | CUBIN, WILLIAM | CASPER | WY |
| **394581** | C00580100 | P80001571 | Trump, Donald J. | FLYNN, DENNIS | GLENROCK | WY |

| | cmte_id | cand_id | cand_nm | contbr_nm | contbr_city | contbr_ |
|---|---|---|---|---|---|---|
| **394582** | C00580100 | P80001571 | Trump, Donald J. | CUNDY, BRAD | MILLS | WY |
| **394583** | C00580100 | P80001571 | Trump, Donald J. | BUSH, BRIAN | GILLETTE | WY |
| **394584** | C00580100 | P80001571 | Trump, Donald J. | COSNER, DIANNA | WRIGHT | WY |
| **394585** | C00580100 | P80001571 | Trump, Donald J. | GABRIELSON, STEVEN | JACKSON | WY |
| **394586** | C00580100 | P80001571 | Trump, Donald J. | GABRIELSON, STEVEN | JACKSON | WY |
| **394587** | C00580100 | P80001571 | Trump, Donald J. | NORRIS, JOHN MR. | RAWLINS | WY |
| **394588** | C00580100 | P80001571 | Trump, Donald J. | MCQUISTEN, ROBERT E MR. | CHEYENNE | WY |
| **394589** | C00580100 | P80001571 | Trump, Donald J. | MAGARITY, RUSSELL | WILSON | WY |
| **394590** | C00580100 | P80001571 | Trump, Donald J. | MAIER, JANICE MRS. | SHERIDAN | WY |
| **394591** | C00580100 | P80001571 | Trump, Donald J. | MEYER, E.G. | LARAMIE | WY |
| **394592** | C00580100 | P80001571 | Trump, Donald J. | WOLF, MARVIN MR. | CHEYENNE | WY |
| **394593** | C00580100 | P80001571 | Trump, Donald J. | MADDEN, SHAWN | TORRINGTON | WY |
| **394594** | C00580100 | P80001571 | Trump, Donald J. | PAYNE, KAREN R MS. | LARAMIE | WY |
| **394595** | C00580100 | P80001571 | Trump, Donald J. | WORMUS, PATSY L MRS. | CASPER | WY |
| **394596** | C00580100 | P80001571 | Trump, Donald J. | WYNN, MARILYN | EVANSTON | WY |
| **394597** | C00580100 | P80001571 | Trump, Donald J. | YOUNG, TAMARA | CODY | WY |
| **394598** | C00580100 | P80001571 | Trump, Donald J. | WOLF, JOY MS. | CODY | WY |

| | cmte_id | cand_id | cand_nm | contbr_nm | contbr_city | contbr_ |
|---|---|---|---|---|---|---|
| **394599** | C00580100 | P80001571 | Trump, Donald J. | WOLF, JOY MS. | CODY | WY |
| **394600** | C00580100 | P80001571 | Trump, Donald J. | MARTTINEN, MIKKO | LAHTI | ZZ |
| **394601** | C00580100 | P80001571 | Trump, Donald J. | JONES, KEVIN | APO | ZZ |
| **394602** | C00580100 | P80001571 | Trump, Donald J. | ZALESKI, HENRYK | HORTEN,3181,NORWAY | ZZ |
| **394603** | C00580100 | P80001571 | Trump, Donald J. | PIERROT, MARC | HONG KONG | ZZ |
| **394604** | C00580100 | P80001571 | Trump, Donald J. | GEDDES, C | AUCKLAND NEW ZEALAND | ZZ |
| **394605** | C00580100 | P80001571 | Trump, Donald J. | GELFAND, BEN | TORONTO | ZZ |
| **394606** | C00580100 | P80001571 | Trump, Donald J. | SYKES, C E R | HUDDERSFIELD ENGLAND | ZZ |
| **394607** | C00580100 | P80001571 | Trump, Donald J. | ZALESKI, HENRYK | HORTEN,3181,NORWAY | ZZ |

394608 rows × 18 columns

In [31]:
```
Clinton_donor = pd.read_csv('Clinton.csv', index_col = False, dtype={'Ma
x. CPC': pd.np.float64})
```

```
/Users/tedlam/anaconda/envs/py27/lib/python2.7/site-packages/IPython/co
re/interactiveshell.py:2717: DtypeWarning: Columns (6,11) have mixed ty
pes. Specify dtype option on import or set low_memory=False.
  interactivity=interactivity, compiler=compiler, result=result)
```

In [32]:
```
Clinton_donor.head()
```

Out[32]:

| | cmte_id | cand_id | cand_nm | contbr_nm | contbr_city | contbr_st | contbr_zip | |
|---|---|---|---|---|---|---|---|---|
| **0** | C00575795 | P00003392 | Clinton, Hillary Rodham | LARSEN, CARLENE | DPO | AA | 3.4005e+08 | |
| **1** | C00575795 | P00003392 | Clinton, Hillary Rodham | ABUNDO, FRANK | MONTREUX | AA | NaN | |
| **2** | C00575795 | P00003392 | Clinton, Hillary Rodham | FRASCARELLI, RAFFAELLA | ROME | AA | NaN | |
| **3** | C00575795 | P00003392 | Clinton, Hillary Rodham | FRIES-LAFFAILLE, SUSAN | PARIS | AA | NaN | |
| **4** | C00575795 | P00003392 | Clinton, Hillary Rodham | ROBINSON, TODD | DPO | AA | 3.4024e+08 | |

In [33]:
```
donor = [Trump_donor, Clinton_donor]
```

In [34]:
```
donor_df = pd.concat(donor)
```

In [35]:
```
donor_df.head()
```

Out[35]:

| | cmte_id | cand_id | cand_nm | contbr_nm | contbr_city | contbr_st | contbr_zip | cor |
|---|---|---|---|---|---|---|---|---|
| **0** | C00580100 | P80001571 | Trump, Donald J. | ECKERSALL, LAWRENCE | DPO | AA | 34004 | L-3 CO VEF AEF |
| **1** | C00580100 | P80001571 | Trump, Donald J. | VANBEUGE, ROBERT | APO | AA | 34022 | U.S OF |
| **2** | C00580100 | P80001571 | Trump, Donald J. | VANBEUGE, ROBERT | APO | AA | 34022 | U.S OF |
| **3** | C00580100 | P80001571 | Trump, Donald J. | VANBEUGE, ROBERT | APO | AA | 34022 | U.S OF |
| **4** | C00580100 | P80001571 | Trump, Donald J. | RIVERA, JOSEPH | APO | AE | 9643 | USA |

In [36]:
```
donor_df['contb_receipt_amt'].value_counts()
```

Out[36]:

| | |
|---|---|
| 25.00 | 324509 |
| 100.00 | 204168 |
| 50.00 | 191429 |
| 5.00 | 169088 |
| 10.00 | 155761 |
| 250.00 | 102702 |
| 19.00 | 88725 |
| 8.00 | 65093 |
| 40.00 | 63633 |
| 80.00 | 56243 |
| 20.00 | 53380 |
| 28.00 | 53261 |
| 38.00 | 50776 |
| 2700.00 | 49041 |
| 75.00 | 45013 |
| 200.00 | 42884 |
| 500.00 | 41372 |
| 1000.00 | 33567 |
| 15.00 | 30262 |
| 3.00 | 25604 |
| 1.00 | 21764 |
| 16.00 | 20028 |
| 150.00 | 13968 |
| 30.00 | 10569 |
| 160.00 | 9354 |
| 24.00 | 8395 |
| 4.00 | 7477 |
| 2.40 | 7312 |
| 11.00 | 6931 |
| 12.00 | 6907 |
| ... | |
| 12000.00 | 1 |
| 496.13 | 1 |
| 262.39 | 1 |
| 4.55 | 1 |
| 261.36 | 1 |
| 239.55 | 1 |
| 52.73 | 1 |
| -48.07 | 1 |
| 52.27 | 1 |
| 112.92 | 1 |
| 1198.55 | 1 |
| 1192.70 | 1 |
| 507.12 | 1 |
| 6.96 | 1 |
| -47.75 | 1 |
| 234.30 | 1 |
| 113.17 | 1 |
| 1188.30 | 1 |
| 34.26 | 1 |
| 234.55 | 1 |
| 113.08 | 1 |
| 232.70 | 1 |
| 1302.87 | 1 |
| 52.77 | 1 |
| 233.05 | 1 |
| 112.83 | 1 |

```
 217.68           1
 266.64           1
 232.45           1
 155.78           1
Name: contb_receipt_amt, dtype: int64
```

In [37]:
```
don_mean = donor_df['contb_receipt_amt'].mean()

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

print 'The average donation was %.2f with a std %.2f' %(don_mean, don_st
d)
```

```
The average donation was 177.01 with a std 12444.56
```

In [38]:
```
top_donor = donor_df['contb_receipt_amt'].copy()

top_donor.sort()

top_donor
```

```
/Users/tedlam/anaconda/envs/py27/lib/python2.7/site-packages/ipykernel/
__main__.py:3: FutureWarning: sort is deprecated, use sort_values(inpla
ce=True) for INPLACE sorting
  app.launch_new_instance()
```

Out[38]:

| | |
|---|---|
| 175921 | -84236.80 |
| 545860 | -20000.00 |
| 381740 | -6574.19 |
| 376112 | -5463.57 |
| 376113 | -5463.57 |
| 381739 | -5459.49 |
| 940085 | -5400.00 |
| 1252498 | -5400.00 |
| 573376 | -5400.00 |
| 908439 | -5400.00 |
| 1217962 | -5400.00 |
| 1217945 | -5400.00 |
| 118675 | -5400.00 |
| 1217794 | -5400.00 |
| 231077 | -5400.00 |
| 1212059 | -5400.00 |
| 1588433 | -5400.00 |
| 265666 | -5400.00 |
| 143123 | -5400.00 |
| 1211966 | -5400.00 |
| 143207 | -5400.00 |
| 930067 | -5400.00 |
| 26220 | -5400.00 |
| 545756 | -5400.00 |
| 1168682 | -5400.00 |
| 70373 | -5200.00 |
| 70139 | -5200.00 |
| 1143210 | -5000.00 |
| 265529 | -5000.00 |
| 1170202 | -5000.00 |
| | ... |
| 1266583 | 5400.00 |
| 586205 | 5400.00 |
| 327652 | 5400.00 |
| 65054 | 5400.00 |
| 558066 | 5400.00 |
| 1558047 | 5400.00 |
| 1237448 | 5400.00 |
| 390485 | 5400.00 |
| 395030 | 7300.00 |
| 91965 | 10000.00 |
| 163957 | 10000.00 |
| 228782 | 10000.00 |
| 210306 | 10000.00 |
| 380550 | 10000.00 |
| 378837 | 10030.24 |
| 255560 | 12000.00 |
| 118669 | 12500.00 |
| 557920 | 20000.00 |
| 175999 | 86936.80 |
| 1333268 | 1467070.94 |
| 1323962 | 1603724.41 |
| 1240153 | 1797624.94 |
| 1312765 | 2069248.84 |
| 1369111 | 3600489.08 |
| 1242816 | 3686373.26 |
| 1367848 | 4575438.56 |

```
1330636     4904860.51
1290608     6358481.89
1314211     7402361.45
1165273    12777705.58
Name: contb_receipt_amt, dtype: float64
```

In [39]:
```
top_donor = top_donor [top_donor > 0]

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

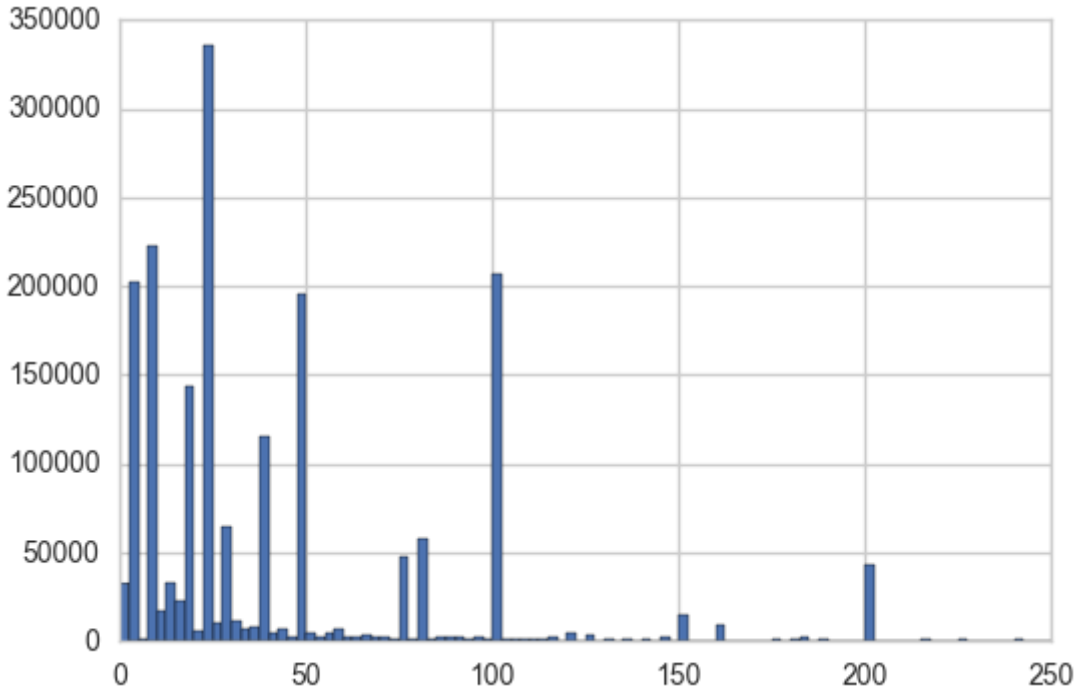
```
/Users/tedlam/anaconda/envs/py27/lib/python2.7/site-packages/ipykernel/
__main__.py:3: FutureWarning: sort is deprecated, use sort_values(inpla
ce=True) for INPLACE sorting
  app.launch_new_instance()
```

Out[39]:
```
25.0     324509
100.0    204168
50.0     191429
5.0      169088
10.0     155761
250.0    102702
19.0      88725
8.0       65093
40.0      63633
80.0      56243
Name: contb_receipt_amt, dtype: int64
```

The top most donation amount

In [40]:
```
com_don = top_donor[top_donor < 250]

com_don.hist(bins = 100)
```

Out[40]: <matplotlib.axes._subplots.AxesSubplot at 0x13bc181d0>




The graph of most common donations

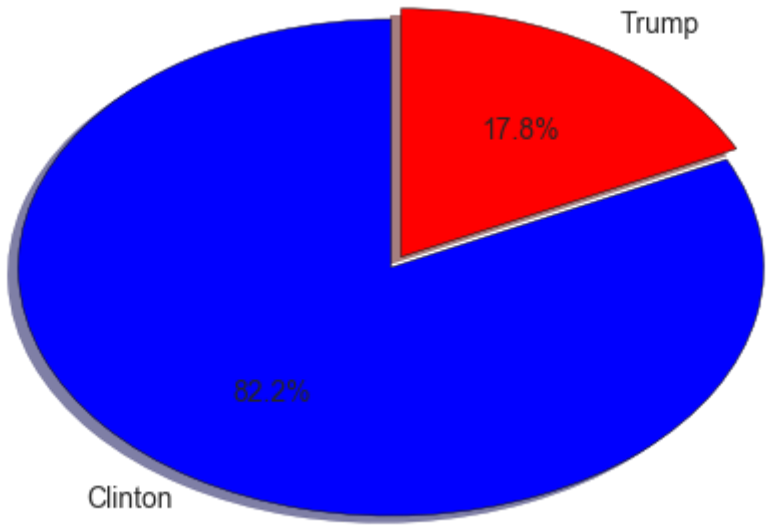
In [41]:
```
donor_df.groupby('cand_nm')['contb_receipt_amt'].count()
```

Out[41]:
```
cand_nm
Clinton, Hillary Rodham    1818440
Trump, Donald J.            394608
Name: contb_receipt_amt, dtype: int64
```

In [42]:
```
from pylab import *
```

In [43]:
```
number = [1818440, 394608]
labels = 'Clinton', 'Trump'
explode=(0, 0.05)
colors = 'b', 'r'
pie(number, colors = colors, explode = explode, labels = labels, autopct
 = '%1.1f%%',shadow = True, startangle = 90)
```

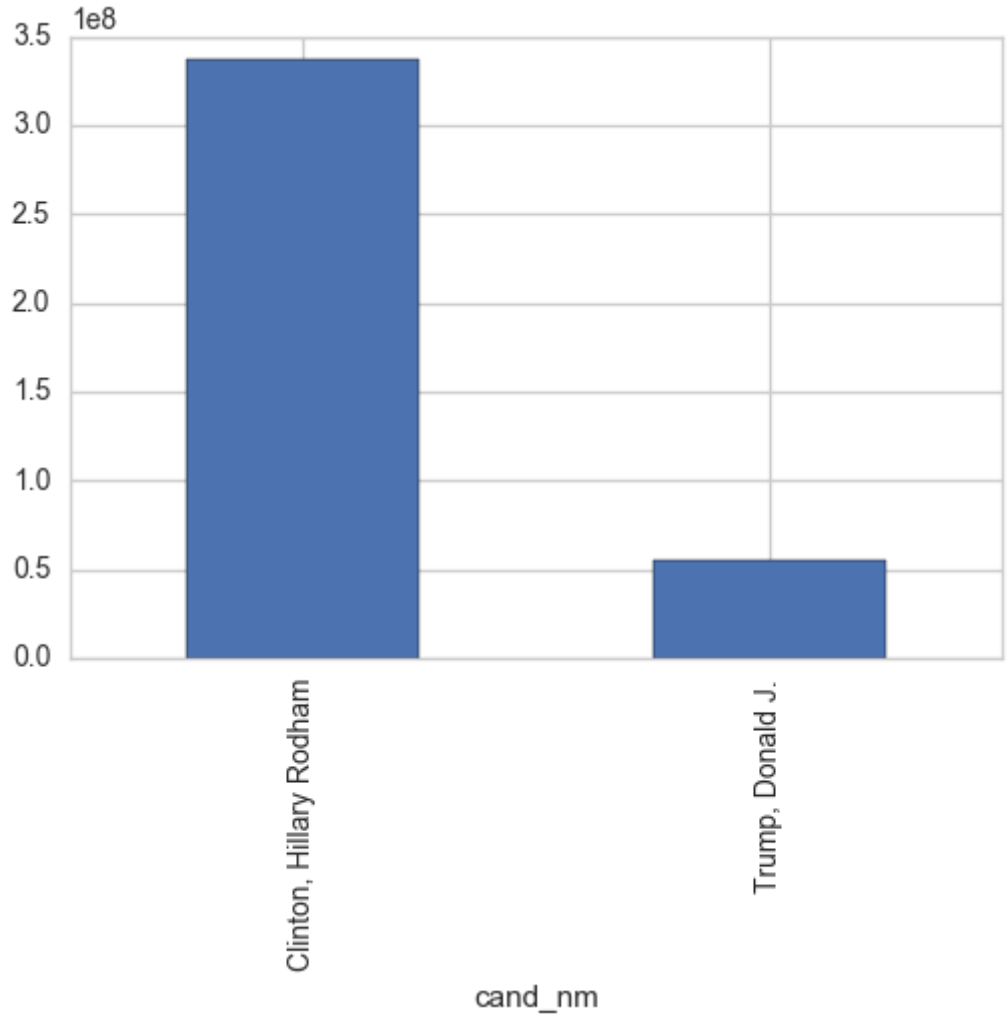
Out[43]:
```
([<matplotlib.patches.Wedge at 0x1186f4850>,
  <matplotlib.patches.Wedge at 0x1188c0bd0>],
 [<matplotlib.text.Text at 0x1188c02d0>,
  <matplotlib.text.Text at 0x1188cc650>],
 [<matplotlib.text.Text at 0x1188c0790>,
  <matplotlib.text.Text at 0x1188cca90>])
```




Clinton has 82.2% of donor

In [44]:
```
cand_amount = donor_df.groupby('cand_nm') ['contb_receipt_amt'].sum()
cand_amount.plot(kind = 'bar')
```

Out[44]: <matplotlib.axes._subplots.AxesSubplot at 0x1198e47d0>




In [46]:
```
pwd
```

Out[46]: u'/Users/tedlam'

In [ ]: