In [1]:
```
import glob
import pandas
import json
import time
print('pandas',pandas.__version__)
import matplotlib.pyplot as plt
import numpy
print('numpy',numpy.__version__)
```

```
pandas 0.23.4
numpy 1.13.3
```

In [6]:
```
# https://stackoverflow.com/questions/17977540/pandas-looking-up-the-list-of-sheets-in-an-excel-file
xl = pandas.ExcelFile("voting_data/Burlington/Burlington/Burlington 2006 Kleppner Burlington IRV election analysis.xls")
```

In [7]:
```
xl.sheet_names  # see all sheet names
```

Out[7]:
```
['Description',
 '1. Turnout',
 '2. Runoff turnout',
 '3. Ballot images',
 '4.  Instant runoff tally',
 '5.  Reconciliation VTS&CPPro']
```

In [8]:
```
xl.sheet_names[3]  # see all sheet names
```

Out[8]:
```
'3. Ballot images'
```

In [11]:
```
df = pandas.read_excel("voting_data/Burlington/Burlington/Burlington 2006 Kleppner Burlington IRV election analysis.xls",
                       sheet_name=xl.sheet_names[3],
                       skiprows=15)
```

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

Out[12]:

	Ward	Mem card	number	1st	2nd	3rd	4th	5th	Rankings	1st.1	2nd.1	Top rank	1st rnd	2nd rnd
0	1.0	11.0	1.0	C04	C02	C01	NaN	NaN	3.0	C04	C04	1.0	valid	eff
1	1.0	11.0	2.0	C03	C04	C05	NaN	NaN	3.0	C03	C03	1.0	valid	eff
2	1.0	11.0	3.0	C04	C03	NaN	NaN	NaN	2.0	C04	C04	1.0	valid	eff
3	1.0	11.0	4.0	C03	C04	NaN	NaN	NaN	2.0	C03	C03	1.0	valid	eff
4	1.0	11.0	5.0	C03	C04	C01	NaN	NaN	3.0	C03	C03	1.0	valid	eff

In [14]:
```
set(list(df['1st.1'].values))
```

Out[14]:
```
{'C01', 'C02', 'C03', 'C04', 'C05', 'C06', nan, 'ov', 'uv'}
```

In [16]:
```
df[df['1st.1']=='ov']
```

Out[16]:

	Ward	Mem card	number	1st	2nd	3rd	4th	5th	Rankings	1st.1	2nd.1	r
9779	2.0	21.0	13.0	C03=C06	C05	C01	C02	NaN	4.0	ov	ov	1
9780	3.0	31.0	172.0	C01=C04	C02	C05	C03	NaN	4.0	ov	ov	1
9781	4.0	41.0	233.0	C03=C04	NaN	NaN	NaN	NaN	1.0	ov	ov	1
9782	4.0	41.0	242.0	C03=C04	NaN	NaN	NaN	NaN	1.0	ov	ov	1
9783	4.0	41.0	379.0	C04=C06	C02	C05	NaN	NaN	3.0	ov	ov	1
9784	3.0	31.0	397.0	C03=C04	NaN	NaN	NaN	NaN	1.0	ov	ov	1
9785	7.0	71.0	640.0	C01=C02	C03	NaN	NaN	NaN	2.0	ov	ov	1
9786	7.0	71.0	641.0	C01=C06	C04	NaN	NaN	NaN	2.0	ov	ov	1
9787	7.0	71.0	646.0	C01=C05	C02=C04	C03	C06	NaN	4.0	ov	ov	1
9788	4.0	42.0	832.0	C01=C04	NaN	NaN	NaN	NaN	1.0	ov	ov	1

In [18]:
```
df[pandas.isna(df['1st.1'])]
```

Out[18]:

	Ward	Mem card	number	1st	2nd	3rd	4th	5th	Rankings	1st.1	2nd.1	Top rank	1st rnd
9778	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
9789	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

In [24]:
```
reslts={}
for entry in set(list(df['1st.1'].values)):
    reslts[entry]=len(df[df['1st.1']==entry])
```

In [25]:
```
for k,v in reslts.items():
    print(k,':',v)
```

```
nan : 0
C04 : 3106
C01 : 119
ov : 10
uv : 77
C02 : 2609
C03 : 3809
C05 : 57
C06 : 78
```

In [31]:
```
list_of_votes=[]
for k,v in reslts.items():
    if str(k).startswith('C'):
        print(k)
        list_of_votes.append(v)
list_of_votes
```

```
C04
C01
C02
C03
C05
C06
```

Out[31]: `[3106, 119, 2609, 3809, 57, 78]`

In [37]:
```
percentages = [x/sum(list_of_votes) for x in list_of_votes]
```

In [38]:
```
def which_cat(percentages):
    if max(percentages)>0.5:
        print("Leading candidate in the first round has greater than 50% first choice votes")
    elif max(percentages)<=0.5 and max(percentages)>=0.45:
        print("Leading candidate in the first round has between 45-50% first choice votes")
    elif max(percentages)<0.45:
        print("Leading candidate in the first round has less than 45% of first choice votes")
    else:
        raise Exception("invalid outcome")
    return
```

In [39]:
```
which_cat(percentages)
```

```
Leading candidate in the first round has less than 45% of first choice votes
```