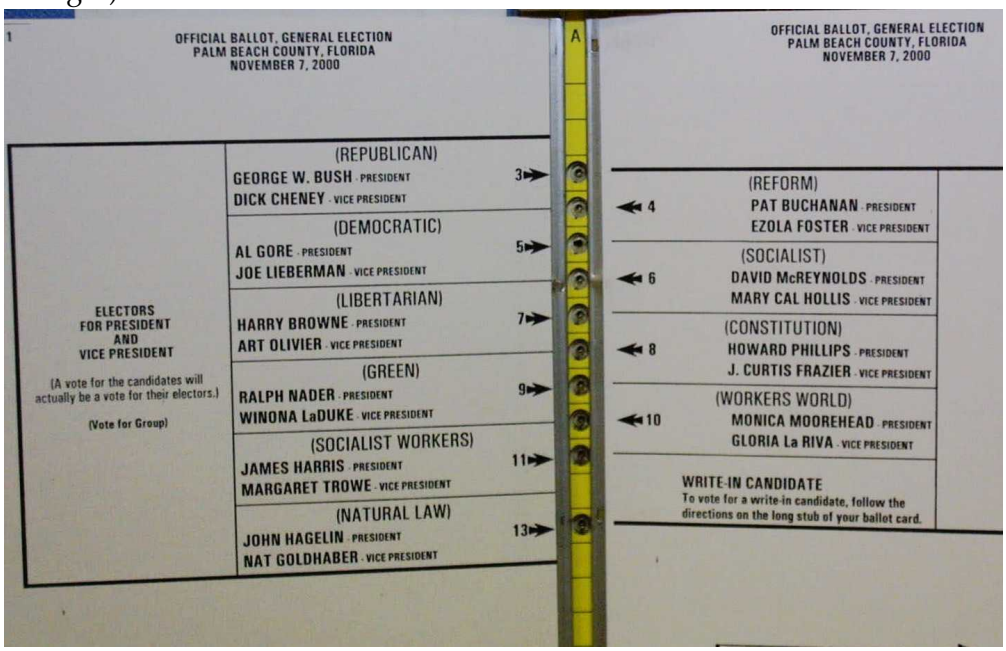
3A. The 2000 U.S. Presidential Election

January 29, 2019

1 The 2000 U.S. Presidential Election

The 2000 presidential election—between Republican George W. Bush, Democrat Al Gore, and other third-party candidates—was one of the closest in American history. The election came down to one state, Florida, which Bush won by just 537 votes (out of nearly 6,000,000 votes cast in the state).

After Election Day, Democrats claimed that the "butterfly ballot" that was used in Palm Beach County confused Gore voters into voting for Reform Party candidate Pat Buchanan. The ballot in question is shown below. To vote for Gore, who is listed second on the left, a voter actually had to punch the third hole (because the second hole is actually a vote for Buchanan, who is listed first on the right).



In this lab, you will evaluate this. The data file `https://raw.githubusercontent.com/dlsun/data-science-` contains county-level information about:

- the number of votes for Gore, Bush, Buchanan (and a few other candidates) in the 2000 presidential election
- the number of votes for Clinton (Democrat), Dole (Republican), and Perot (Reform) in the 1996 presidential election
- the number of votes for Buchanan in the 1996 primary

- the number of registered Reform voters and the total number of registered voters

Using this data, evaluate the claim that many voters in Palm Beach County voted for Buchanan when they intended to vote for Gore. (*Hint:* You should check whether Palm Beach County fits the general pattern of the other counties in Florida. Visualizations will likely be more helpful than summary statistics.) Then, craft a story that guides the reader through your discoveries. Your story should contain both figures and explanations.

In [1]:
```
%matplotlib inline
import matplotlib.pyplot as plt
import pandas as pd
from altair import *

election = pd.read_csv("https://raw.githubusercontent.com/dlsun/data-science-book/maste
election.head()
```

Out[1]:

	county	buchanan2000	gore2000	bush2000	nader2000	browne2000	total2000	\
0	ALACHUA	262	47,300	34,062	3,215	658	85,235	
1	BAKER	73	2,392	5,610	53	17	8,072	
2	BAY	248	18,850	38,637	828	171	58,486	
3	BRADFORD	65	3,072	5,413	84	28	8,597	
4	BREVARD	570	97,318	115,185	4,470	643	217,616	

	clinton96	dole96	perot96	buchanan96p	reform.reg	total.reg
0	40,144	25,303	8,072	2,151	91	120,867
1	2,273	3,684	667	73	4	12,352
2	17,020	28,290	5,922	1,816	55	92,749
3	3,356	4,038	819	155	3	13,547
4	80,416	87,980	25,249	7,927	148	283,680

In [2]:
```
correct_palm_beach = election.iloc[49]["total.reg"]
correct_palm_beach

part1 = election[0:49]
part1

part2 = election[49:]
part2["total.reg"] = part2["total.reg"].shift(-1)
```

```
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:8: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html
```

We have read in the data and are now going to explore whether or not the butterfly ballot had an effect on Palm Beach County voters in the 2000 U.S. Election. However, upon further investigation of the data set, it has become apparent that the Total Registration Voters is incorrect. To correct

the problem, we have shifted the values in the total registration column up one position starting from the incorrect position and have reassigned Palm Beach County its correct total number of registered voters.

```
In [3]: election = part1.append(part2)

In [4]: election["total.reg"][66] = correct_palm_beach

/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html
  """Entry point for launching an IPython kernel.


In [5]: election = election.set_index("county")

In [6]: election["buchanan2000"] = election["buchanan2000"].str.replace(",", "")
        election["buchanan2000"] = pd.to_numeric(election["buchanan2000"])

        election["gore2000"] = election["gore2000"].str.replace(",", "")
        election["gore2000"] = pd.to_numeric(election["gore2000"])

        election["bush2000"] = election["bush2000"].str.replace(",", "")
        election["bush2000"] = pd.to_numeric(election["bush2000"])

        election["total2000"] = election["total2000"].str.replace(",", "")
        election["total2000"] = pd.to_numeric(election["total2000"])
        election["total2000"] = election["total2000"] + election["buchanan2000"]

        election["total.reg"] = election["total.reg"].str.replace(",","")
        election["total.reg"] = pd.to_numeric(election["total.reg"])
```

In order to analyze the data, we must make the values numeric by replacing the comma in the entries and then converting the columns to numeric. Now that the data is clean and correctly formatted, we can start our analysis.

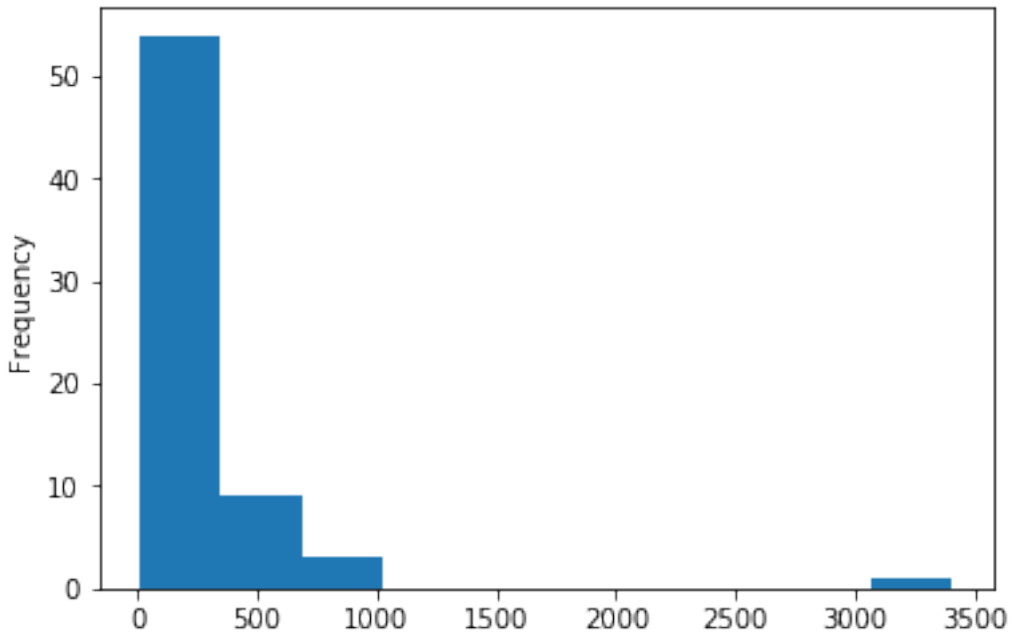
Let's first take a look at the values of each variable for Palm Beach County. Notice that Al Gore and George Bush dominate the votes, but what really sticks out is that Pat Buchanan received 3407 votes, but there were only 9 registered Reform voters. Perhaps, many Reform voters registered at other counties than which they voted at, so we will first explore if the distribution of Buchanan's votes is normal.

```
In [7]: election.loc["PALM BEACH"]

Out[7]: buchanan2000       3407
        gore2000         268945
        bush2000         152846
        nader2000         5,564
        browne2000          743
```

```
total2000        431505
clinton96       230,621
dole96          133,762
perot96          30,739
buchanan96p       8,788
reform.reg            9
total.reg        656694
Name: PALM BEACH, dtype: object
```

```
In [8]: election.buchanan2000.plot.hist()
```

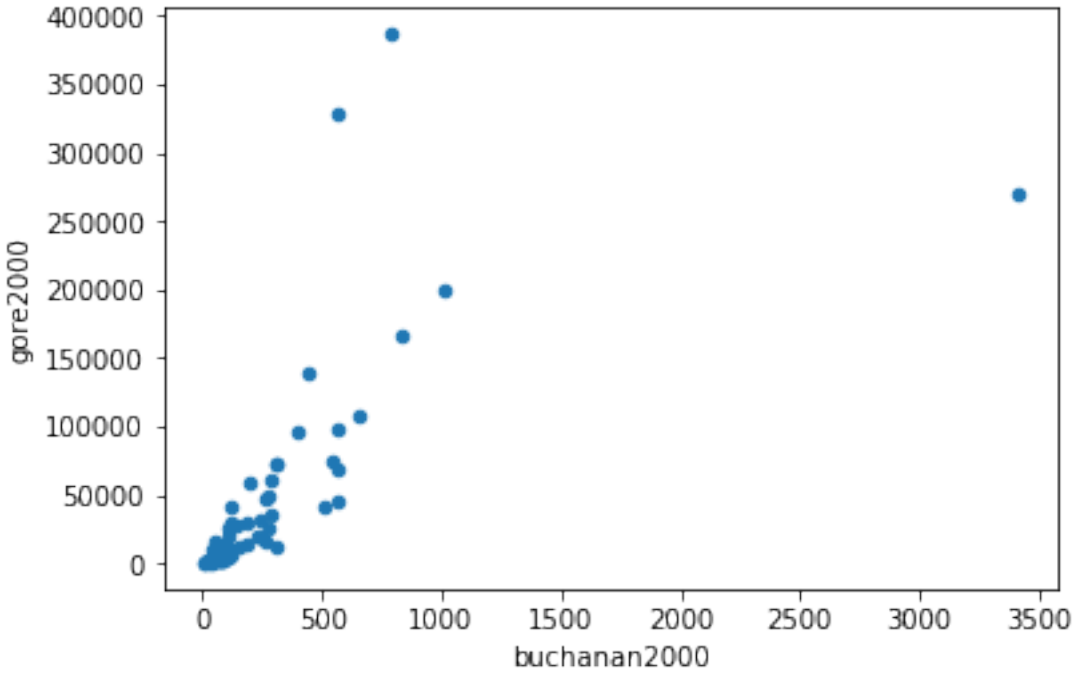
```
Out[8]: <matplotlib.axes._subplots.AxesSubplot at 0x7effbeecbf28>
```



From this histogram, we see that there is one value past 3000 votes - Palm Beach County. This does raise some red flags and it appears that the Democrats may be correct. The scatter plot below shows as well that Al Gore was receiving more votes proportionally than Pat Buchanan in other counties in Florida, except for Palm Beach. It might be worth be considering to look at the proportion of votes for Pat Buchanan across all counties in Florida before we make any conclusions.

```
In [9]: election.plot.scatter("buchanan2000", "gore2000")
```

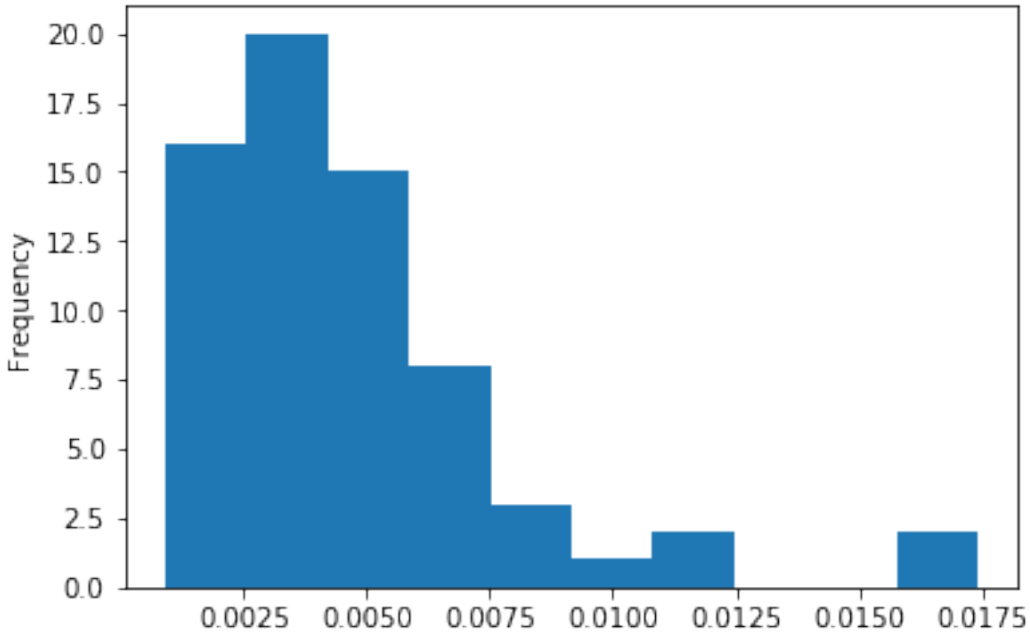
```
Out[9]: <matplotlib.axes._subplots.AxesSubplot at 0x7effbed8e4a8>
```




```
In [10]: election["buchanan_percent"] = election.buchanan2000/election.total2000

         election["gore_percent"] = election.gore2000/election.total2000

         election["bush_percent"] = election.bush2000/election.total2000
```

We have now made proportion variables in our data set and can explore the distribution of votes as a proportion relative to the total number of voters in each county. This will hopefully give us more evidence that the Butterfly Ballot did affect the way voters voted in Palm Beach County.

```
In [11]: election.buchanan_percent.plot.hist()
         election.buchanan_percent.describe()

Out[11]: count    67.000000
         mean      0.004665
         std       0.003210
         min       0.000898
         25%       0.002611
         50%       0.003978
         75%       0.005554
         max       0.017418
         Name: buchanan_percent, dtype: float64
```



We see that the distribution of proportions of votes for Buchanan is skewed right and that there is an observation that seems relatively further away from the rest of the data, also known as an outlier. According to the percent of Buchanan voters below from our summary of Palm Beach, only 0.7% of voters voted for Pat Buchanan. This means that Palm Beach County is not the furthest right point from the histogram above, but the summary statistics above does tell us that the proportion of people who voted for Buchanan from Palm Beach out of all counties in Florida, is in the upper 75% of all proportions.

So far, all signs are pointing to the idea that more people did vote for Buchanan than they were supposed to have in Palm Beach County. Earlier, we mentioned that there were only 9 registered Reform Voters so as one last test, let's see how big the difference is between registered reform voters and the number of people who voted for Buchanan in all counties.

```
In [12]: election.loc["PALM BEACH"]
```

```
Out[12]: buchanan2000              3407
         gore2000                268945
         bush2000                152846
         nader2000                5,564
         browne2000                 743
         total2000               431505
         clinton96               230,621
         dole96                  133,762
         perot96                  30,739
         buchanan96p               8,788
         reform.reg                    9
         total.reg                656694
         buchanan_percent     0.00789562
```

```
gore_percent          0.623272
bush_percent          0.354216
Name: PALM BEACH, dtype: object
```

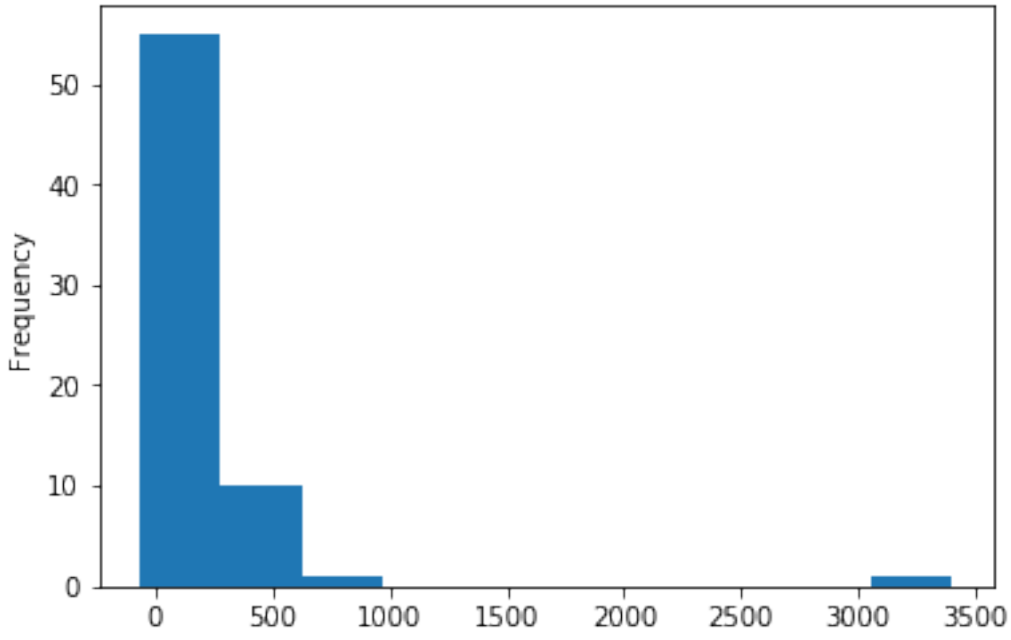
```
In [13]: election["reform_expected_votes"] = (
             election["reform.reg"]/election["total.reg"])*election["total2000"]
```

```
In [14]: election["reform_expected_votes"].loc["PALM BEACH"]
```

```
Out[14]: 5.9137817613683081
```

```
In [15]: (election["buchanan2000"] -
              election["reform_expected_votes"]).plot.hist()
```

```
Out[15]: <matplotlib.axes._subplots.AxesSubplot at 0x7effbc5fa8d0>
```



In order to calculate the difference in expected number of Reform votes and actual Reform votes, we should make an Expected Reform Votes Variable. We have taken the number of registered reform voters in each county, divided it by the total number of registered voters in each county, and have finally multiplied the result to the total number of votes. We see that the expected number of votes for Pat Buchanan in Palm Beach is a mere 6, but Buchanan actually received 3407 votes. According to the distribution of the Difference in Buchanan's votes and Reform Expected Votes, we see that Palm Beach County is a huge outlier.

Therefore, we conclude it is possible that the Butterfly Ballot did have an effect of how people voted in 2000. We cannot say for certain due to the discrepancies in our data. Is 9 reform registered voters actually accurate? How accurate are all the data values? Maybe these voters intended to vote for someone else other than Al Gore. At the end of the day, we will never really know, but from this analysis, it does appear that the Democrats had something to argue for.

2 Submission Instructions

Once you are finished, follow these steps:

1. Restart the kernel and re-run this notebook from beginning to end by going to `Kernel > Restart Kernel and Run All Cells`.
2. If this process stops halfway through, that means there was an error. Correct the error and repeat Step 1 until the notebook runs from beginning to end.
3. Double check that there is a number next to each code cell and that these numbers are in order.

Then, submit your lab as follows:

1. Go to `File > Export Notebook As > PDF`.
2. Double check that the entire notebook, from beginning to end, is in this PDF file. (If the notebook is cut off, try first exporting the notebook to HTML and printing to PDF.)
3. Upload the PDF to PolyLearn.