SurveyPrep

March 22, 2016

0.1 Upload data for all candidates

In [1]:
```
from textstat.textstat import textstat
import csv
import pandas
import matplotlib
#matplotlib.style.use('ggplot')
%matplotlib inline
import ast
pandas.options.display.max_colwidth = 10000
```

In [2]:
```
trump_df = pandas.read_csv('data/all_trump_w_topics.csv')
trump_df['candidate'] = 'trump'
clinton_df = pandas.read_csv('data/all_clinton_w_topics.csv')
clinton_df['candidate'] = 'clinton'
sanders_df = pandas.read_csv('data/all_sanders_w_topics.csv')
sanders_df['candidate'] = 'sanders'
cruz_df = pandas.read_csv('data/all_cruz_w_topics.csv')
cruz_df['candidate'] = 'cruz'
ORGS = ['nyt', 'wsj', 'cnn', 'fox', 'ap', 'reuters', 'politico', 'mcclatchy', 'buzzfeed', 'huff
```

In [3]:
```
n = len(clinton_df)
clinton_df.index = xrange(len(trump_df), (len(trump_df) + n))
m = len(sanders_df)
sanders_df.index = xrange(max(clinton_df.index), max(clinton_df.index) + m)
c = len(cruz_df)
cruz_df.index = xrange(max(sanders_df.index), max(sanders_df.index) + c)
```

In [4]:
```
all_df = pandas.concat([trump_df,clinton_df, sanders_df, cruz_df])
all_df['gunning_fog'] = all_df['body'].apply(lambda x: textstat.gunning_fog(x) if type(x) == st
all_df['flesch'] = all_df['body'].apply(lambda x: textstat.flesch_kincaid_grade(x) if type(x) =
all_df['readability'] = all_df['body'].apply(lambda x: textstat.flesch_reading_ease(x) if type(
```

0.2 Topic Processing

In [5]:
```
all_df['topic_dict'] = all_df['topic'].apply(lambda d: ast.literal_eval(d))
all_df['top_topics'] = all_df['topic_dict'].apply(lambda d: {k:v for k, v in d.iteritems() if v
all_df['topic_list'] = all_df['top_topics'].apply(lambda d: d.keys())
all_df['top_topic'] = all_df['topic_dict'].apply(lambda d: max(d, key=lambda i: d[i]))
```

0.3 Remove Personality Topic (it's uncategorized)

In [8]:
```
all_nop = all_df[all_df['top_topic']!= "Personality"]
t = len(all_nop)
```

```
        print "TOTAL, NO P:", t
        low_nop = all_nop[all_nop['flesch'] < 6]
        high_nop = all_nop[all_nop['flesch'] > 12]
        mid_nop = all_nop[(all_nop['flesch'] > 8) & (all_nop['flesch'] < 10)]
        print "LESS THAN 6:", len(low_nop), '%.2f' % (100* len(low_nop)/(len(all_nop) * 1.0)), '%'
        print "GREATER THAN 12:", len(high_nop), '%.2f' % (100* len(high_nop)/(len(all_nop) * 1.0)), '%
        print "MIDDLE 8-10:", len(mid_nop), '%.2f' % (100* len(mid_nop)/(len(all_nop) * 1.0)), '%'
```

```
TOTAL, NO P: 2108
LESS THAN 6: 86 4.08 %
GREATER THAN 12: 190 9.01 %
MIDDLE 8-10: 775 36.76 %
```

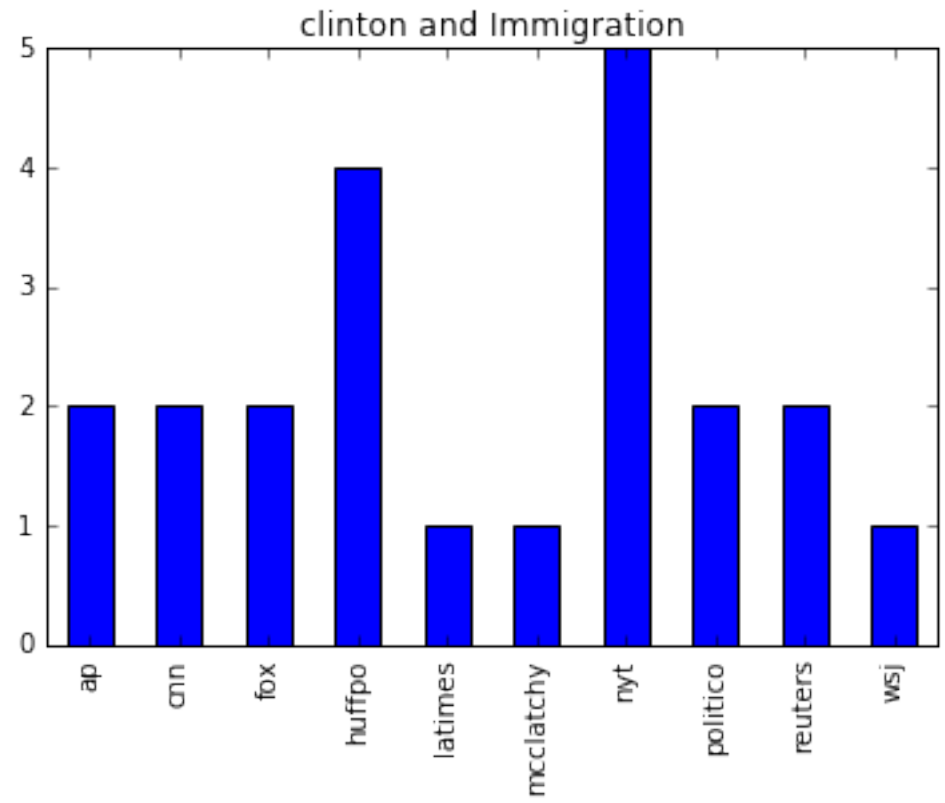
```
In [13]: all_nop.to_csv('data/all_candidates_nop.csv')
```

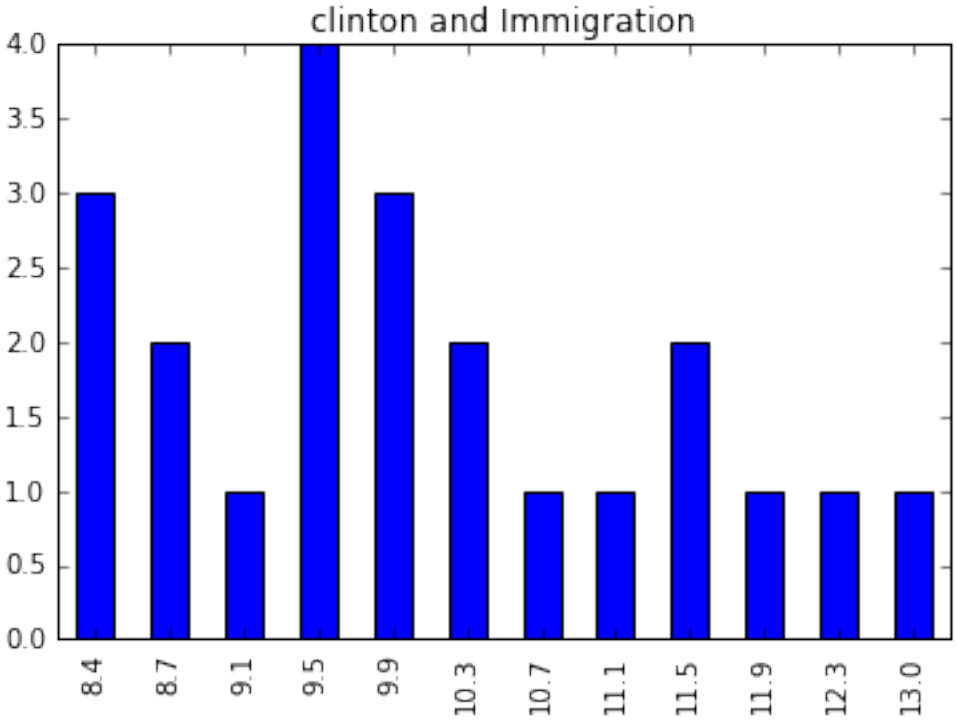
0.4 Split into Topic and +/-/o per candidate

```
In [44]: CANDIDATES = ['clinton', 'sanders', 'trump', 'cruz']
         TOPICS = ['Immigration', 'Campaign Finance', 'Foreign Policy/National Security',
                   'Abortion', 'Health Care', 'Economy', 'Ethics', 'Racial Issues',
                   'Financial Regulation', 'LGBT Issues']
```

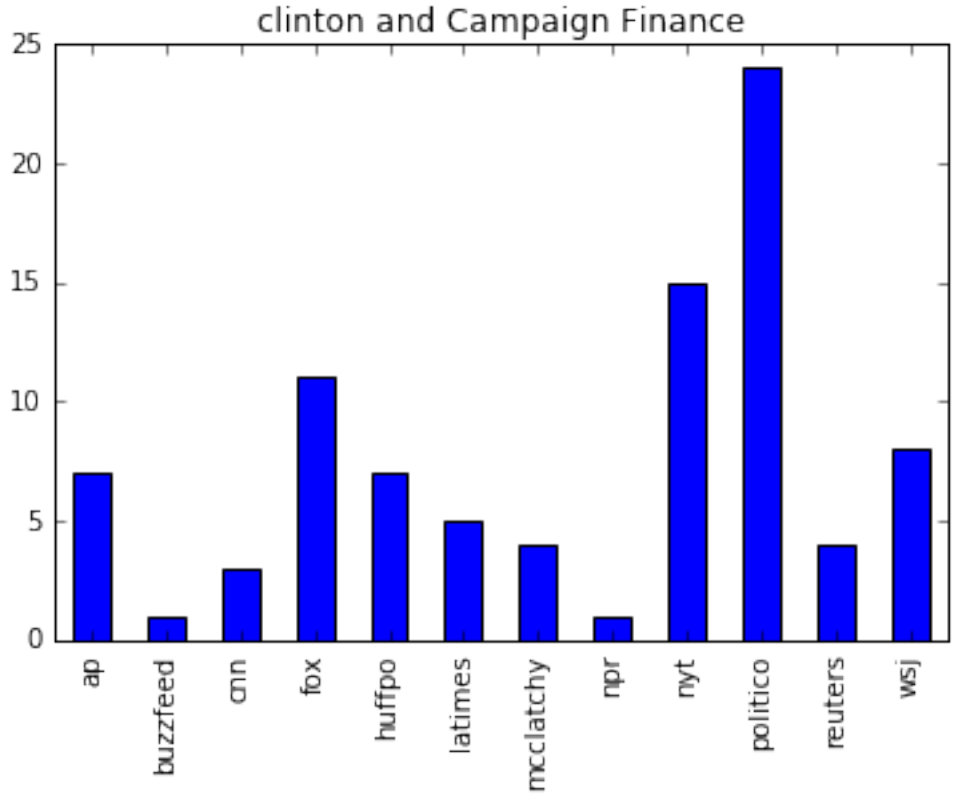
```
In [60]: for c in CANDIDATES:
             #all_nop[all_nop['candidate'] == c]['top_topic'].value_counts().plot(kind="bar", title="To
             #matplotlib.pyplot.show()
             print c
             print
             for t in TOPICS:
                 print t
                 #print all_nop[(all_nop['candidate'] == c) & (all_nop['top_topic'] == t)]['org'].value
                 all_nop[(all_nop['candidate'] == c) & (all_nop['top_topic'] == t)]['org'].value_counts
                                     .sort_index()\
                                     .plot(kind="bar", title= c + " and " + t)
                 matplotlib.pyplot.show()
                 all_nop[(all_nop['candidate'] == c) & (all_nop['top_topic'] == t)]['flesch'].value_cou
                                     .sort_index()\
                                     .plot(kind="bar", title= c + " and " + t)
                 matplotlib.pyplot.show()
                 print
```

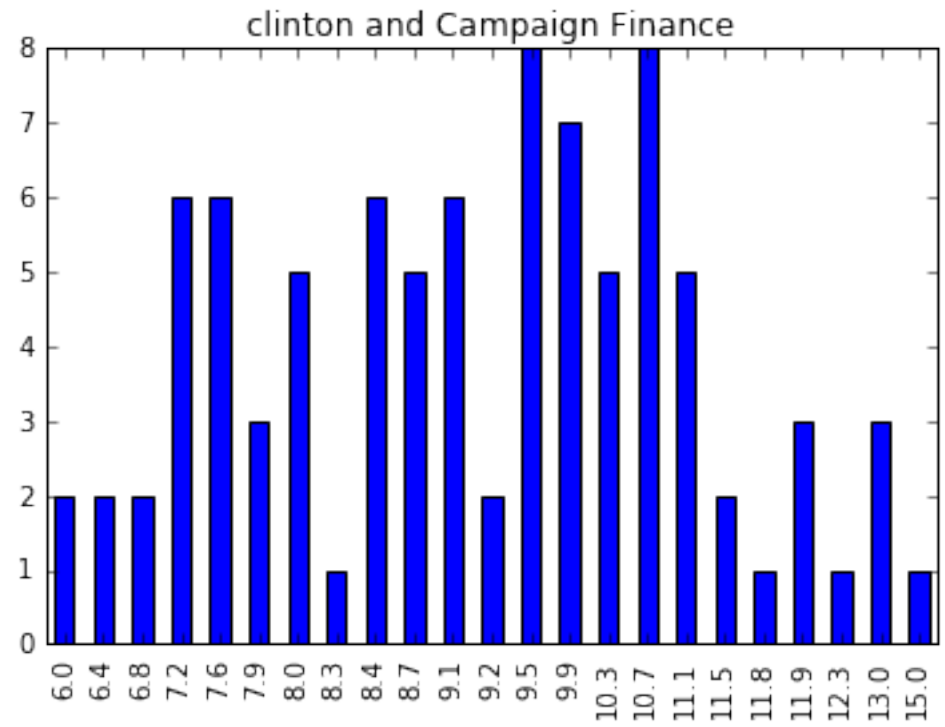
```
clinton

Immigration
```


clinton and Immigration
5
4
3
2
1
0
ap
cnn
fox
huffpo
latimes
mcclatchy
nyt
politico
reuters
wsj

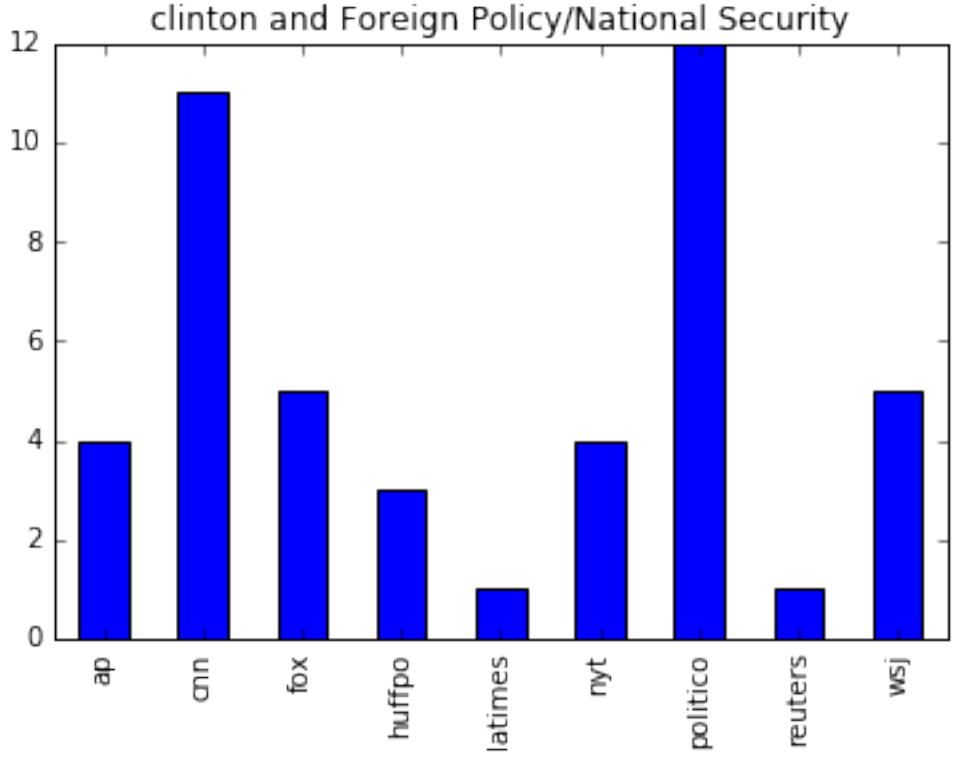

clinton and Immigration
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
8.4
8.7
9.1
9.5
9.9
10.3
10.7
11.1
11.5
11.9
12.3
13.0

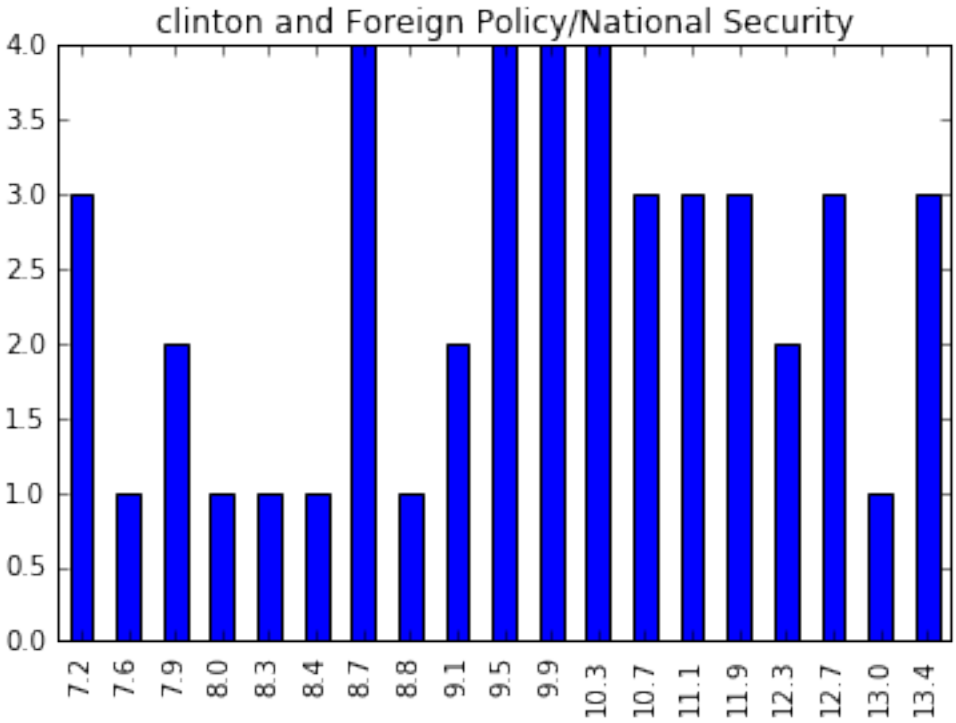
Campaign Finance



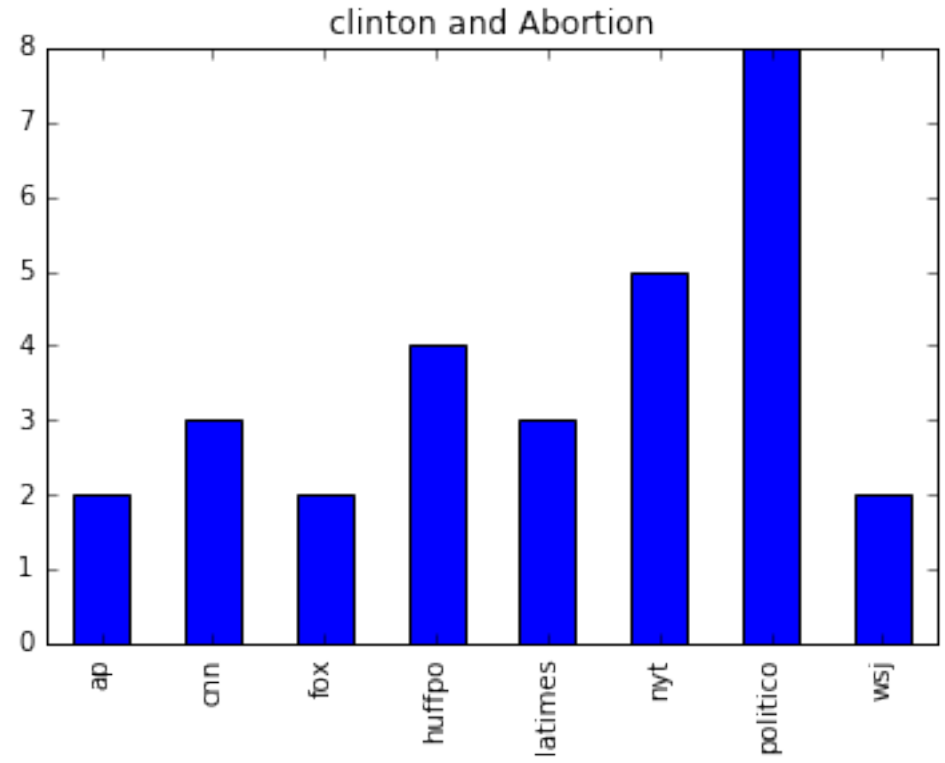


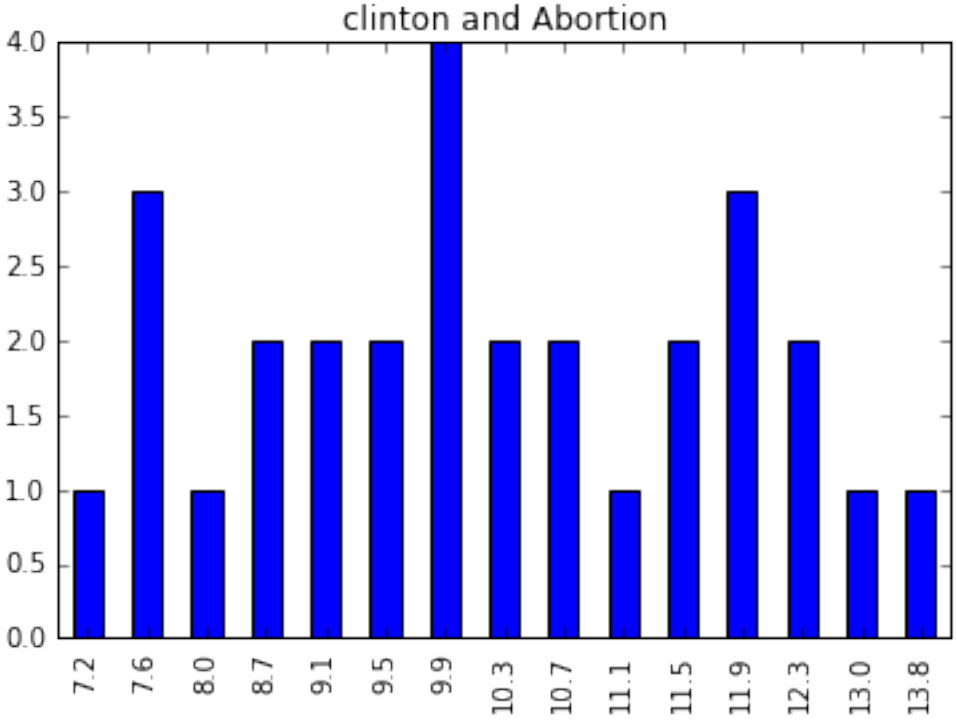
Foreign Policy/National Security



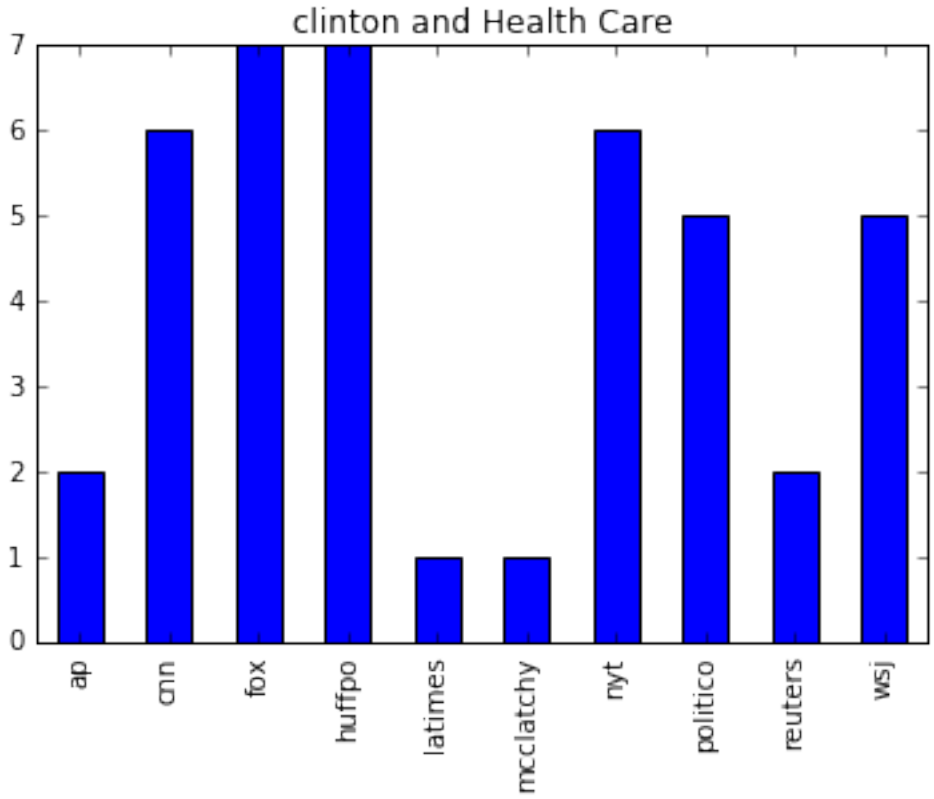

clinton and Foreign Policy/National Security
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
7.2
7.6
7.9
8.0
8.3
8.4
8.7
8.8
9.1
9.5
9.9
10.3
10.7
11.1
11.9
12.3
12.7
13.0
13.4

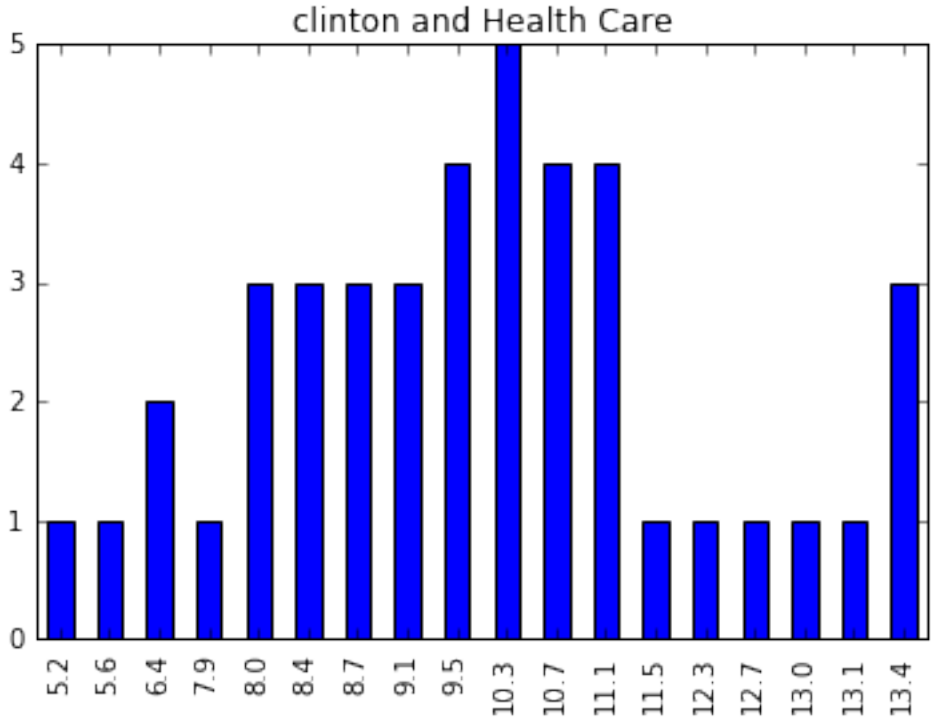
Abortion


clinton and Abortion
8
7
6
5
4
3
2
1
0
ap
cnn
fox
huffpo
latimes
nyt
politico
wsj

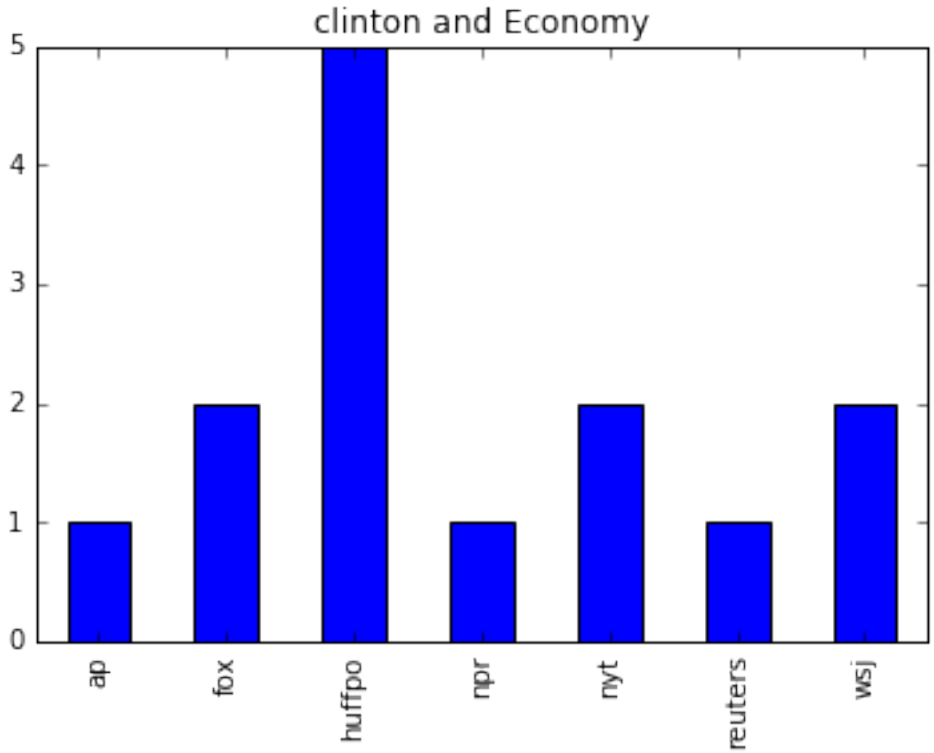

clinton and Abortion
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
7.2
7.6
8.0
8.7
9.1
9.5
9.9
10.3
10.7
11.1
11.5
11.9
12.3
13.0
13.8

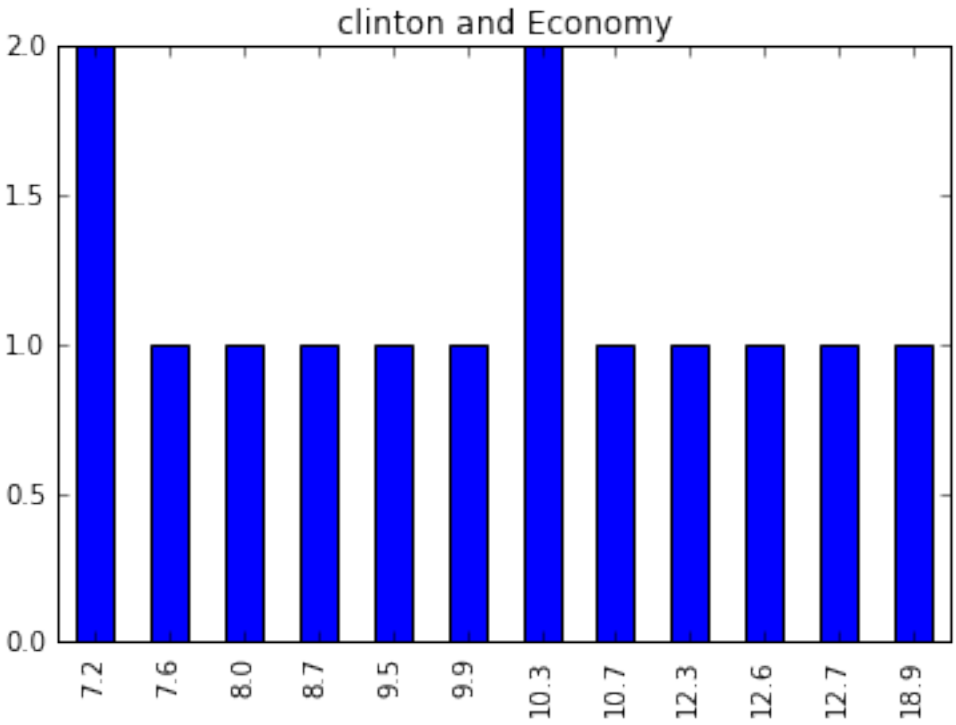
Health Care


clinton and Health Care
0
1
2
3
4
5
6
7
ap
cnn
fox
huffpo
latimes
mcclatchy
nyt
politico
reuters
wsj

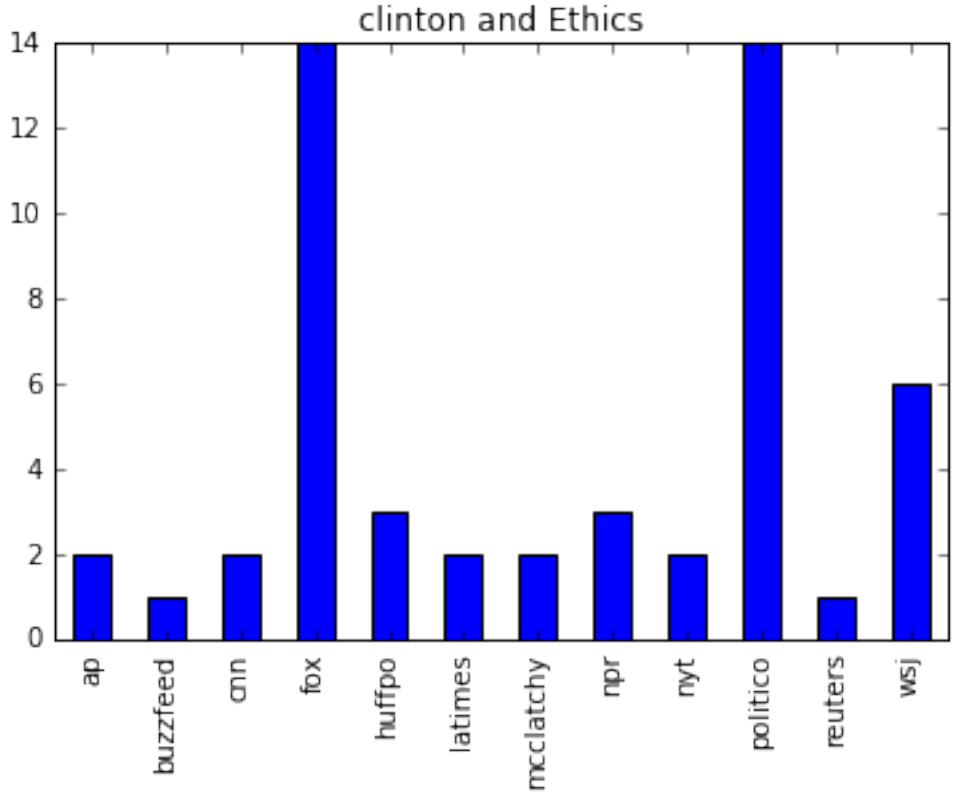

clinton and Health Care
0
1
2
3
4
5
5.2
5.6
6.4
7.9
8.0
8.4
8.7
9.1
9.5
10.3
10.7
11.1
11.5
12.3
12.7
13.0
13.1
13.4

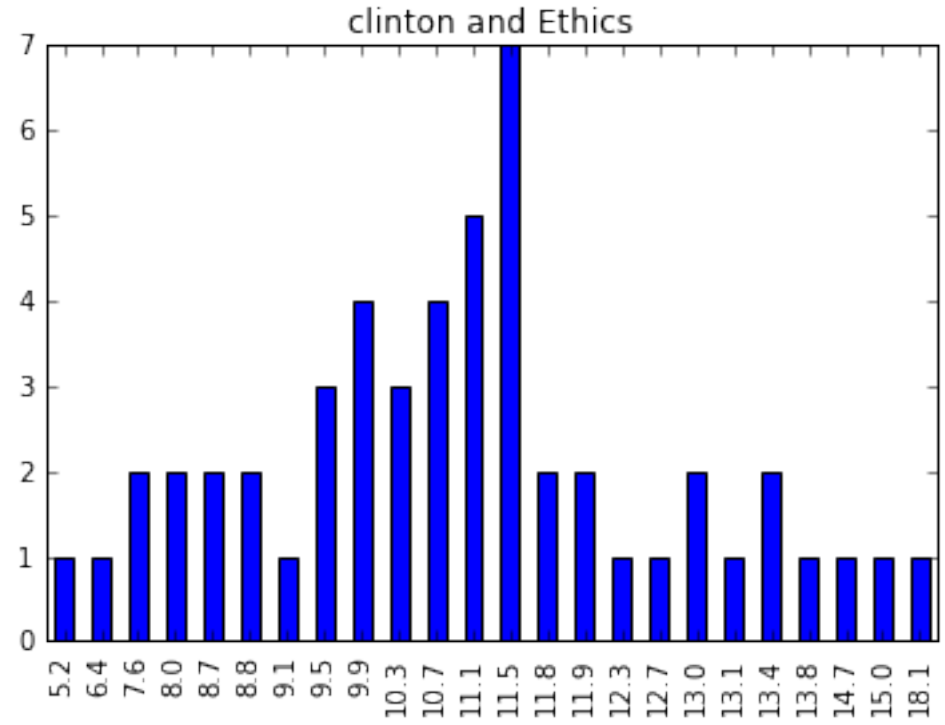
Economy


clinton and Economy
5
4
3
2
1
0
ap
fox
huffpo
npr
nyt
reuters
wsj

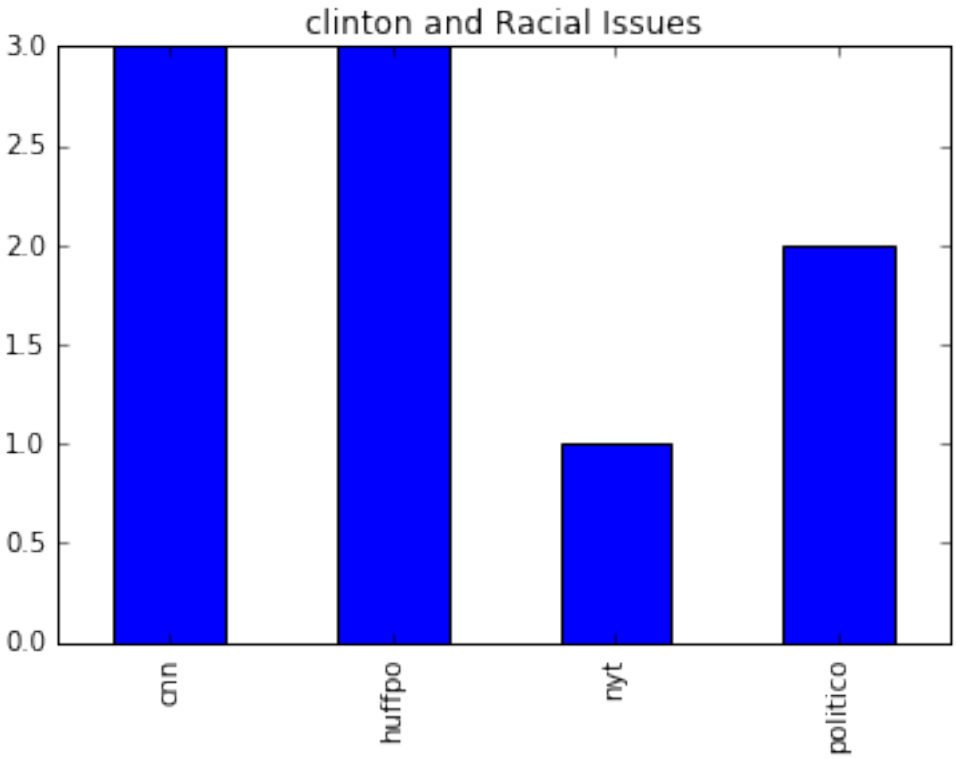

clinton and Economy
2.0
1.5
1.0
0.5
0.0
7.2
7.6
8.0
8.7
9.5
9.9
10.3
10.7
12.3
12.6
12.7
18.9

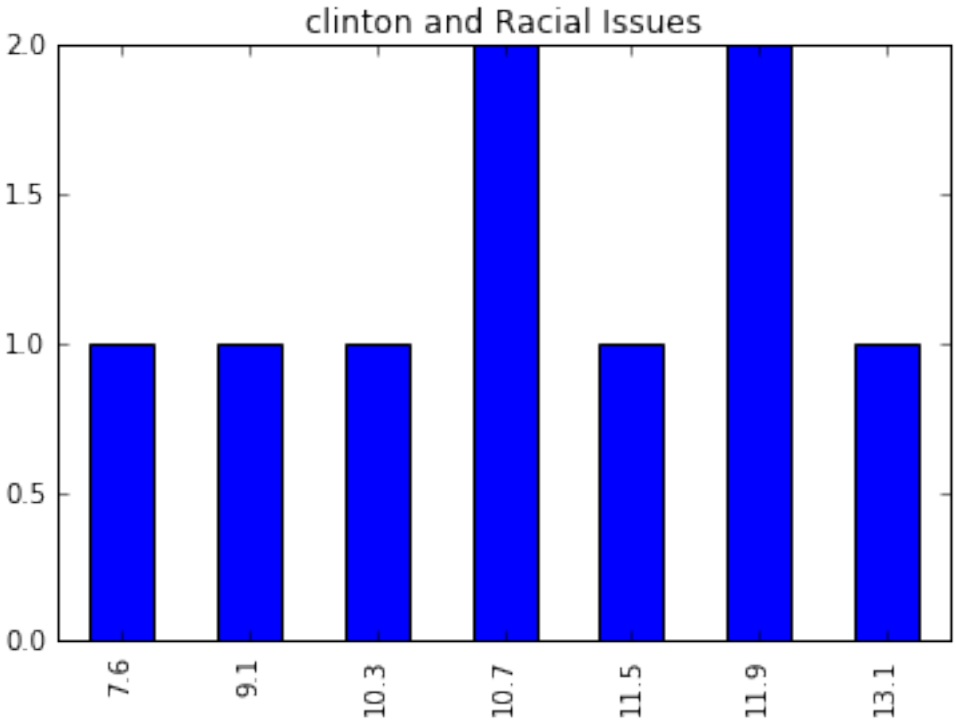
Ethics


clinton and Ethics
14
12
10
8
6
4
2
0
ap
buzzfeed
cnn
fox
huffpo
latimes
mcclatchy
npr
nyt
politico
reuters
wsj

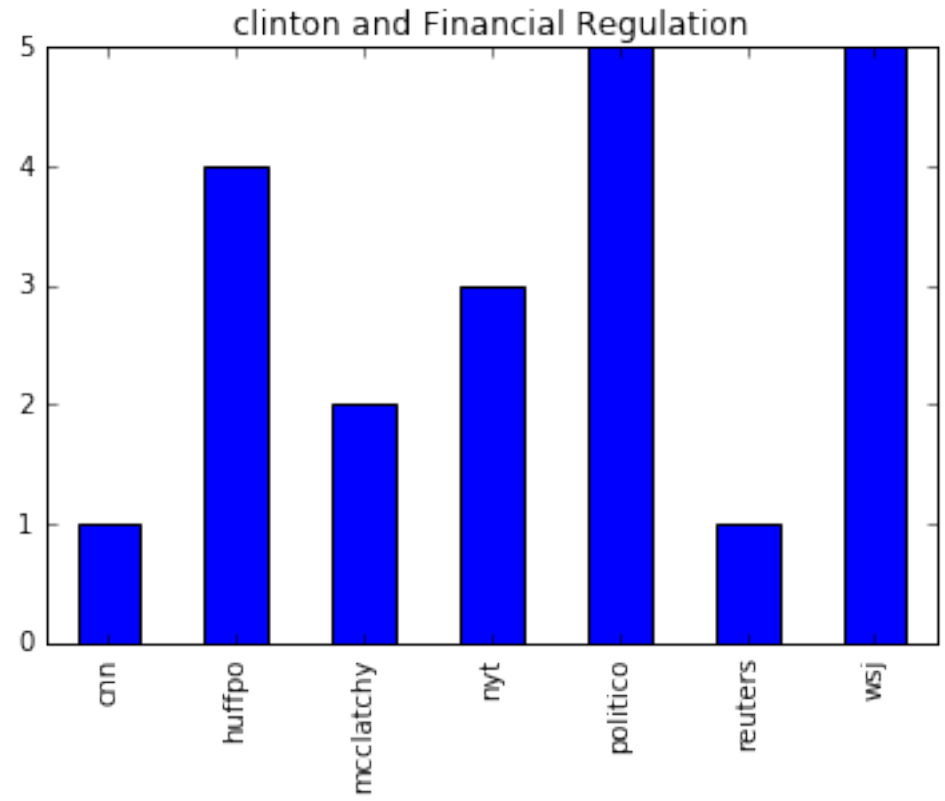


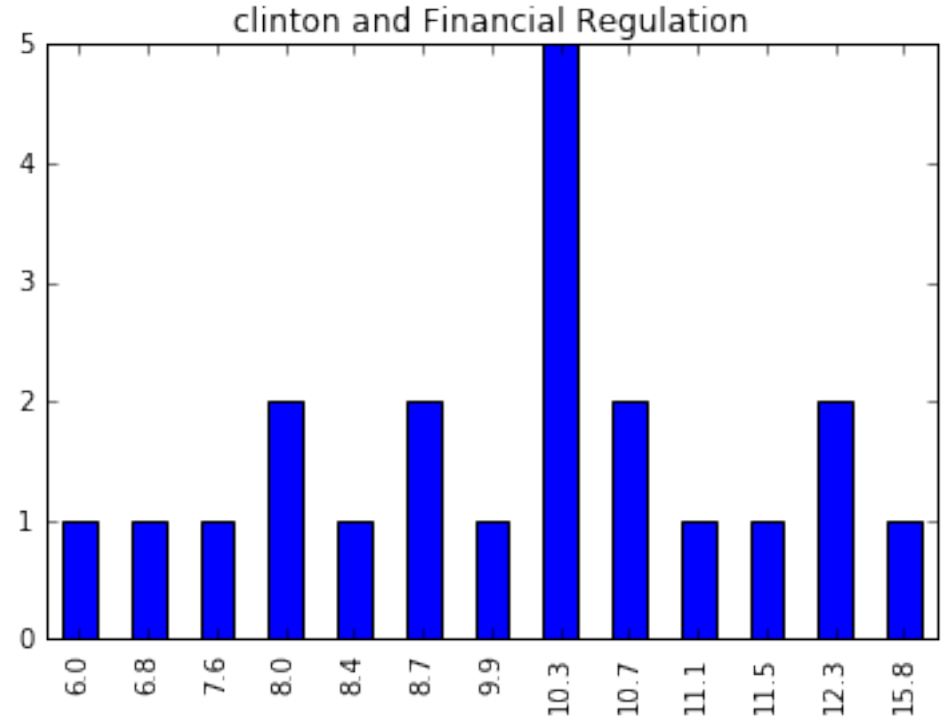
Racial Issues


clinton and Racial Issues
3.0
2.5
2.0
1.5
1.0
0.5
0.0
cnn
huffpo
nyt
politico

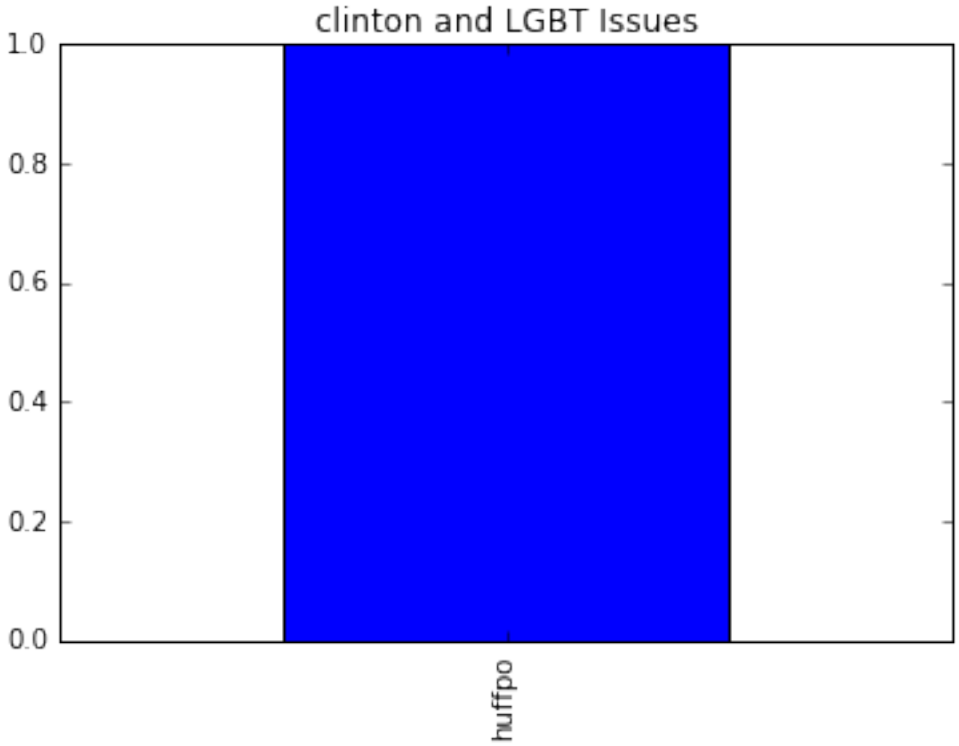

clinton and Racial Issues
2.0
1.5
1.0
0.5
0.0
7.6
9.1
10.3
10.7
11.5
11.9
13.1

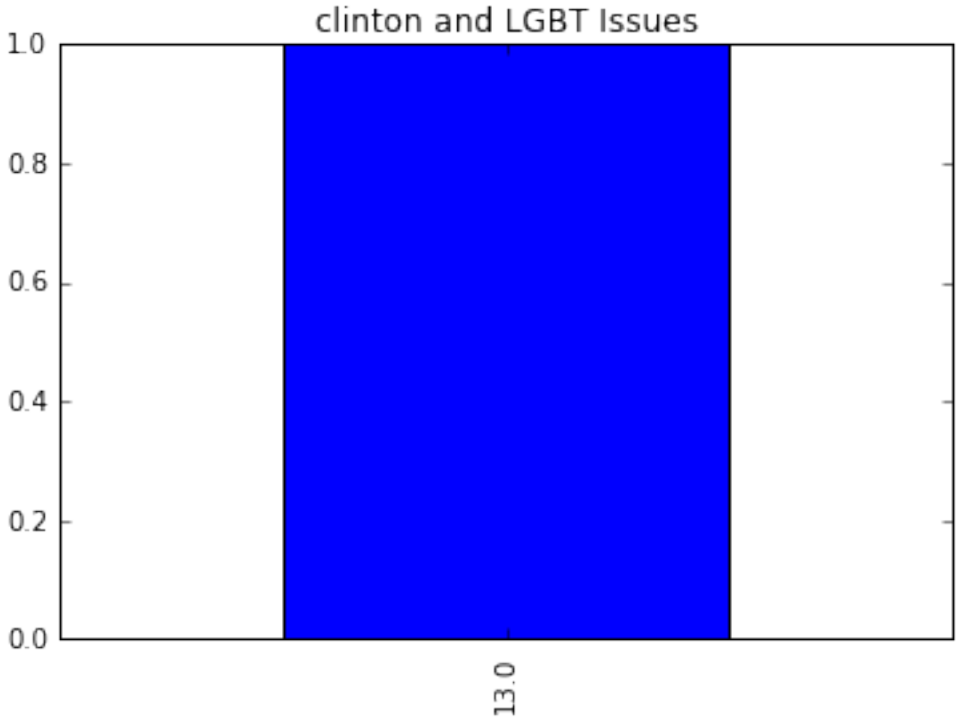
Financial Regulation





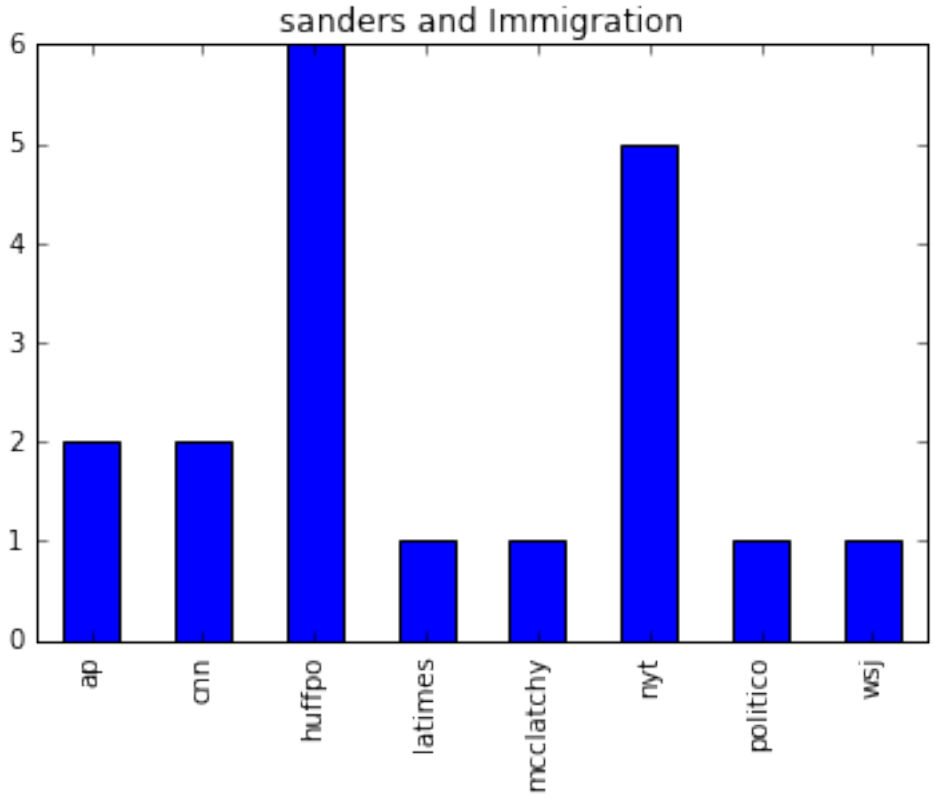
LGBT Issues

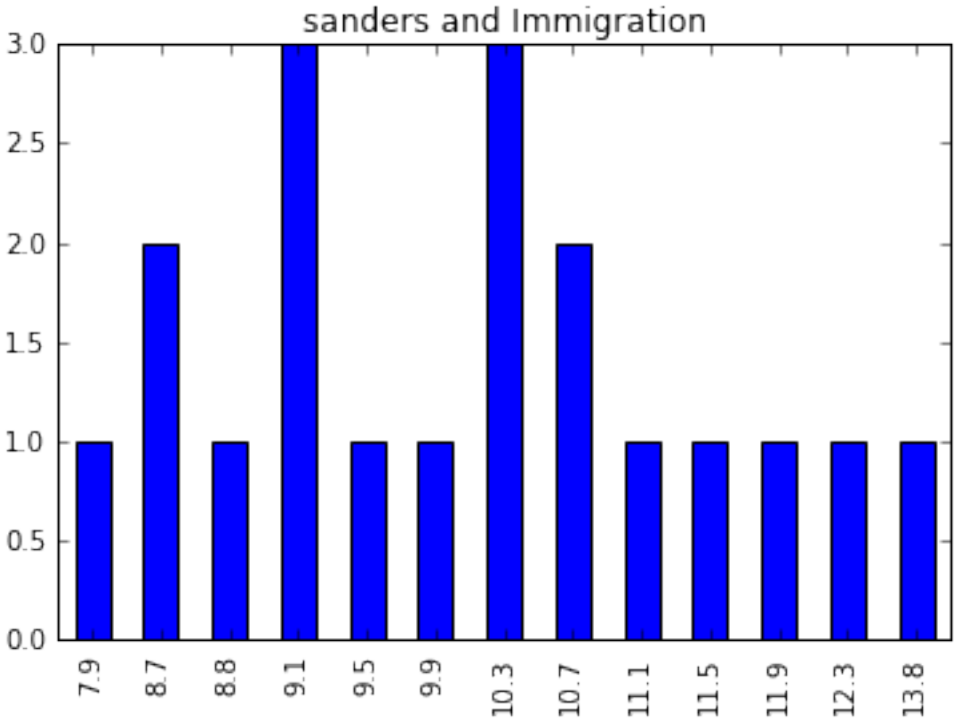

clinton and LGBT Issues
1.0
0.8
0.6
0.4
0.2
0.0
huffpo


clinton and LGBT Issues
1.0
0.8
0.6
0.4
0.2
0.0
13.0

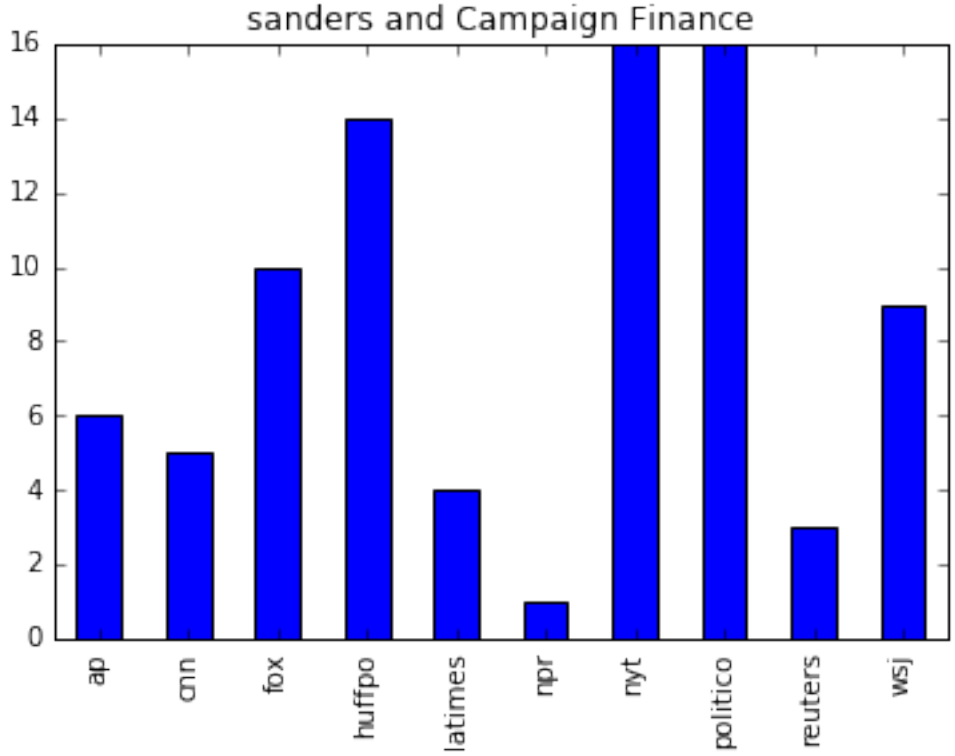
sanders

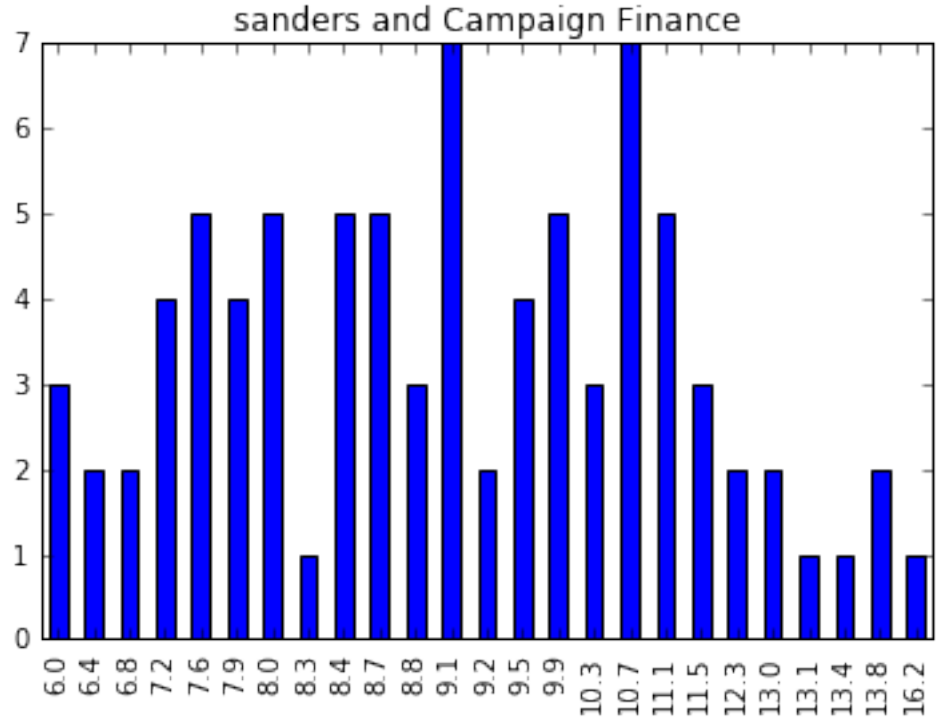
Immigration



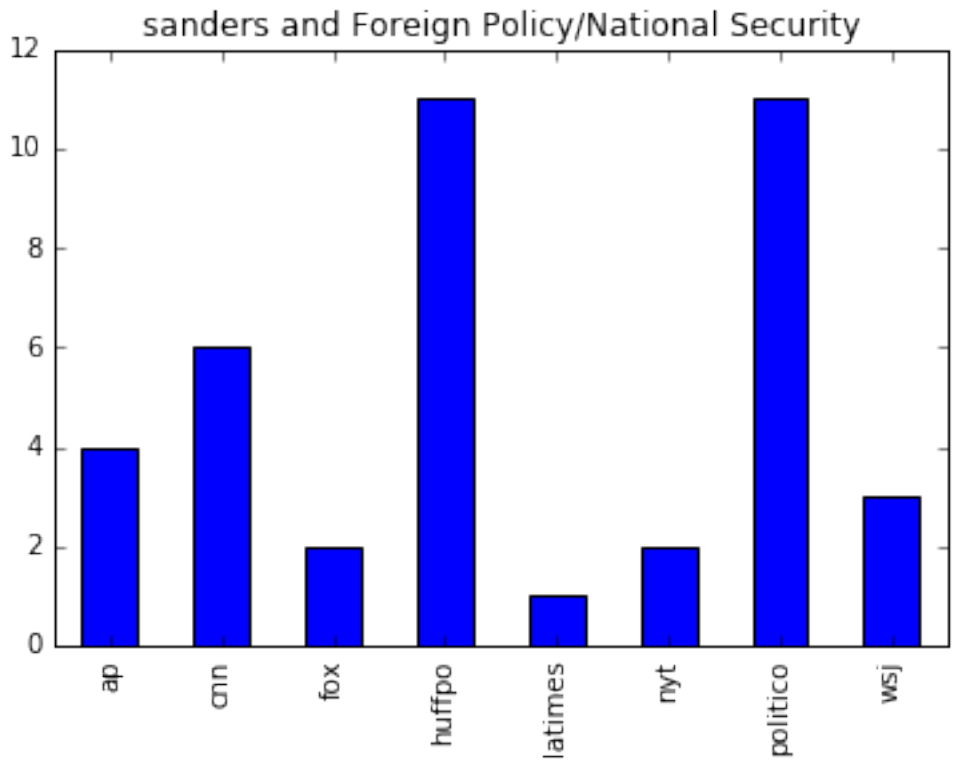


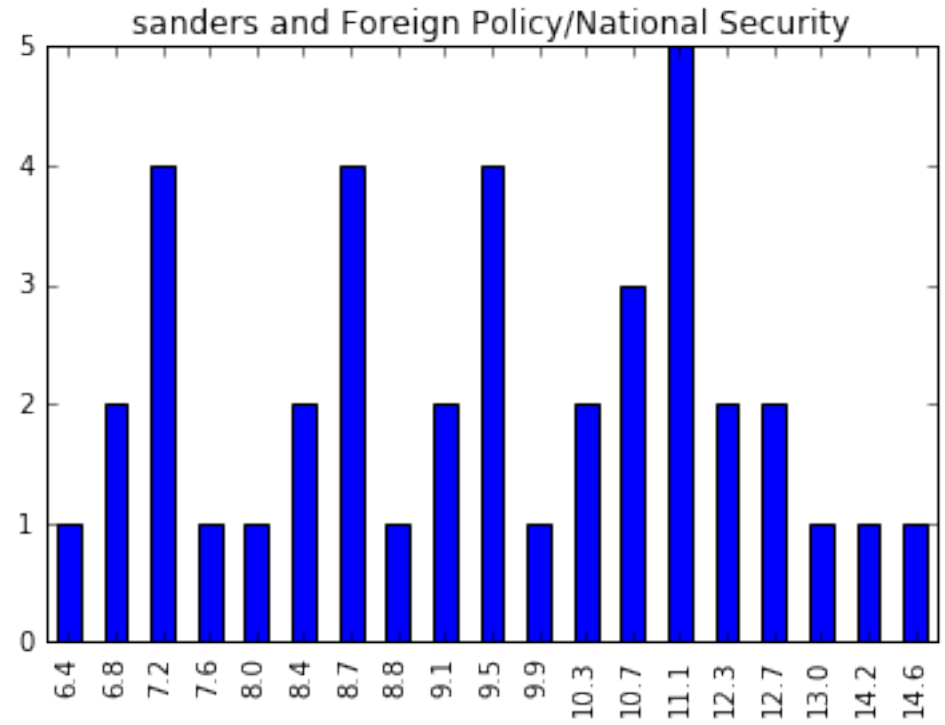
Campaign Finance



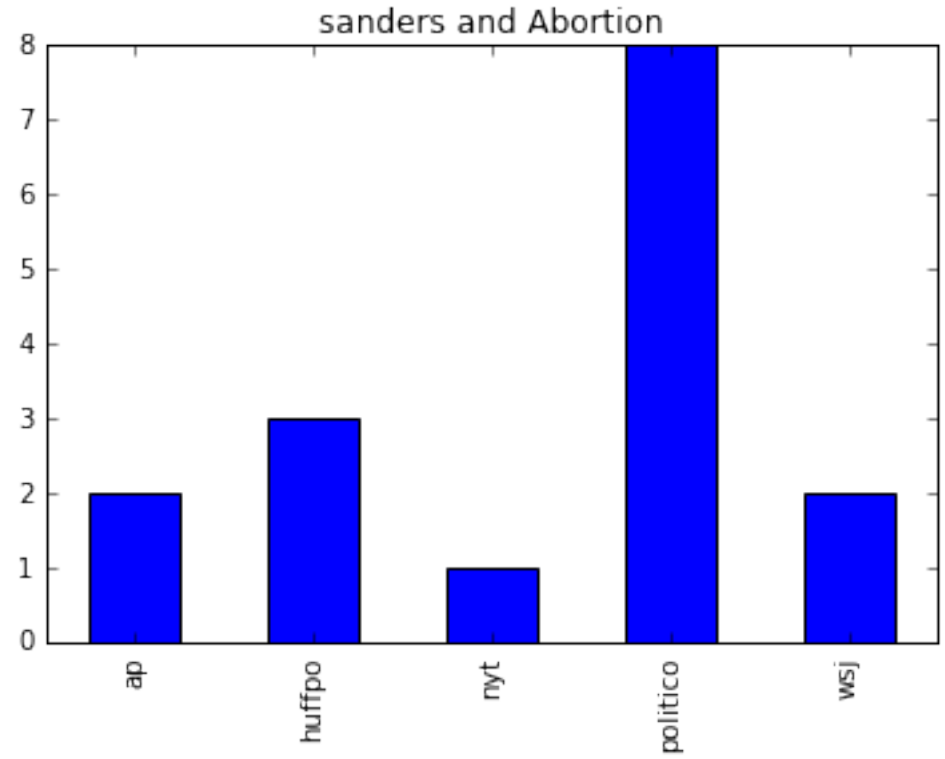


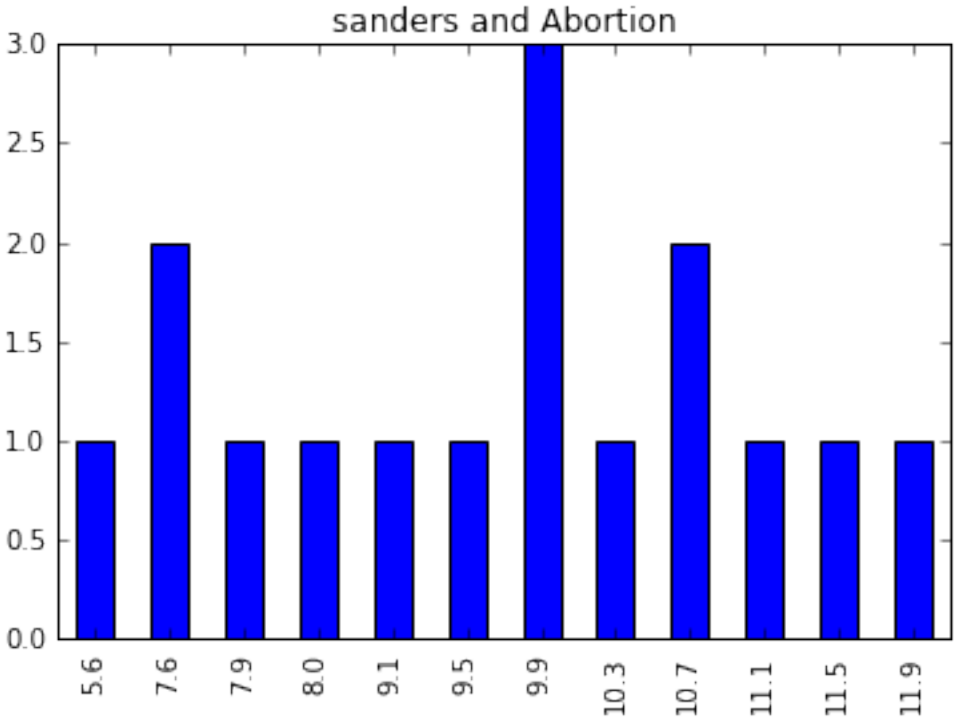
Foreign Policy/National Security


sanders and Foreign Policy/National Security
12
10
8
6
4
2
0
ap
cnn
fox
huffpo
latimes
nyt
politico
wsj

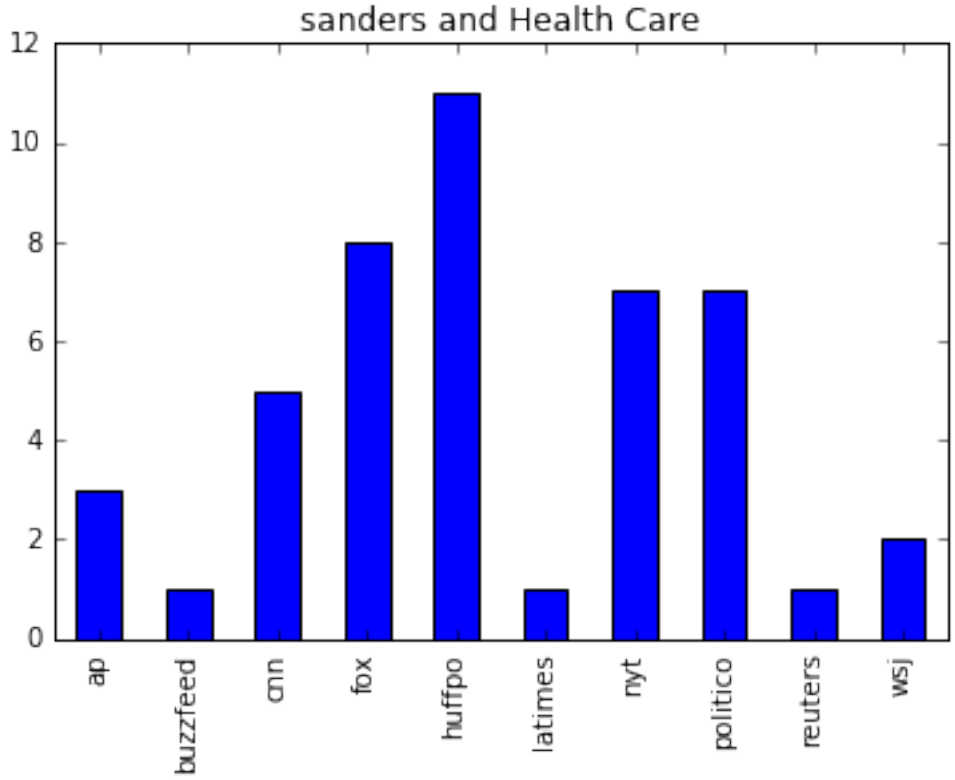

sanders and Foreign Policy/National Security
5
4
3
2
1
0
6.4
6.8
7.2
7.6
8.0
8.4
8.7
8.8
9.1
9.5
9.9
10.3
10.7
11.1
12.3
12.7
13.0
14.2
14.6

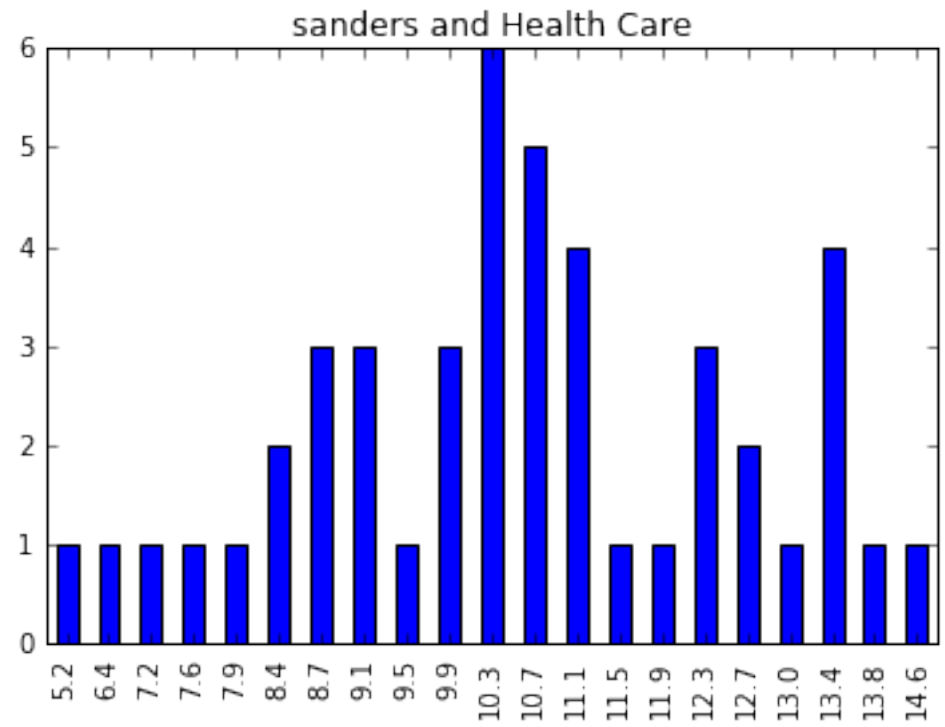
Abortion


sanders and Abortion
8
7
6
5
4
3
2
1
0
ap
huffpo
nyt
politico
wsj

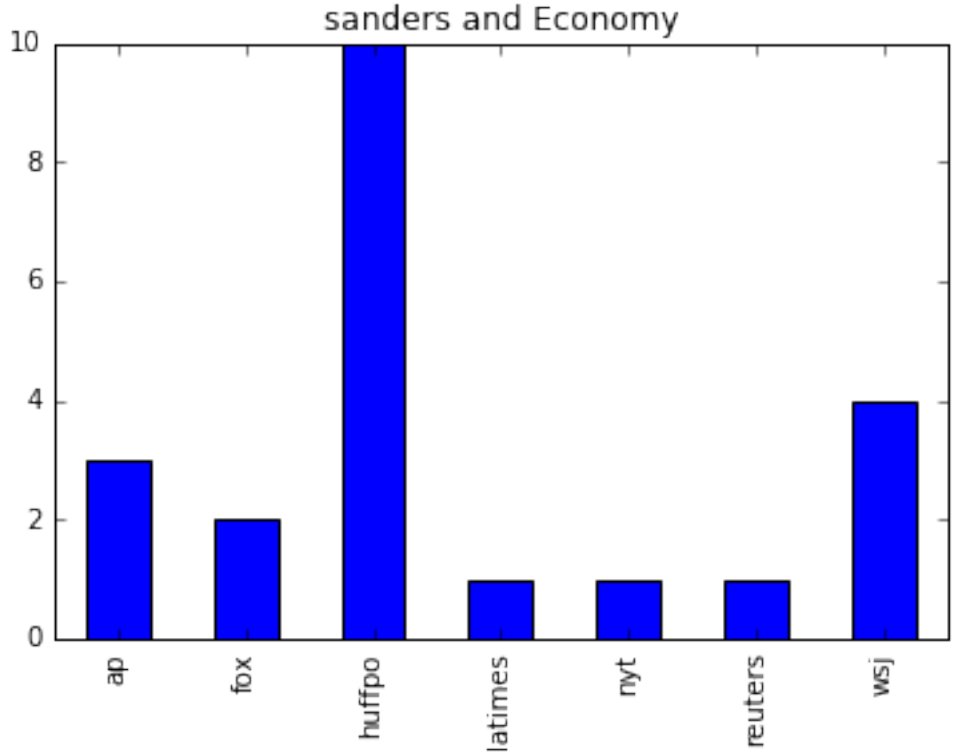

sanders and Abortion
3.0
2.5
2.0
1.5
1.0
0.5
0.0
5.6
7.6
7.9
8.0
9.1
9.5
9.9
10.3
10.7
11.1
11.5
11.9

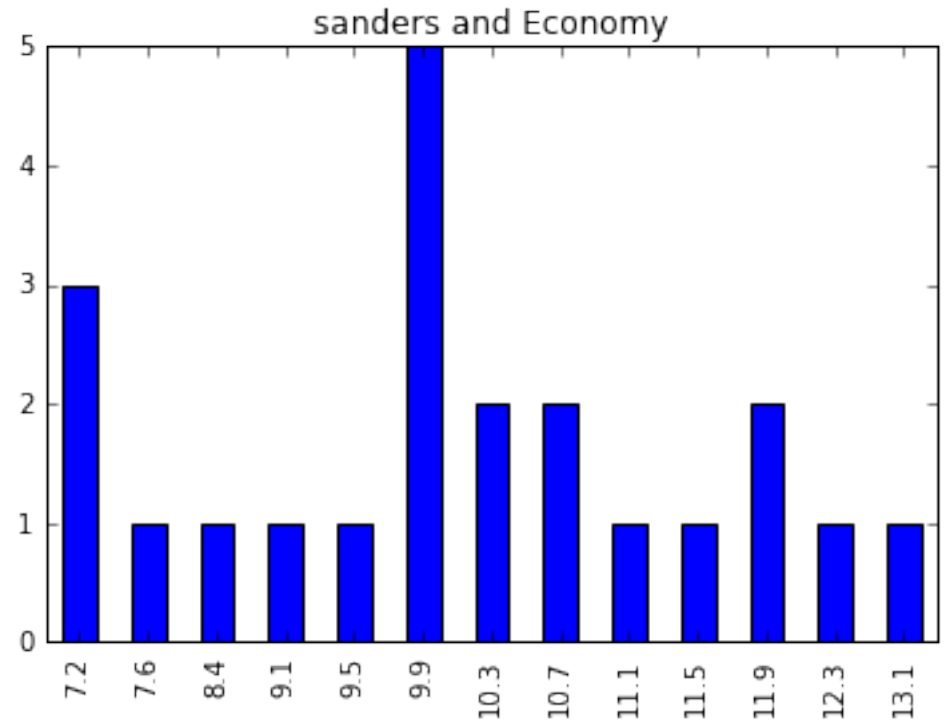
Health Care


sanders and Health Care
12
10
8
6
4
2
0
ap
buzzfeed
cnn
fox
huffpo
latimes
nyt
politico
reuters
wsj

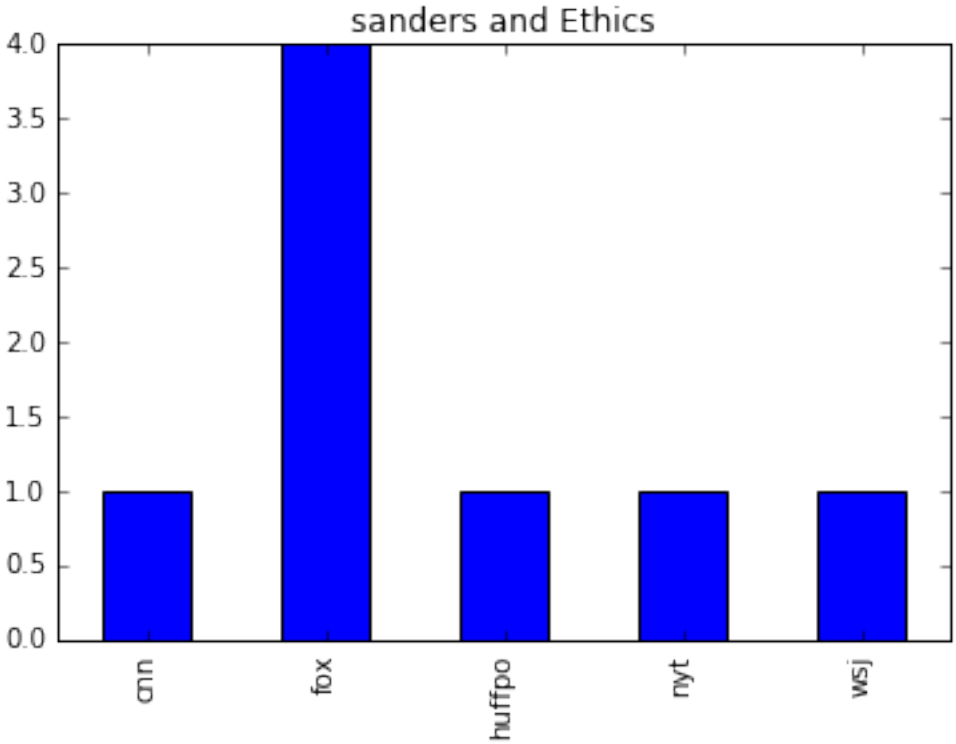

sanders and Health Care
0
1
2
3
4
5
6
5.2
6.4
7.2
7.6
7.9
8.4
8.7
9.1
9.5
9.9
10.3
10.7
11.1
11.5
11.9
12.3
12.7
13.0
13.4
13.8
14.6

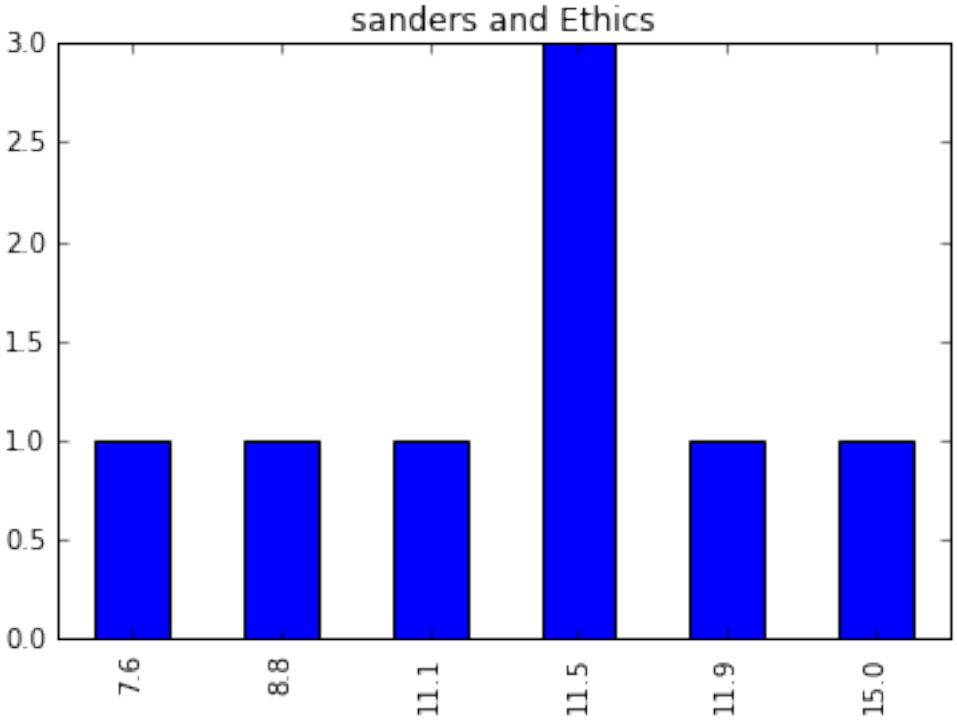
Economy


sanders and Economy
0
2
4
6
8
10
ap
fox
huffpo
latimes
nyt
reuters
wsj

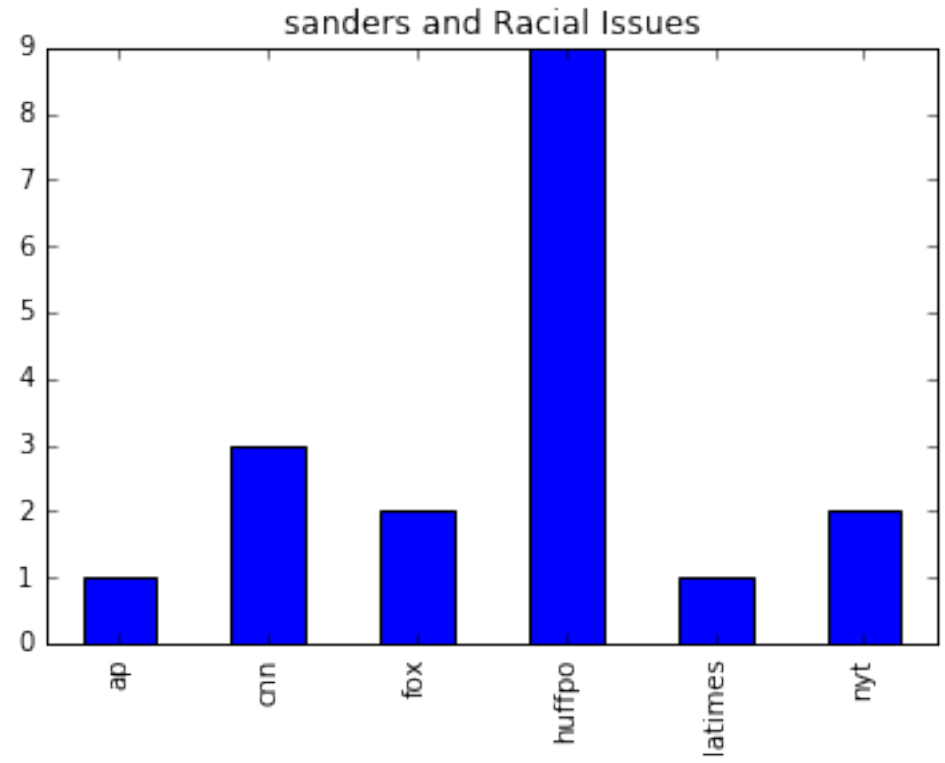

sanders and Economy
5
4
3
2
1
0
7.2
7.6
8.4
9.1
9.5
9.9
10.3
10.7
11.1
11.5
11.9
12.3
13.1

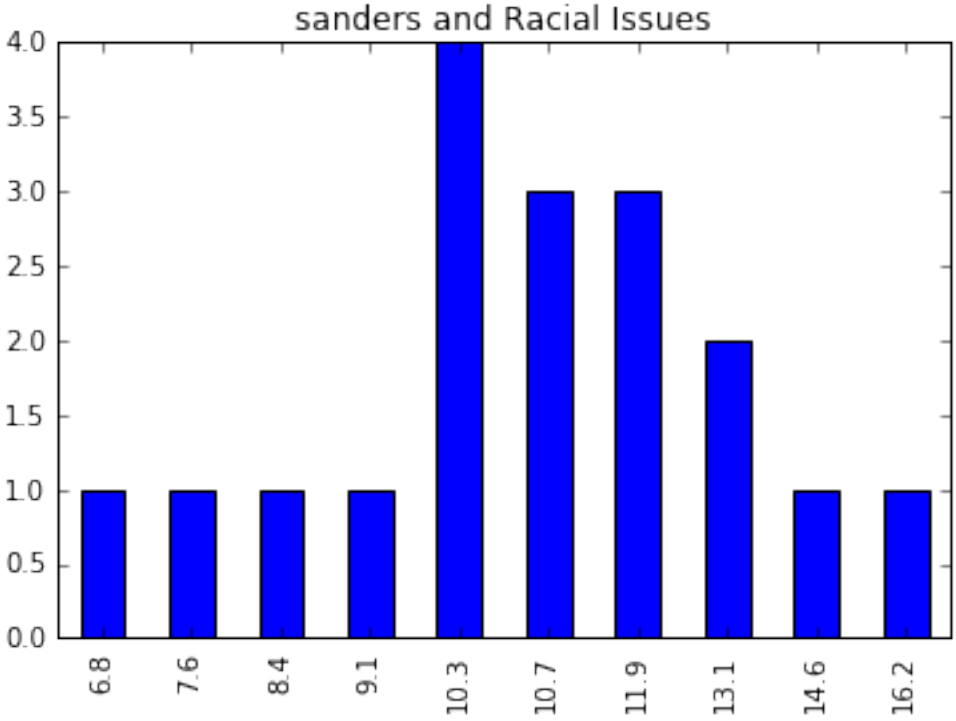
Ethics


sanders and Ethics
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
cnn
fox
huffpo
nyt
wsj

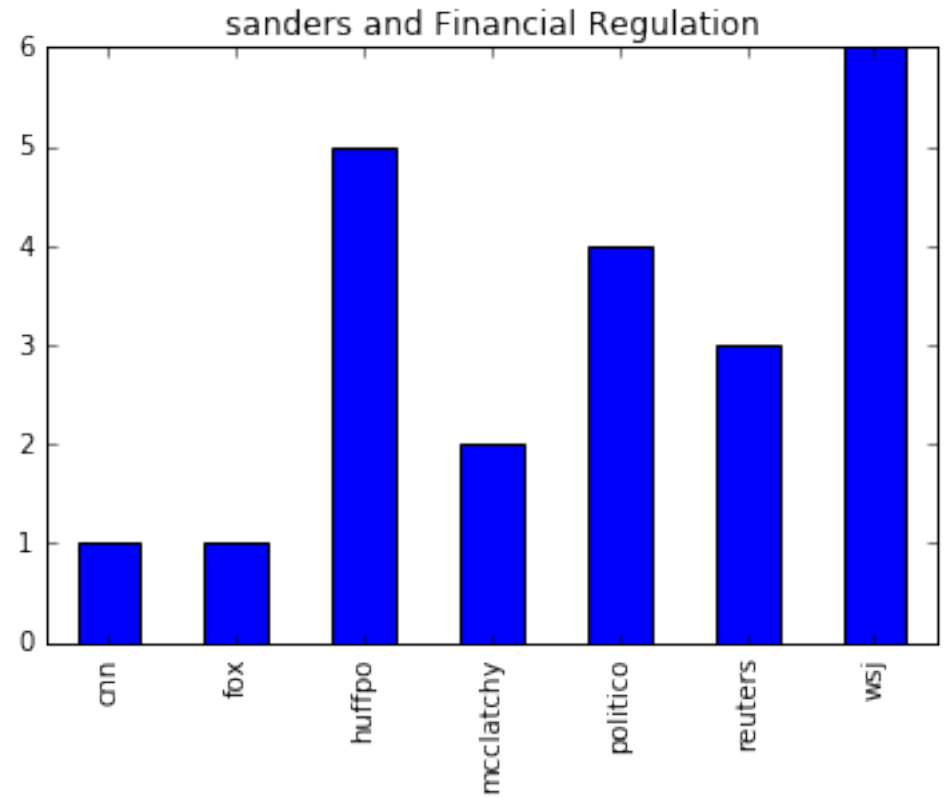


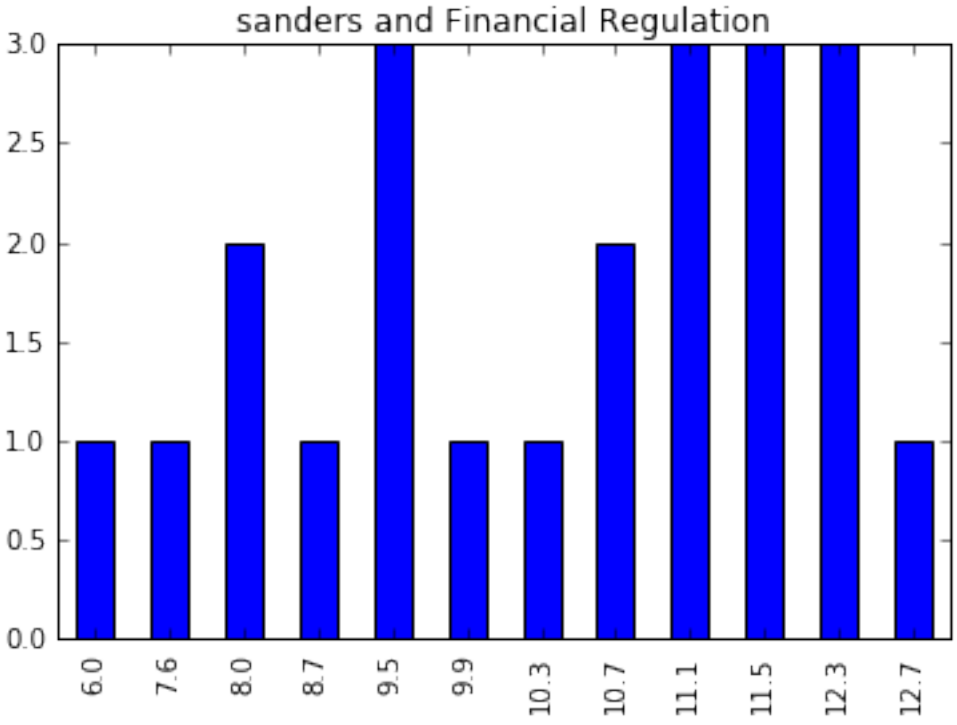
Racial Issues


sanders and Racial Issues
9
8
7
6
5
4
3
2
1
0
ap
cnn
fox
huffpo
latimes
nyt


sanders and Racial Issues
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
6.8
7.6
8.4
9.1
10.3
10.7
11.9
13.1
14.6
16.2

Financial Regulation





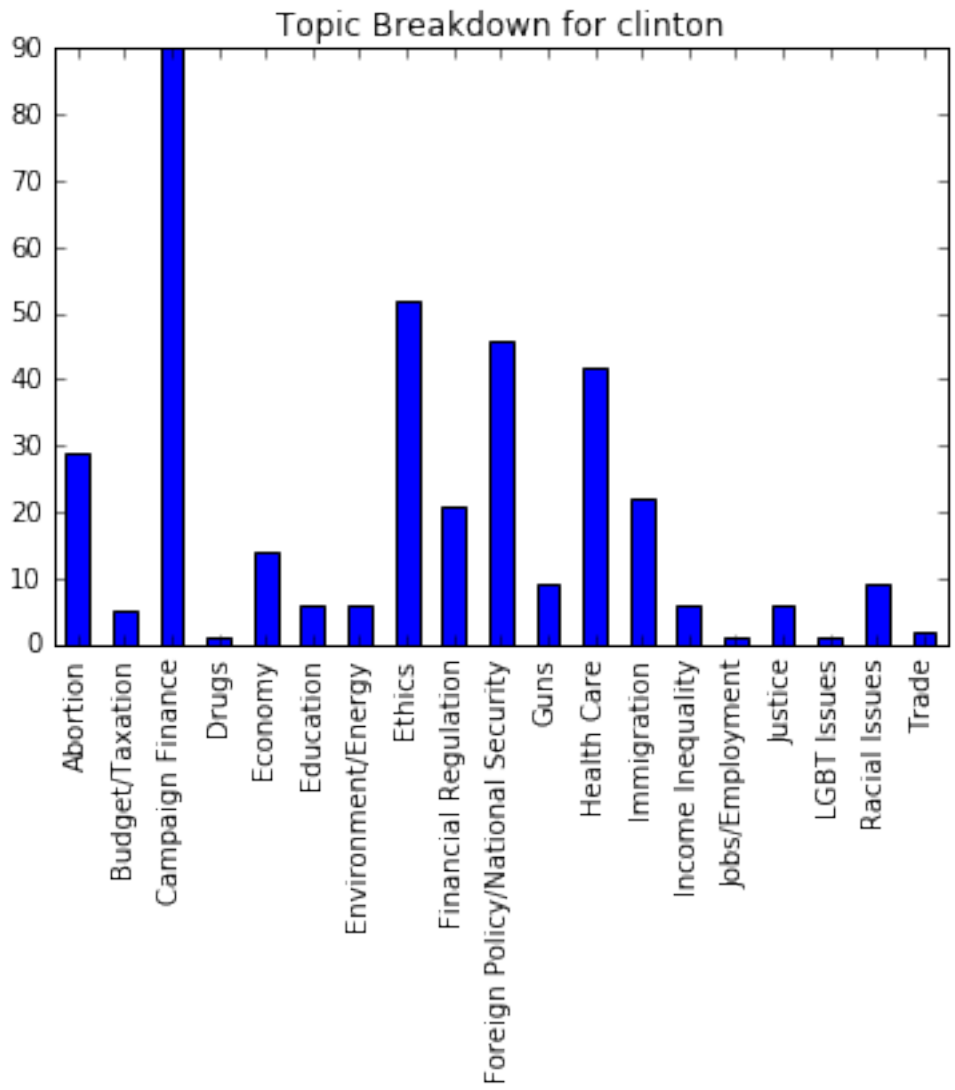
LGBT Issues

```
---------------------------------------------------------------------------

TypeError                                 Traceback (most recent call last)

<ipython-input-60-a33e10b3435e> in <module>()
      7         print t
      8         #print all_nop[(all_nop['candidate'] == c) & (all_nop['top_topic'] == t)]['org'].valu
----> 9         all_nop[(all_nop['candidate'] == c) & (all_nop['top_topic'] == t)]['org'].value_count
     10         matplotlib.pyplot.show()
     11         all_nop[(all_nop['candidate'] == c) & (all_nop['top_topic'] == t)]['flesch'].value_co


/usr/local/lib/python2.7/site-packages/pandas/tools/plotting.pyc in __call__(self, kind, ax, figs
   3495                             colormap=colormap, table=table, yerr=yerr,
   3496                             xerr=xerr, label=label, secondary_y=secondary_y,
-> 3497                             **kwds)
   3498     __call__.__doc__ = plot_series.__doc__
   3499


/usr/local/lib/python2.7/site-packages/pandas/tools/plotting.pyc in plot_series(data, kind, ax,
   2585                  yerr=yerr, xerr=xerr,
   2586                  label=label, secondary_y=secondary_y,
-> 2587                  **kwds)
   2588
   2589


/usr/local/lib/python2.7/site-packages/pandas/tools/plotting.pyc in _plot(data, x, y, subplots,
   2382         plot_obj = klass(data, subplots=subplots, ax=ax, kind=kind, **kwds)
   2383
-> 2384     plot_obj.generate()
   2385     plot_obj.draw()
   2386     return plot_obj.result


/usr/local/lib/python2.7/site-packages/pandas/tools/plotting.pyc in generate(self)
    983     def generate(self):
    984         self._args_adjust()
--> 985         self._compute_plot_data()
    986         self._setup_subplots()
    987         self._make_plot()


/usr/local/lib/python2.7/site-packages/pandas/tools/plotting.pyc in _compute_plot_data(self)
   1092         if is_empty:
   1093             raise TypeError('Empty {0!r}: no numeric data to '
-> 1094                             'plot'.format(numeric_data.__class__.__name__))
   1095
```

```
1096         self.data = numeric_data


TypeError: Empty 'DataFrame': no numeric data to plot
```

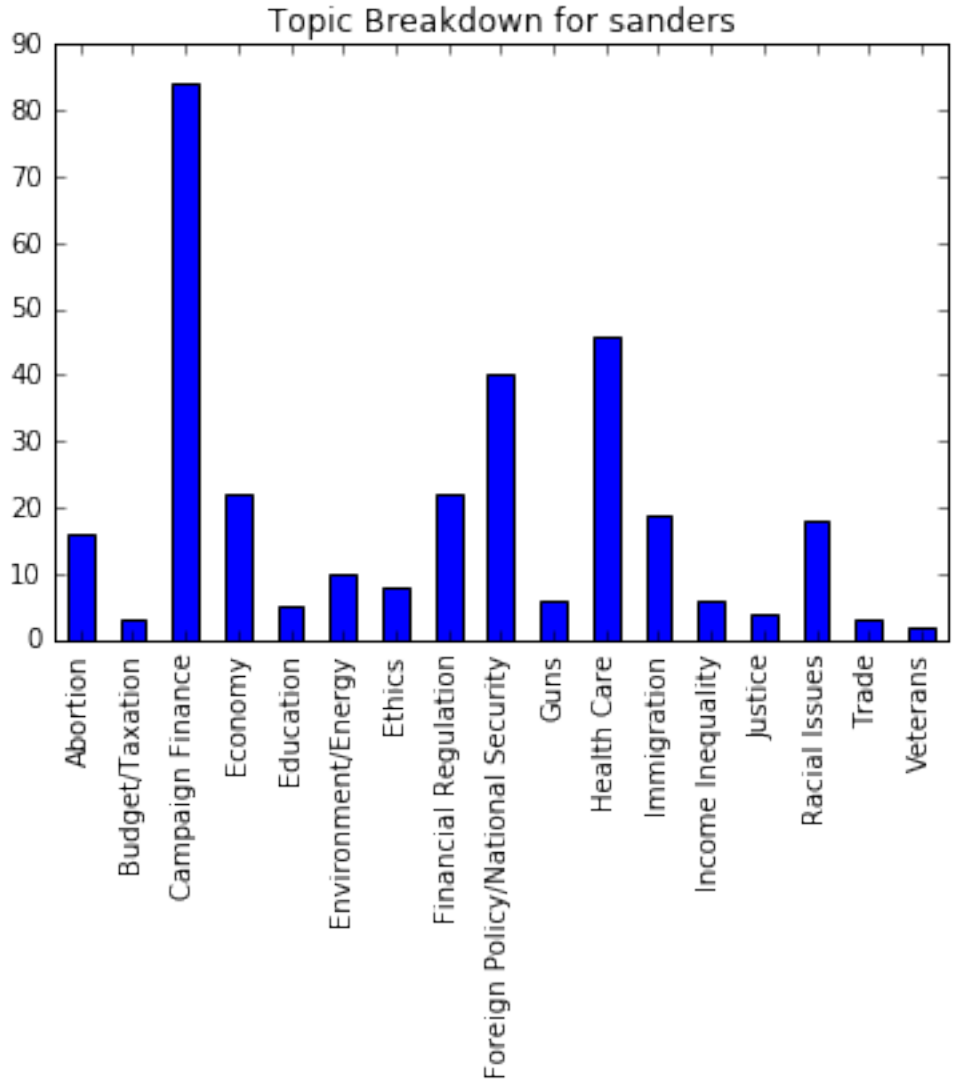
0.5 Topic Breakdown Per Candidate

```
In [81]: for c in CANDIDATES:
             #all_nop[all_nop['candidate'] == c]['top_topic'].value_counts().plot(kind="bar", title="To
             #matplotlib.pyplot.show()
             print
             all_nop[(all_nop['candidate'] == c)]['top_topic'].value_counts()\
                                     .sort_index().plot(kind="bar", title="Topic Breakdown for " + c)
             matplotlib.pyplot.show()
             print all_nop[(all_nop['candidate'] == c)]['top_topic'].value_counts()\
                                     .sort_index()
             print
```



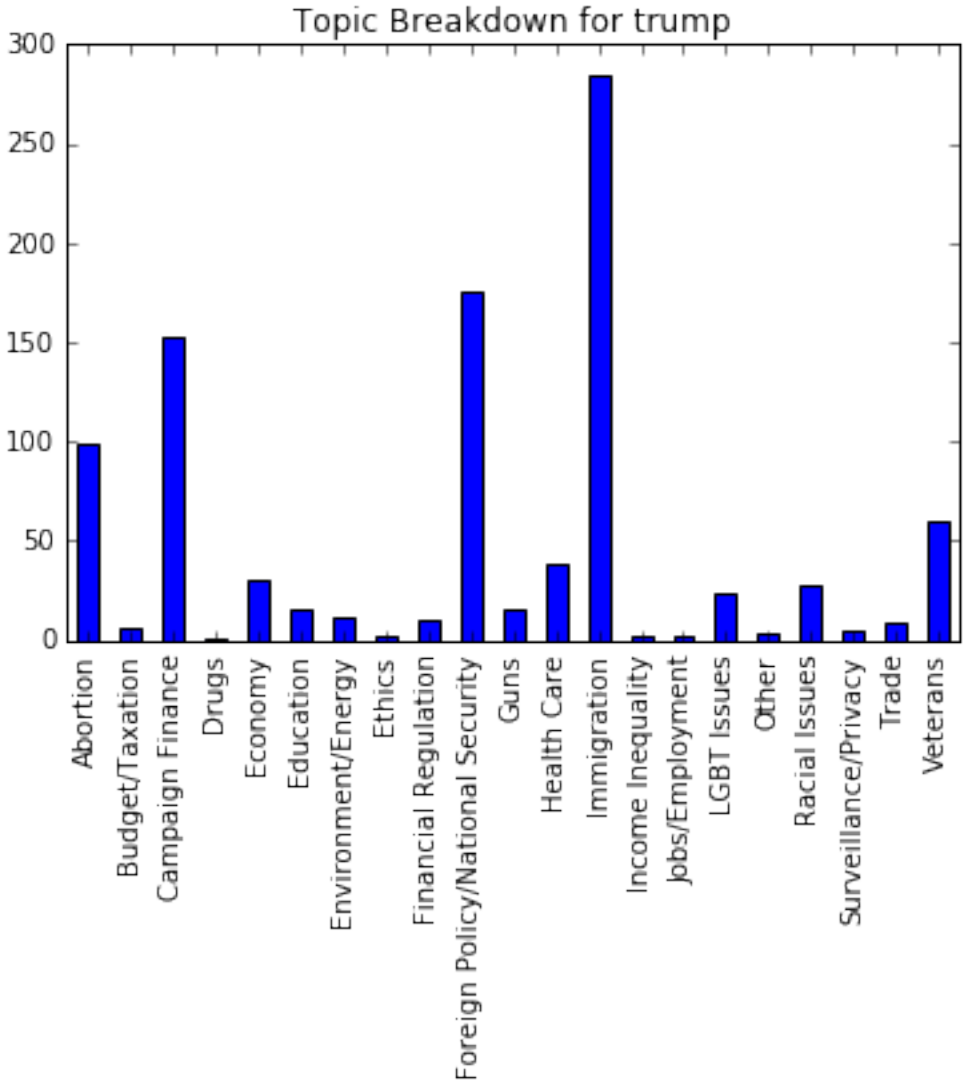
Abortion	29
Budget/Taxation	5
Campaign Finance	90
Drugs	1
Economy	14
Education	6
Environment/Energy	6
Ethics	52
Financial Regulation	21
Foreign Policy/National Security	46
Guns	9
Health Care	42
Immigration	22
Income Inequality	6
Jobs/Employment	1
Justice	6

```
LGBT Issues                          1
Racial Issues                        9
Trade                                2
Name: top_topic, dtype: int64
```



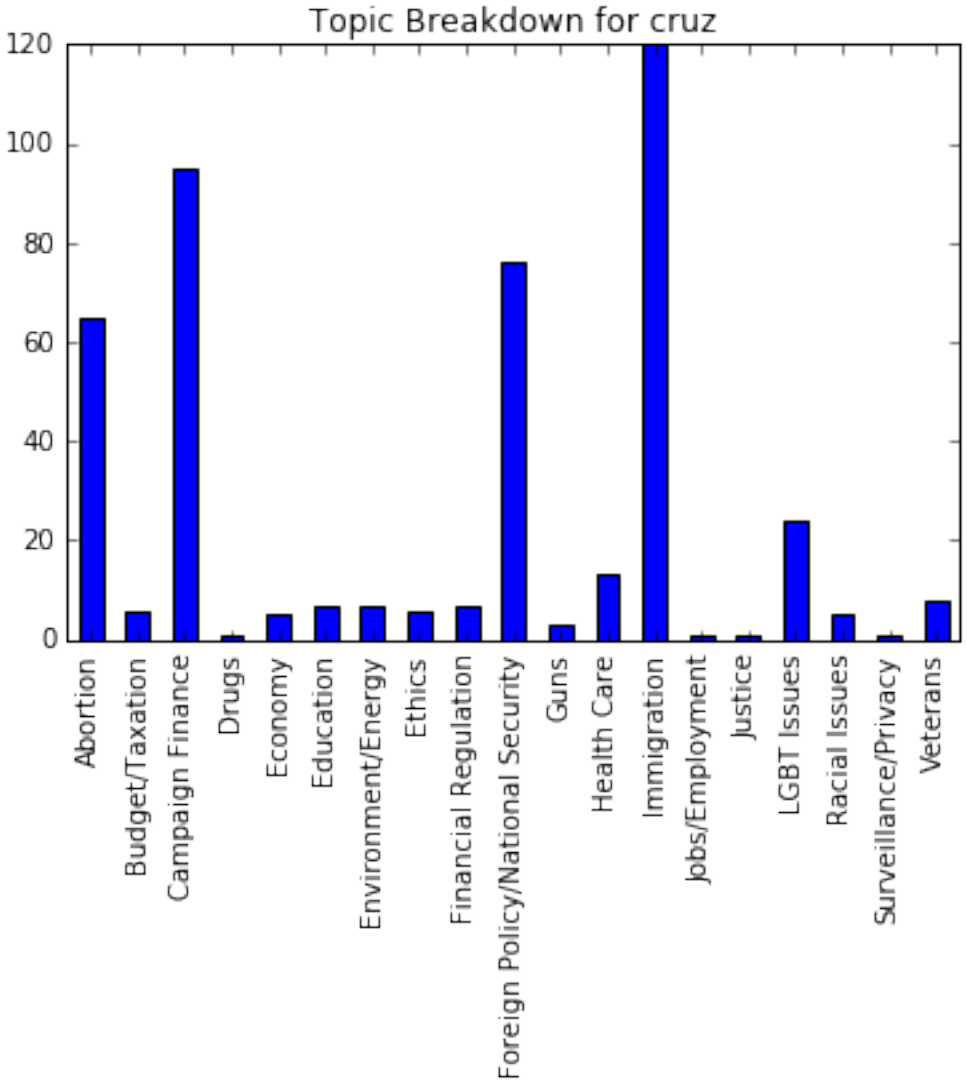

```
Abortion                            16
Budget/Taxation                      3
Campaign Finance                    84
Economy                             22
Education                            5
Environment/Energy                  10
Ethics                               8
Financial Regulation                22
Foreign Policy/National Security    40
Guns                                 6
Health Care                         46
```

```
Immigration                 19
Income Inequality            6
Justice                      4
Racial Issues               18
Trade                        3
Veterans                     2
Name: top_topic, dtype: int64
```




```
Abortion                    99
Budget/Taxation              6
Campaign Finance           153
Drugs                        1
Economy                     30
Education                   16
Environment/Energy          12
Ethics                       2
```

```
Financial Regulation                10
Foreign Policy/National Security   175
Guns                                16
Health Care                         38
Immigration                        284
Income Inequality                   2
Jobs/Employment                     2
LGBT Issues                        24
Other                               3
Racial Issues                      28
Surveillance/Privacy                5
Trade                               9
Veterans                           60
Name: top_topic, dtype: int64
```




```
Abortion                65
Budget/Taxation          6
```

```
Campaign Finance                     95
Drugs                                 1
Economy                               5
Education                             7
Environment/Energy                    7
Ethics                                6
Financial Regulation                  7
Foreign Policy/National Security     76
Guns                                  3
Health Care                          13
Immigration                         120
Jobs/Employment                       1
Justice                               1
LGBT Issues                          24
Racial Issues                         5
Surveillance/Privacy                  1
Veterans                              8
Name: top_topic, dtype: int64
```

In [ ]:

In [ ]: