data_summary_report

October 20, 2016

1 Data Summary Report

This report is run to understand the data that was extracted for use in the algorithm.

```
In [11]: %matplotlib inline
         import numpy as np
         import pandas as pd
         import matplotlib.pyplot as plt
```

1.1 Survey Questions Included

This is the list of questions included in the thermometer dataset.

```
In [8]: # read in the data
        df = pd.read_hdf('../data/anes_timeseries_thermometer.h5', 'main')
```

```
In [15]: # list out survey questions
         dfinfo = pd.read_excel('../data/thermometer_vars.xlsx')
         print(dfinfo)
```

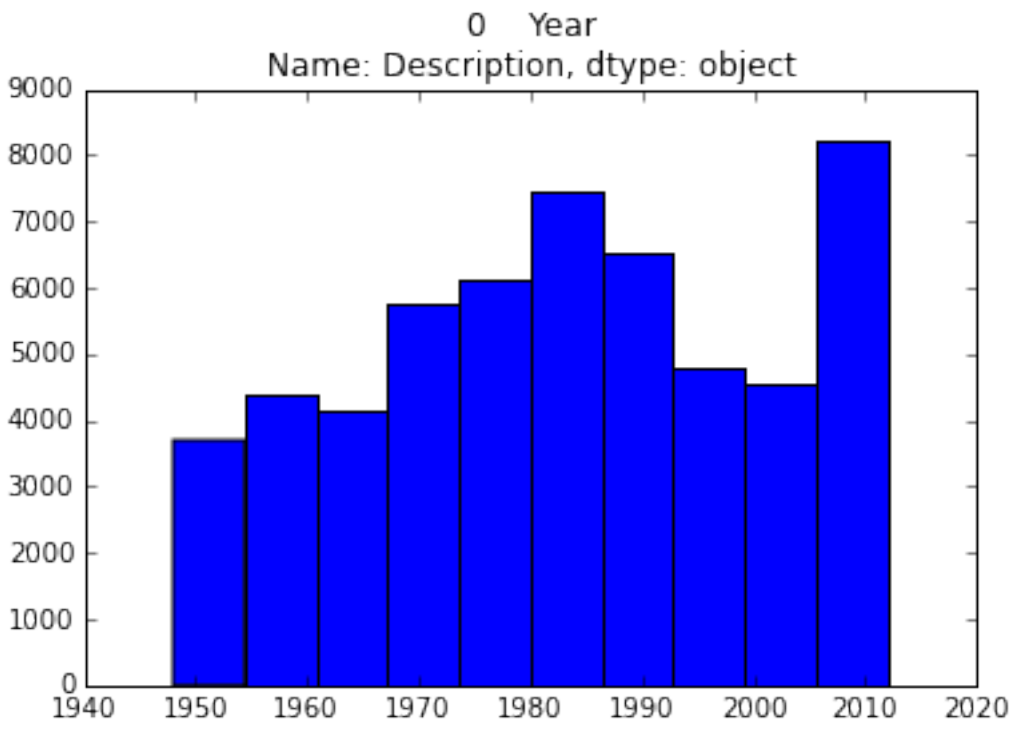
```
VarName                                  Description
0    VCF0004                                     Year
1   VCF0006a                                Respondant ID
2    VCF0013                          Pre-Election Survey
3    VCF0014                         Post-Election Survey
4   VCF0140a                         Respondant Education
5    VCF0301                             Party Identifier
6    VCF0201                 GROUP THERMOMETER: Democrats
7    VCF0202               GROUP THERMOMETER: Republicans
8    VCF0203               GROUP THERMOMETER: Protestants
9    VCF0204                 GROUP THERMOMETER: Catholics
10   VCF0205                      GROUP THERMOMETER: Jews
11   VCF0206                   GROUP THERMOMETER: Blacks
12   VCF0207                   GROUP THERMOMETER: Whites
13   VCF0208               GROUP THERMOMETER: Southerners
14   VCF0209              GROUP THERMOMETER: Big Business
15   VCF0210              GROUP THERMOMETER: Labor Unions
16   VCF0211                  GROUP THERMOMETER: Liberals
17   VCF0212            GROUP THERMOMETER: Conservatives
18   VCF0213                 GROUP THERMOMETER: Military
19   VCF0214                GROUP THERMOMETER: Policemen
20   VCF0215          GROUP THERMOMETER: Black Militants
21   VCF0216     GROUP THERMOMETER: Civil Rights Leaders
22   VCF0217       GROUP THERMOMETER: Chicanos/Hispanics
```

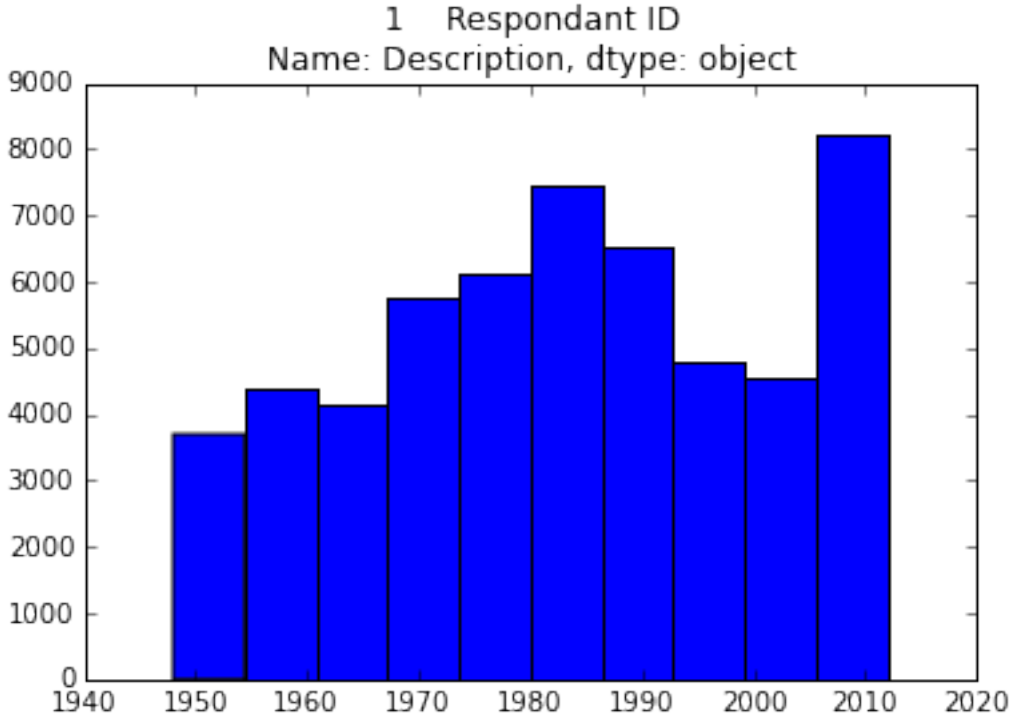
```
23   VCF0218           GROUP THERMOMETER: Democratic Party
24   VCF0219        GROUP THERMOMETER: Middle Class People
25   VCF0220          GROUP THERMOMETER: People on Welfare
26   VCF0221     GROUP THERMOMETER: Political Independents
27   VCF0222          GROUP THERMOMETER: Political Parties
28   VCF0223                GROUP THERMOMETER: Poor People
29   VCF0224           GROUP THERMOMETER: Republican Party
30   VCF0225             GROUP THERMOMETER: Womens Libbers
31   VCF0226               GROUP THERMOMETER: Young People
32   VCF0227            GROUP THERMOMETER: Asian-Americans
33   VCF0228                   GROUP THERMOMETER: Congress
34   VCF0229          GROUP THERMOMETER: Environmentalists
35   VCF0230          GROUP THERMOMETER: Anti-Abortionists
36   VCF0231         GROUP THERMOMETER: Federal Government
37   VCF0232          GROUP THERMOMETER: Gays and Lesbians
38   VCF0233             GROUP THERMOMETER: Illegal Aliens
39   VCF0234  GROUP THERMOMETER: Christian Fundamentalists
40   VCF0235           GROUP THERMOMETER: Radical Students
41   VCF0236                    GROUP THERMOMETER: Farmers
42   VCF0253                  GROUP THERMOMETER: Feminists
```

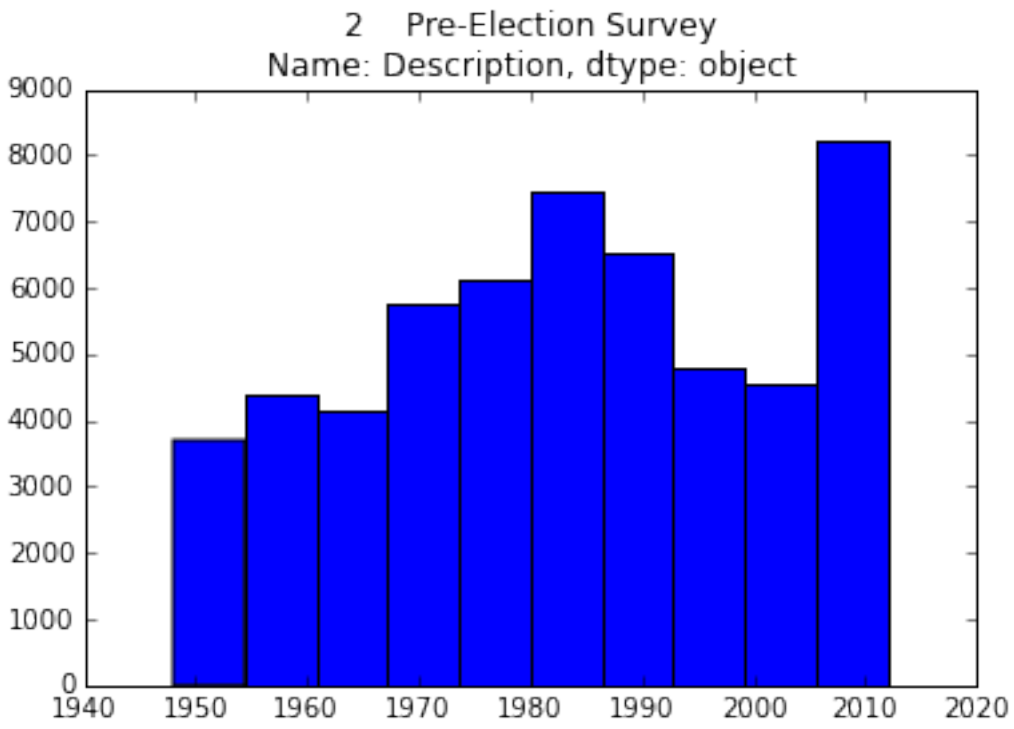
1.2 Missing Data Report

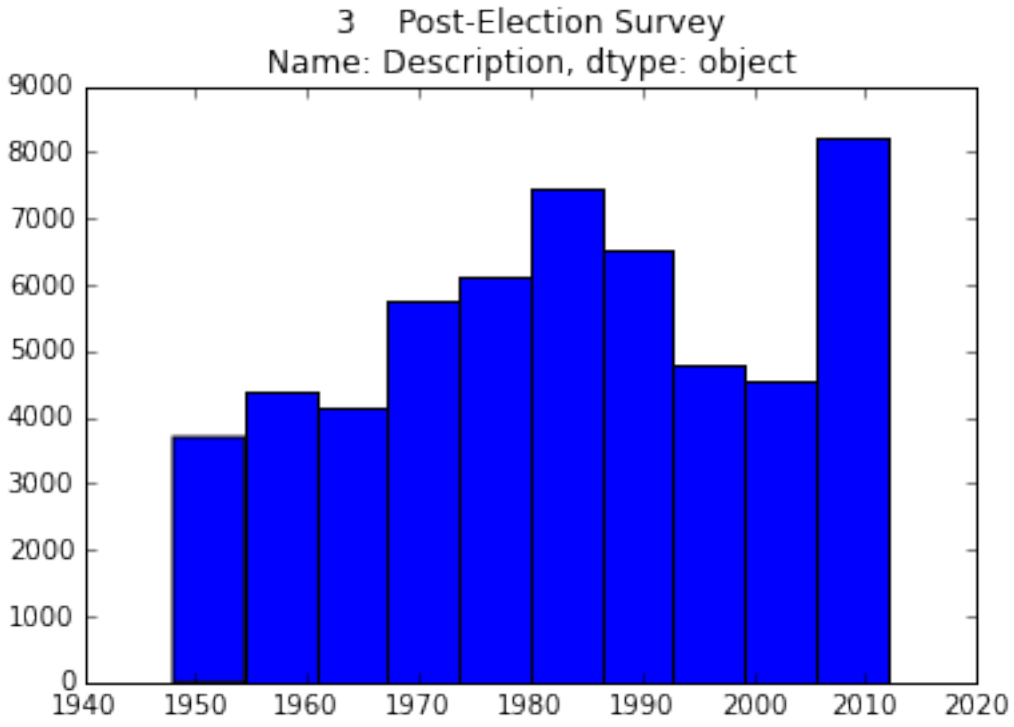
This section will look through each question and examine the number of responses over the years.

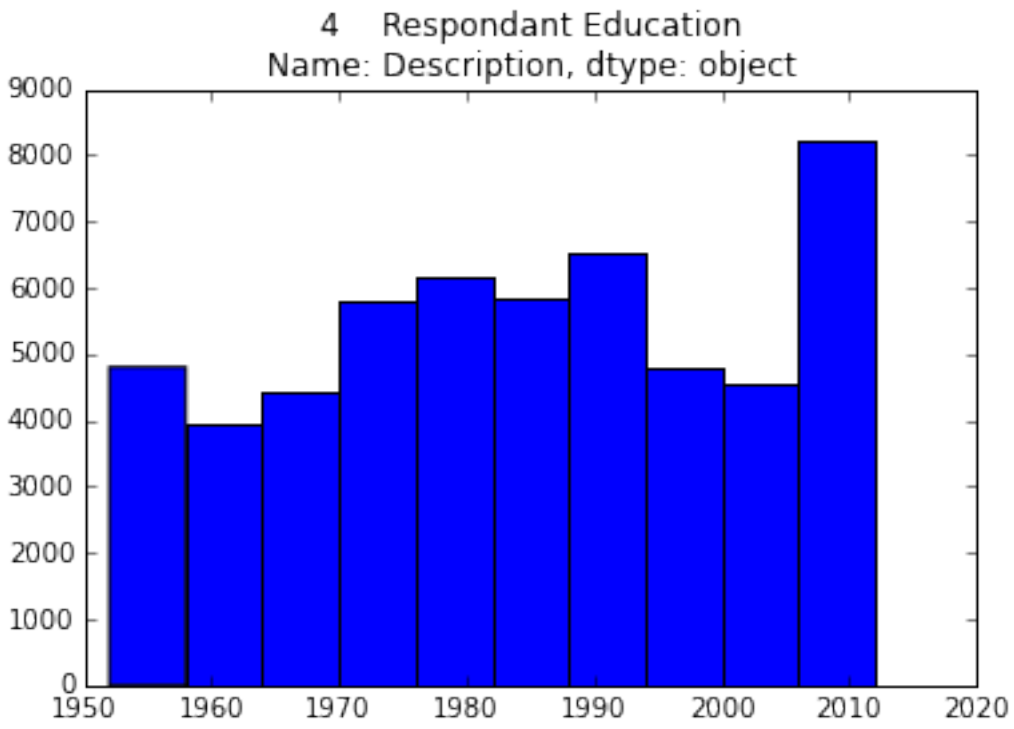
```
In [23]: # check how much data is missing
         for var in dfinfo['VarName']:
             years = df['VCF0004'][np.logical_not(pd.isnull(df[var]))]
             plt.hist(years.values)

             plt.title(dfinfo['Description'][dfinfo['VarName'] == var])

             plt.show()
```

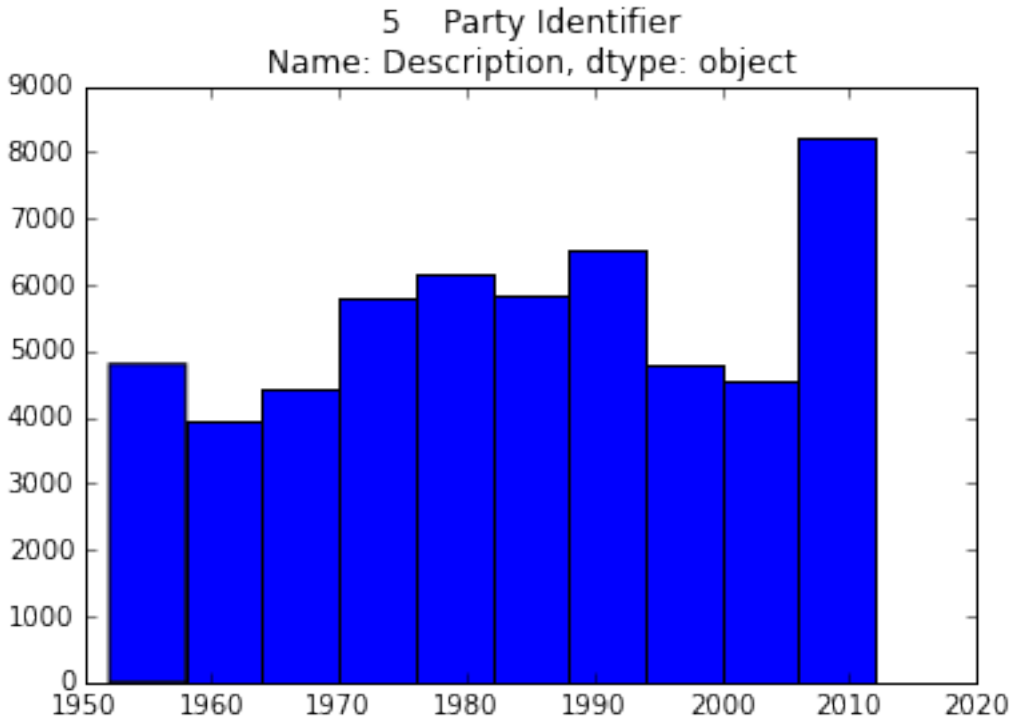

0 Year
Name: Description, dtype: object
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
1940
1950
1960
1970
1980
1990
2000
2010
2020


1 Respondant ID
Name: Description, dtype: object
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
1940
1950
1960
1970
1980
1990
2000
2010
2020

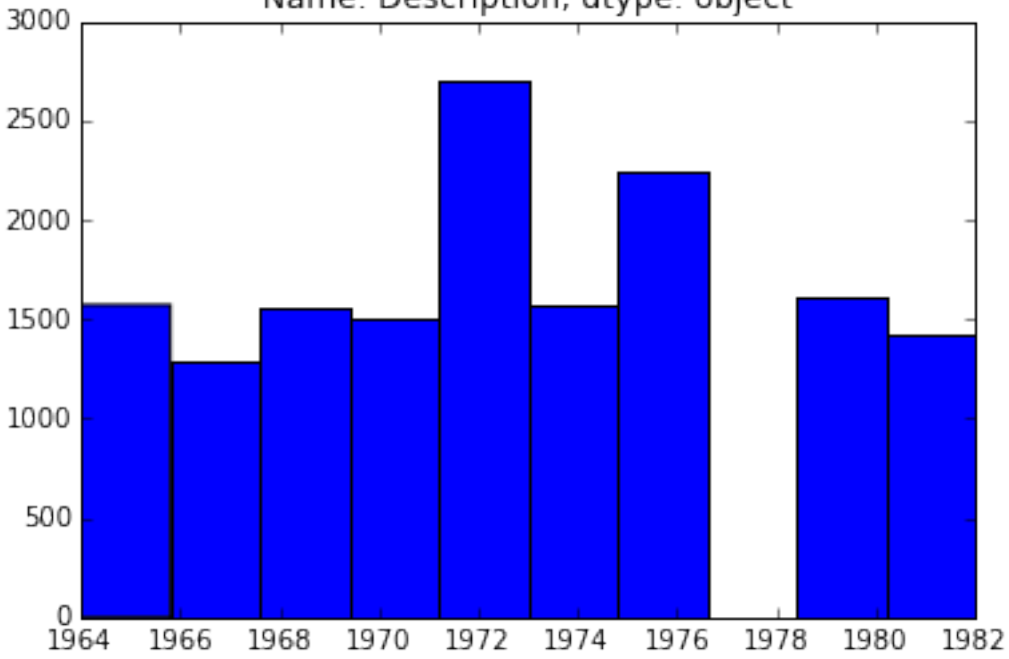

2 Pre-Election Survey
Name: Description, dtype: object
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
1940
1950
1960
1970
1980
1990
2000
2010
2020


3 Post-Election Survey
Name: Description, dtype: object
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
1940
1950
1960
1970
1980
1990
2000
2010
2020

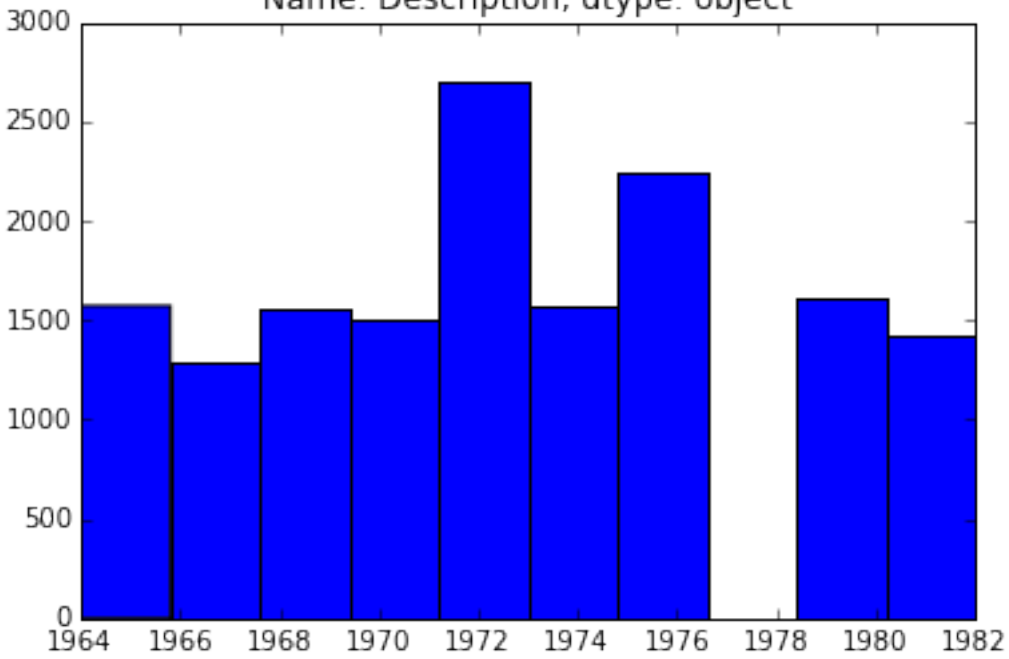

4 Respondant Education
Name: Description, dtype: object
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
1950
1960
1970
1980
1990
2000
2010
2020

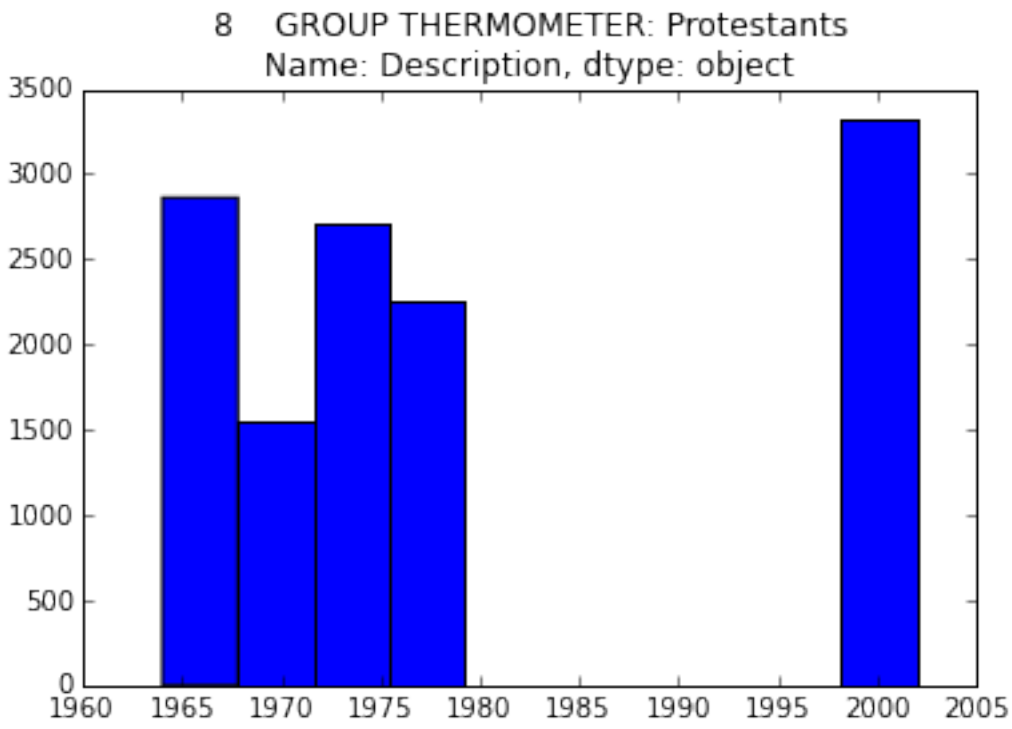

5 Party Identifier
Name: Description, dtype: object
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
1950
1960
1970
1980
1990
2000
2010
2020

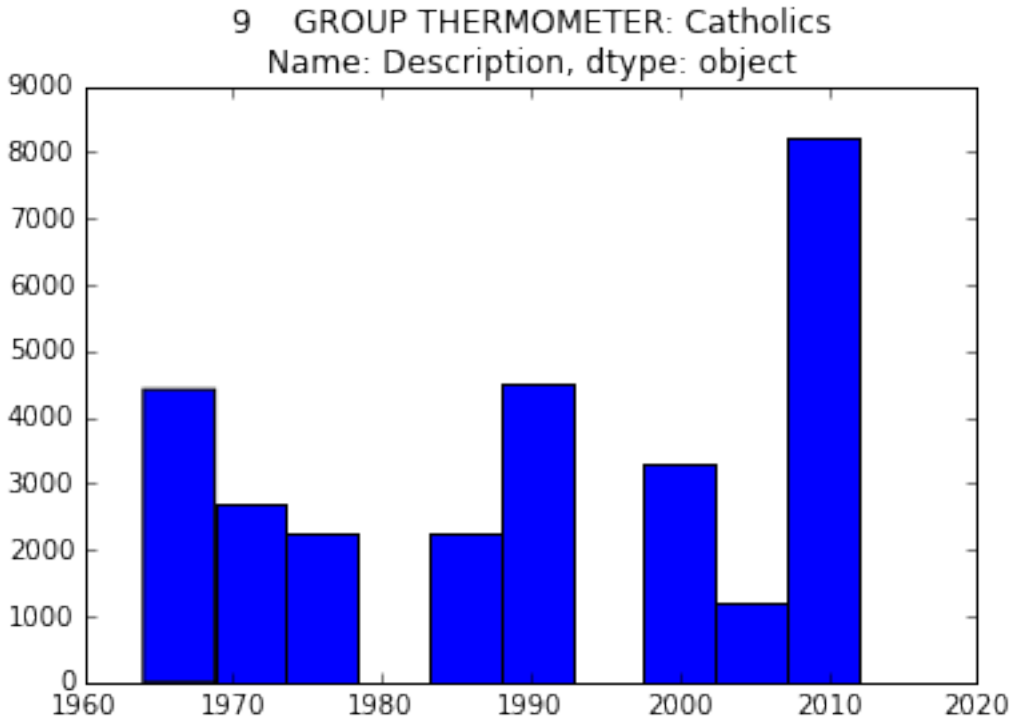
6 GROUP THERMOMETER: Democrats
Name: Description, dtype: object

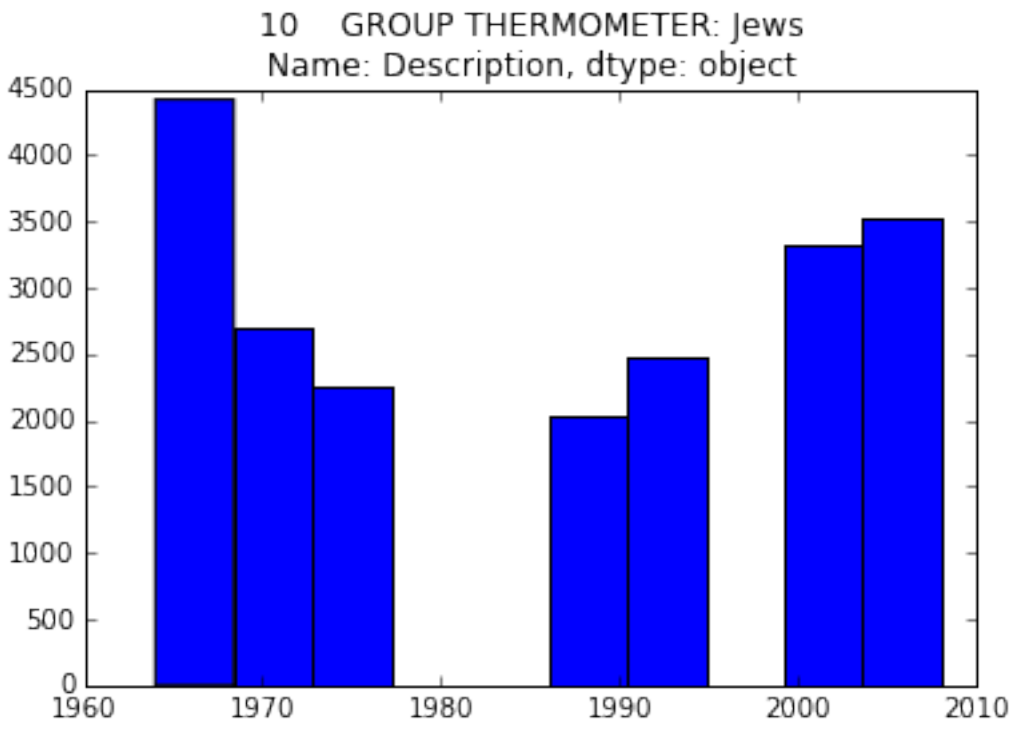


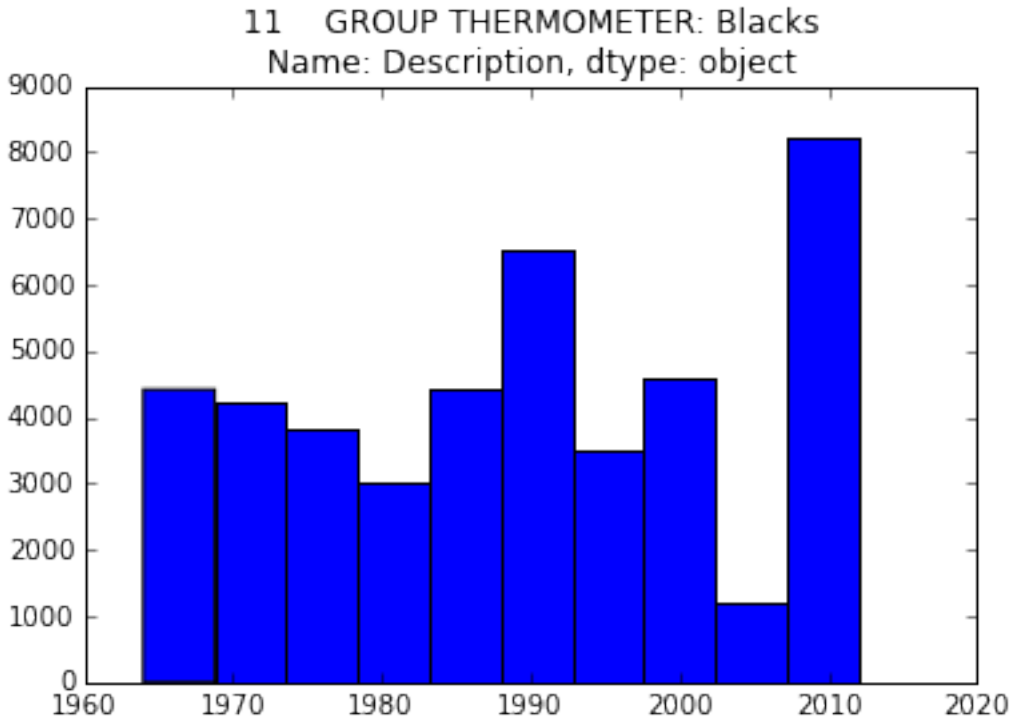
7 GROUP THERMOMETER: Republicans
Name: Description, dtype: object

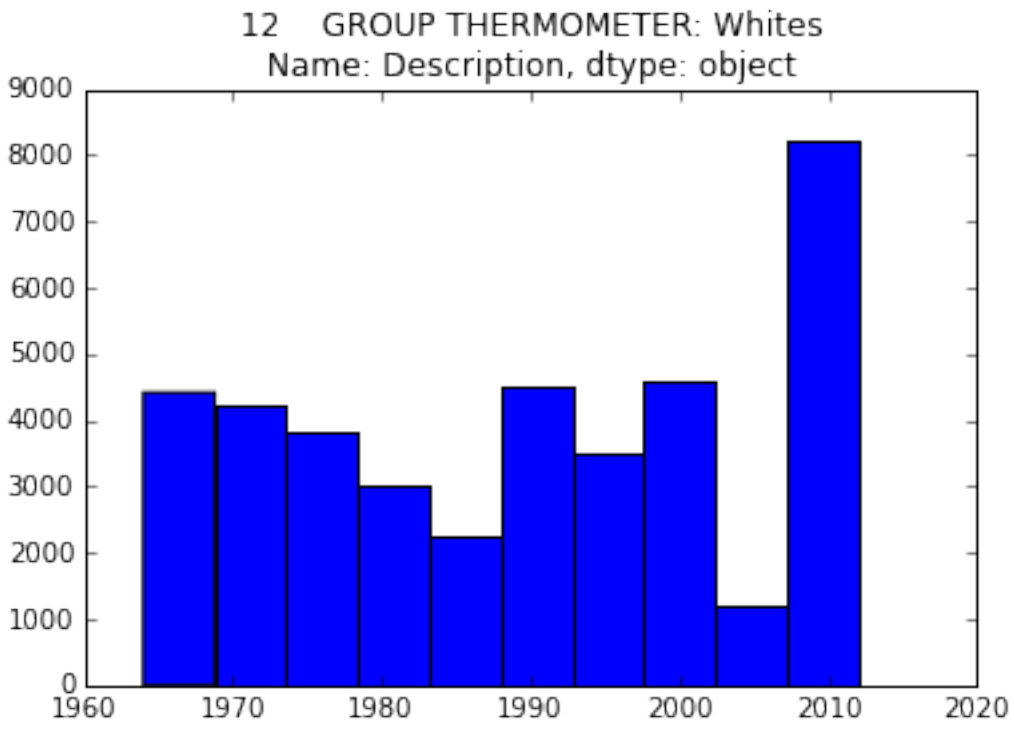


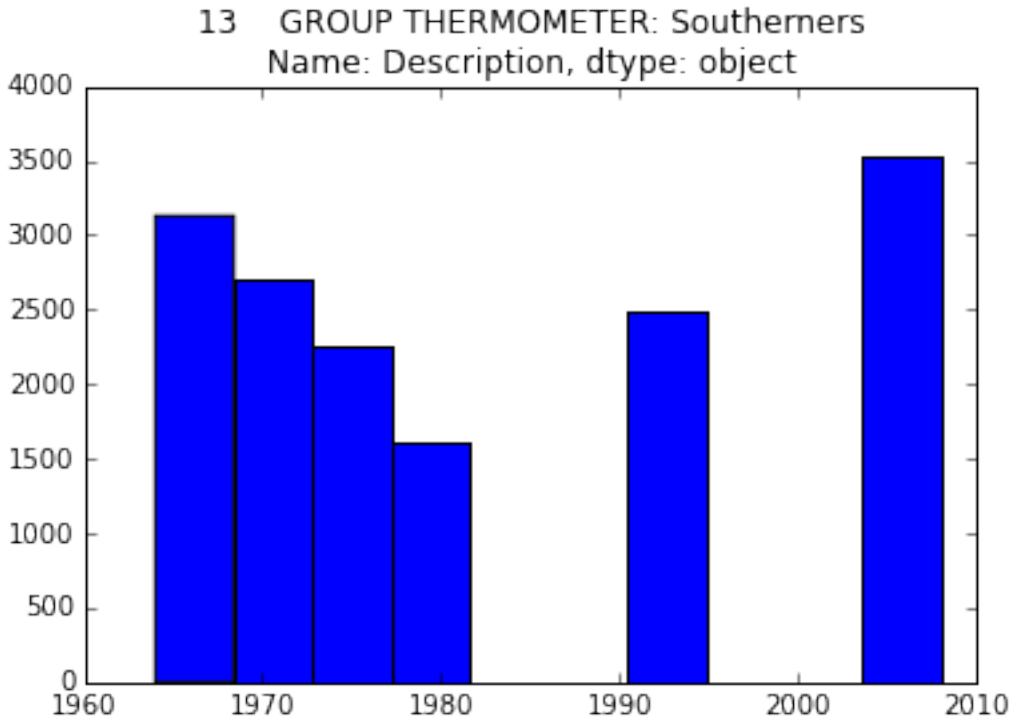

8 GROUP THERMOMETER: Protestants
Name: Description, dtype: object
3500
3000
2500
2000
1500
1000
500
0
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005


9 GROUP THERMOMETER: Catholics
Name: Description, dtype: object
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
1960
1970
1980
1990
2000
2010
2020

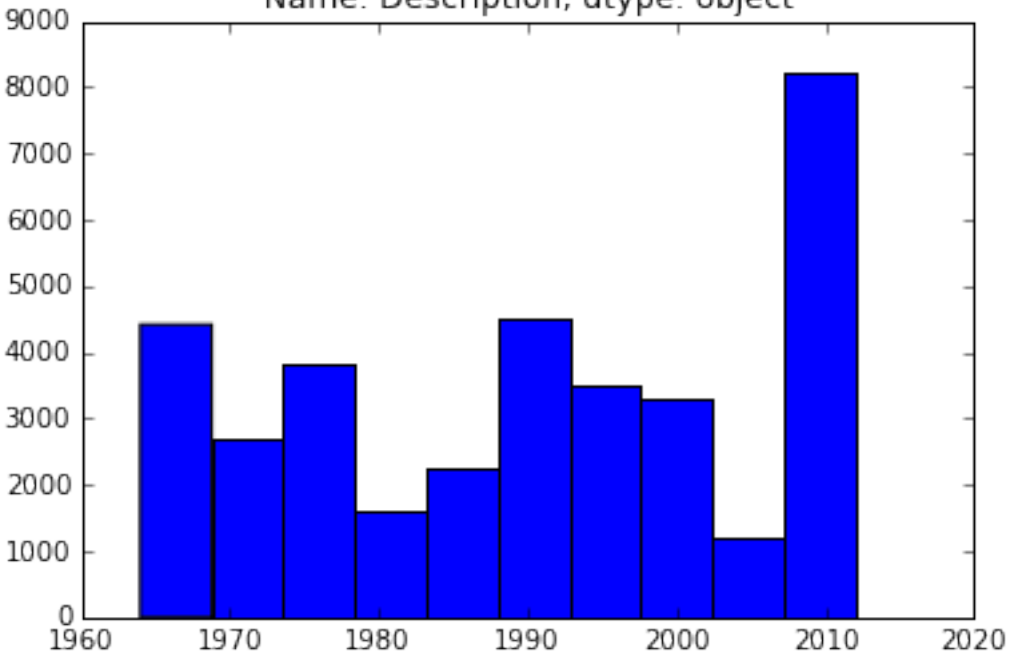

10 GROUP THERMOMETER: Jews
Name: Description, dtype: object
4500
4000
3500
3000
2500
2000
1500
1000
500
0
1960
1970
1980
1990
2000
2010


11 GROUP THERMOMETER: Blacks
Name: Description, dtype: object
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
1960
1970
1980
1990
2000
2010
2020

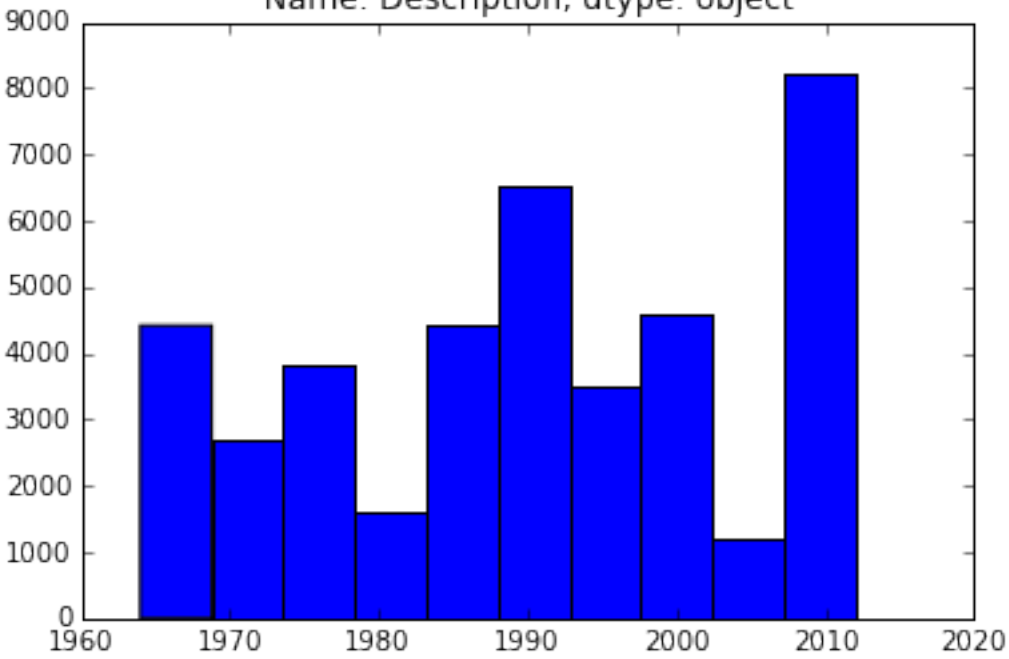

12 GROUP THERMOMETER: Whites
Name: Description, dtype: object
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
1960
1970
1980
1990
2000
2010
2020


13 GROUP THERMOMETER: Southerners
Name: Description, dtype: object
4000
3500
3000
2500
2000
1500
1000
500
0
1960
1970
1980
1990
2000
2010

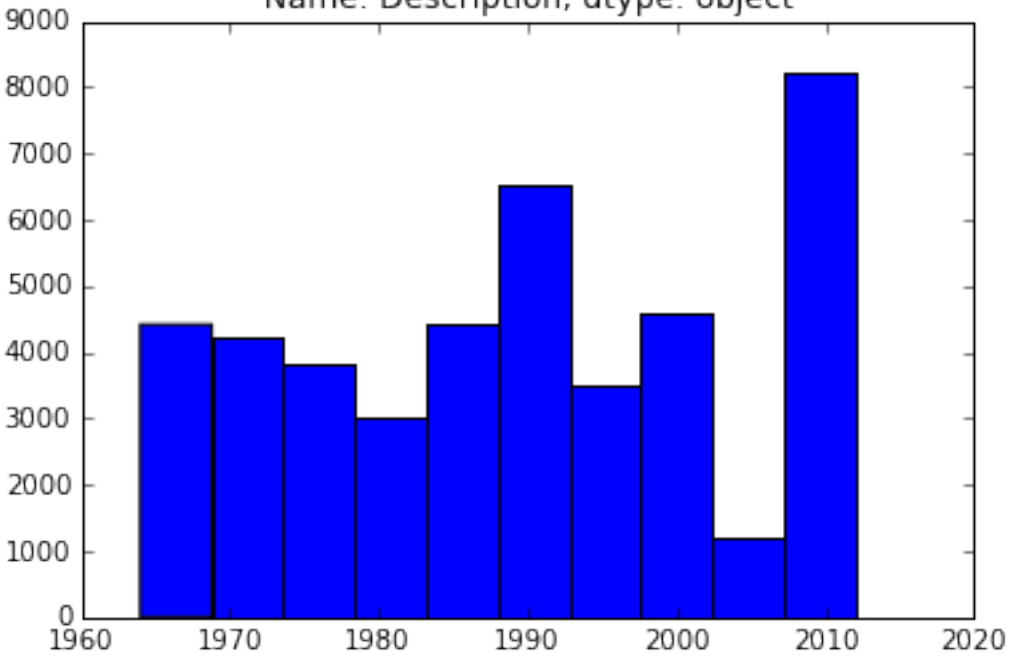
14 GROUP THERMOMETER: Big Business
Name: Description, dtype: object



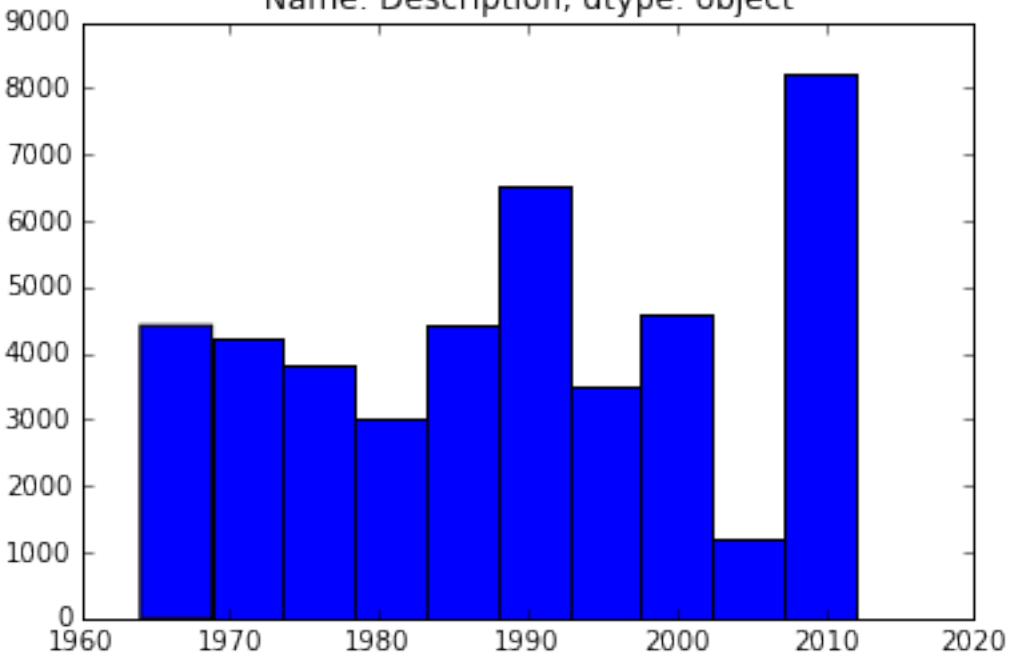
15 GROUP THERMOMETER: Labor Unions
Name: Description, dtype: object

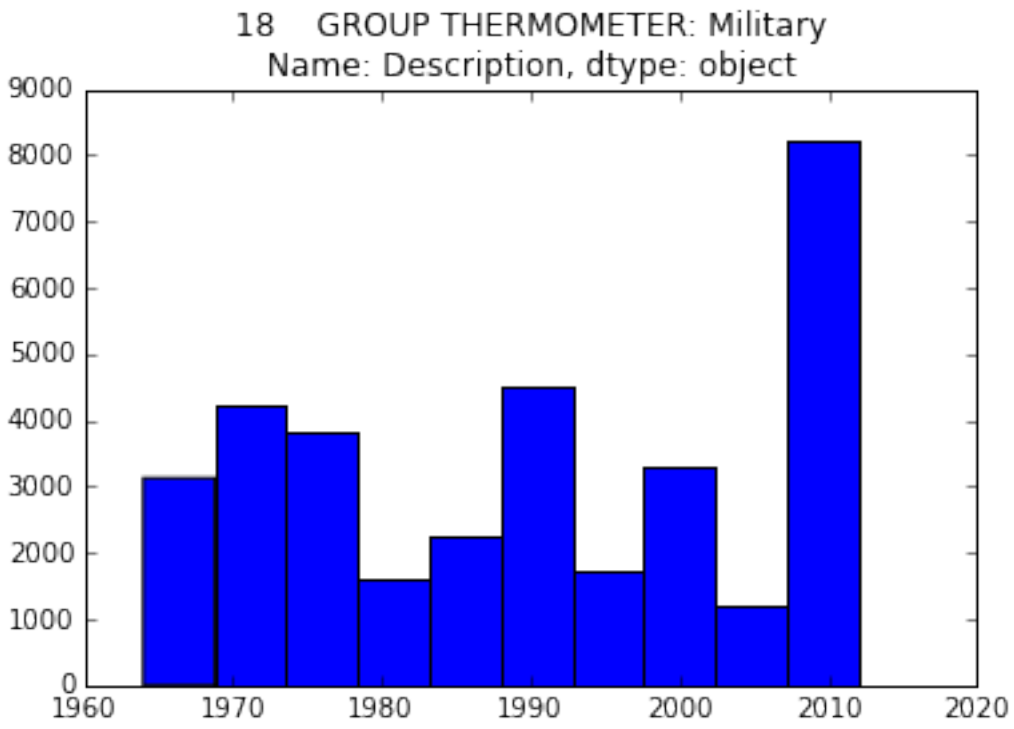


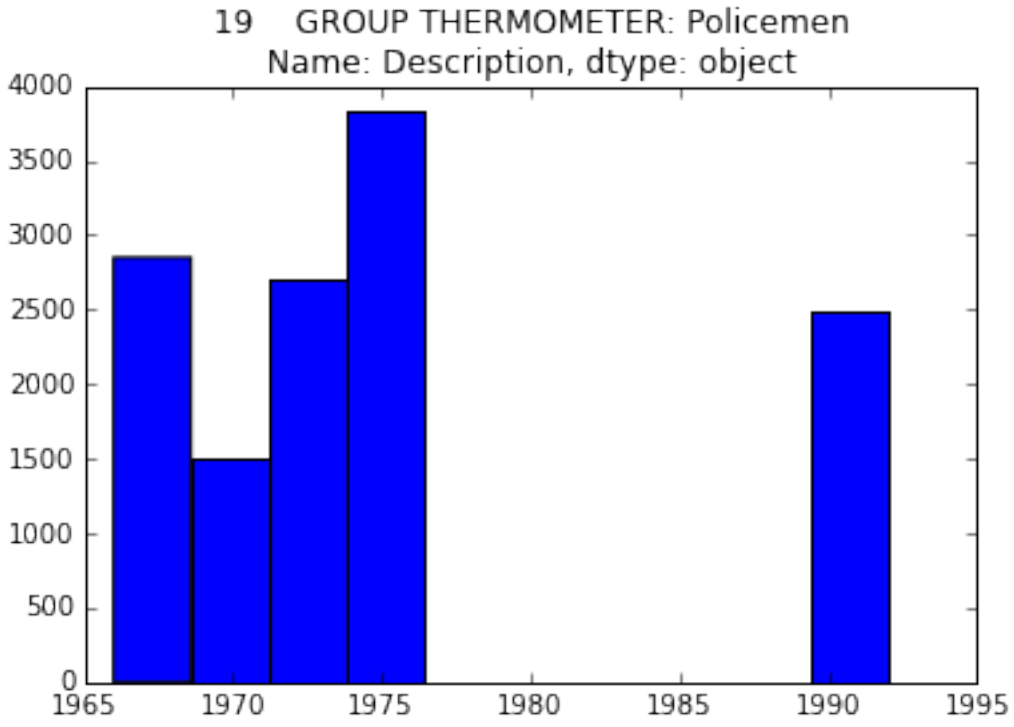
16 GROUP THERMOMETER: Liberals
Name: Description, dtype: object

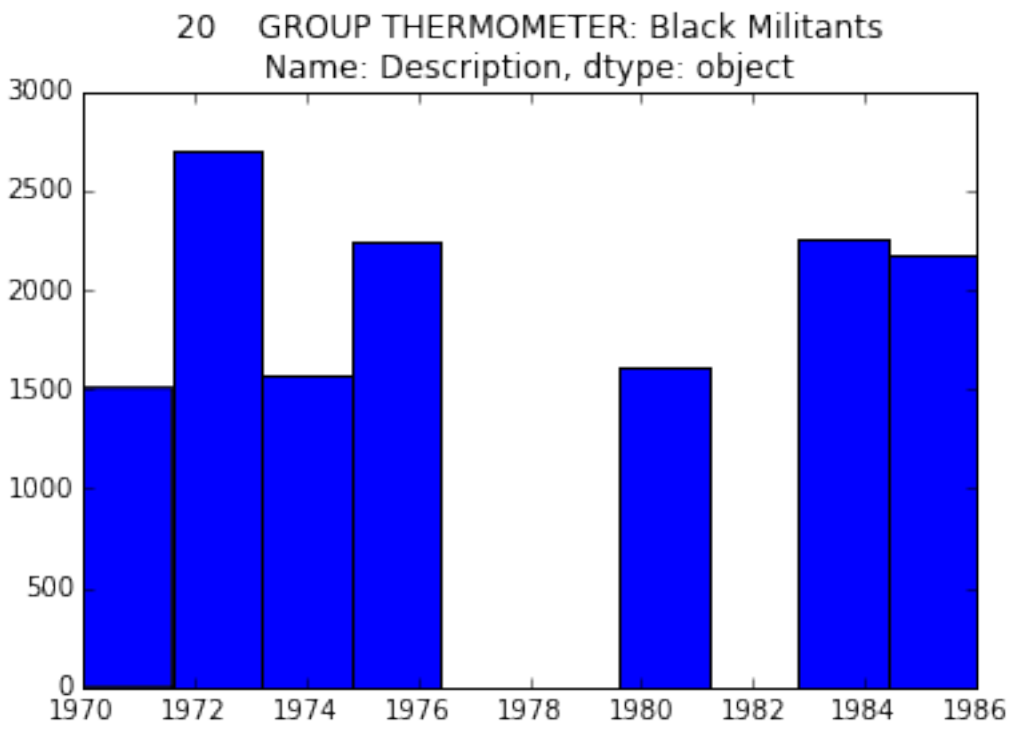


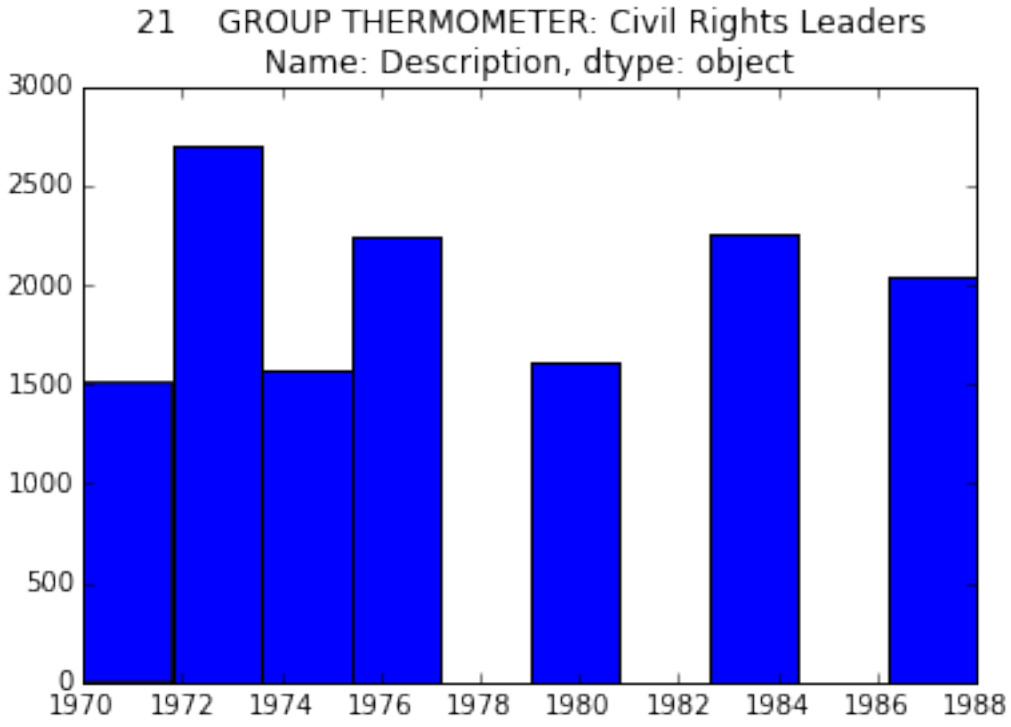
17 GROUP THERMOMETER: Conservatives
Name: Description, dtype: object



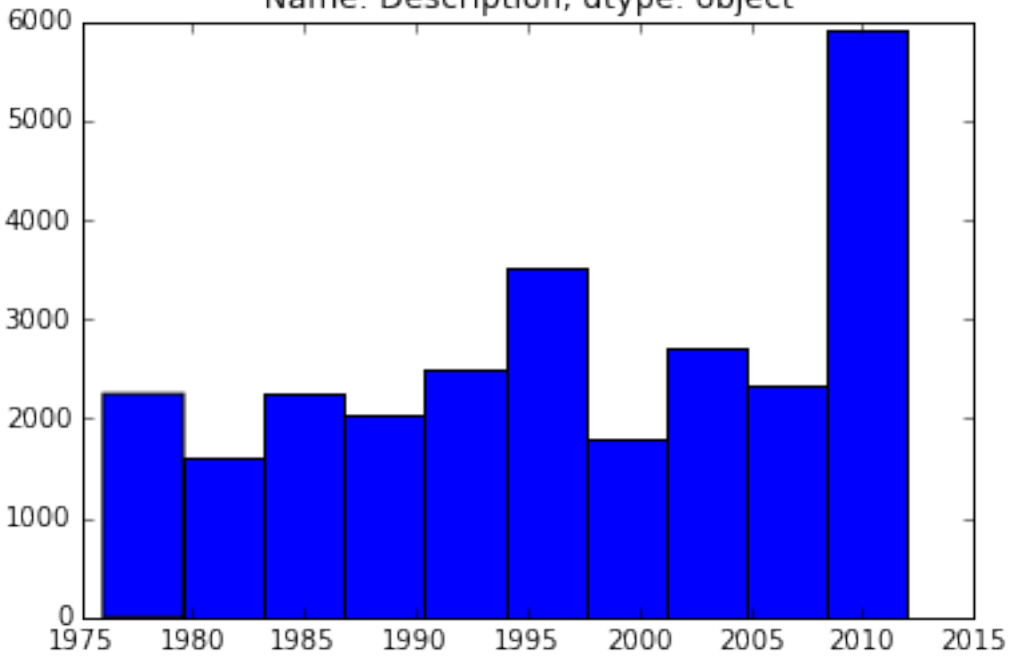

18 GROUP THERMOMETER: Military
Name: Description, dtype: object
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
1960
1970
1980
1990
2000
2010
2020


19 GROUP THERMOMETER: Policemen
Name: Description, dtype: object
4000
3500
3000
2500
2000
1500
1000
500
0
1965
1970
1975
1980
1985
1990
1995

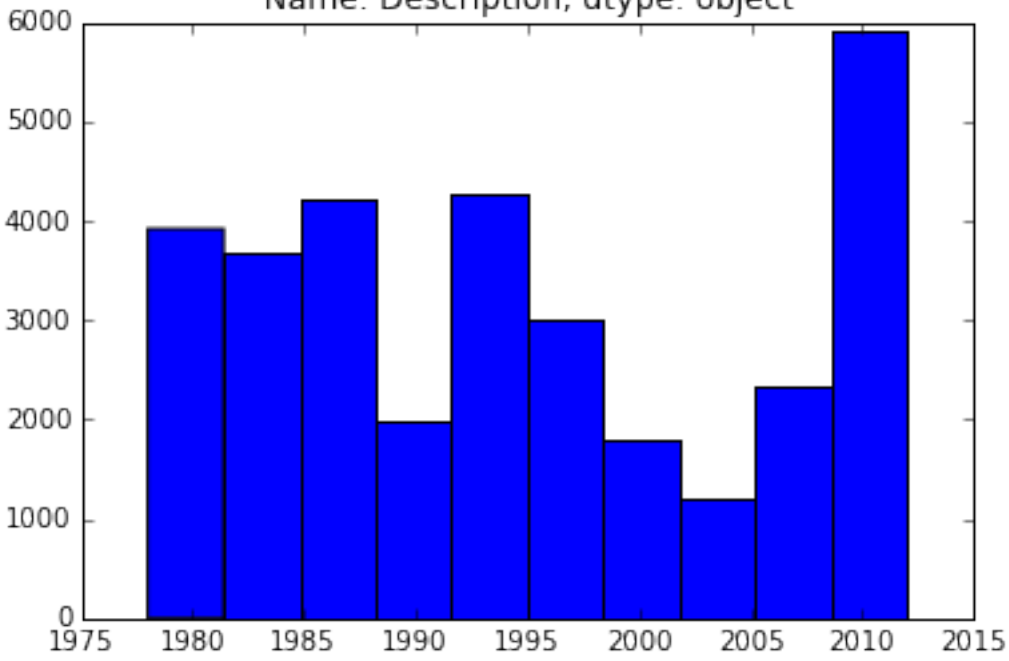

20 GROUP THERMOMETER: Black Militants
Name: Description, dtype: object
3000
2500
2000
1500
1000
500
0
1970
1972
1974
1976
1978
1980
1982
1984
1986


21 GROUP THERMOMETER: Civil Rights Leaders
Name: Description, dtype: object
3000
2500
2000
1500
1000
500
0
1970
1972
1974
1976
1978
1980
1982
1984
1986
1988

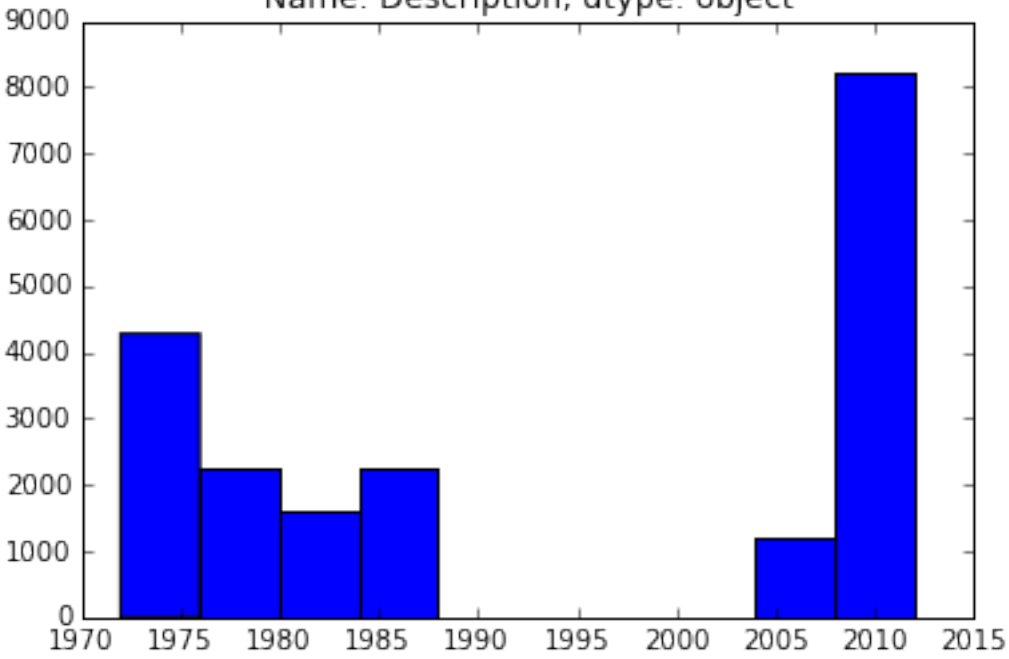
22 GROUP THERMOMETER: Chicanos/Hispanics
Name: Description, dtype: object



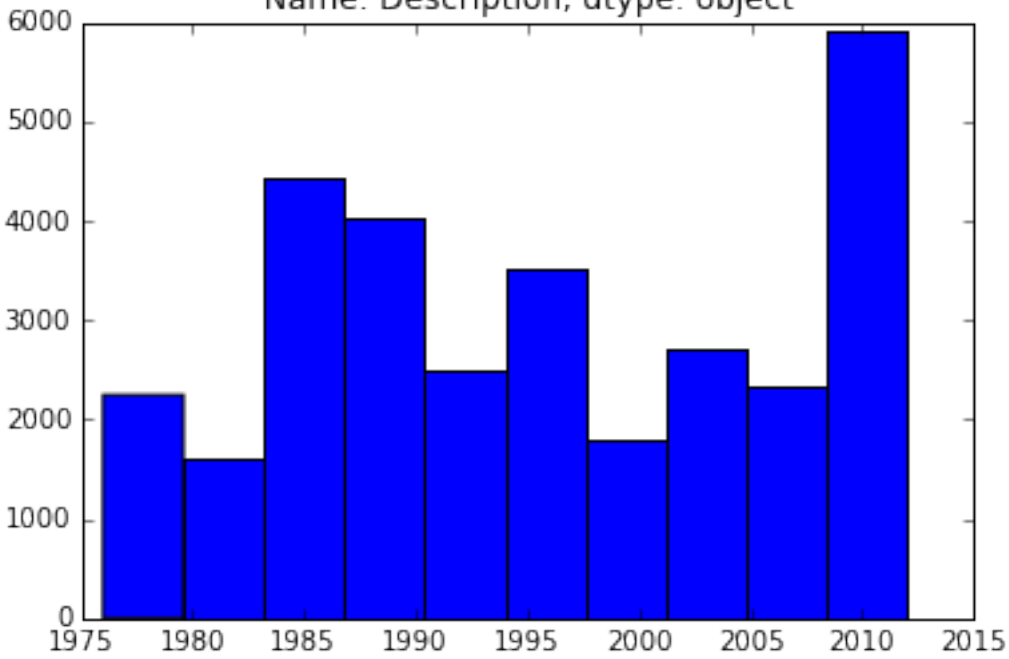
23 GROUP THERMOMETER: Democratic Party
Name: Description, dtype: object

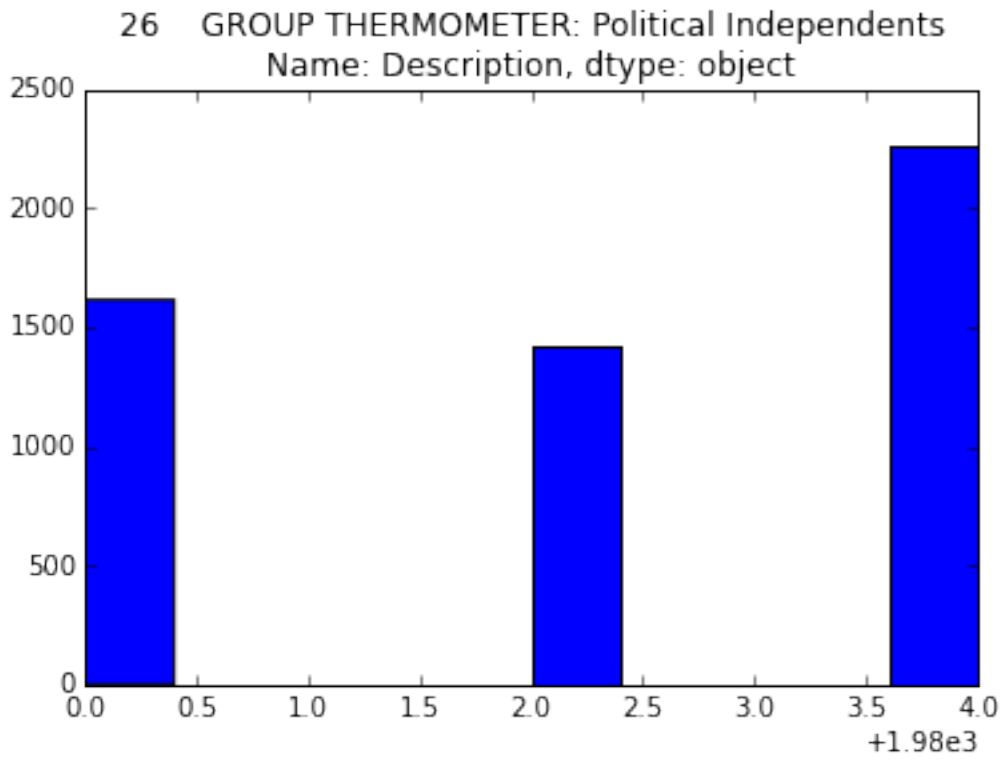


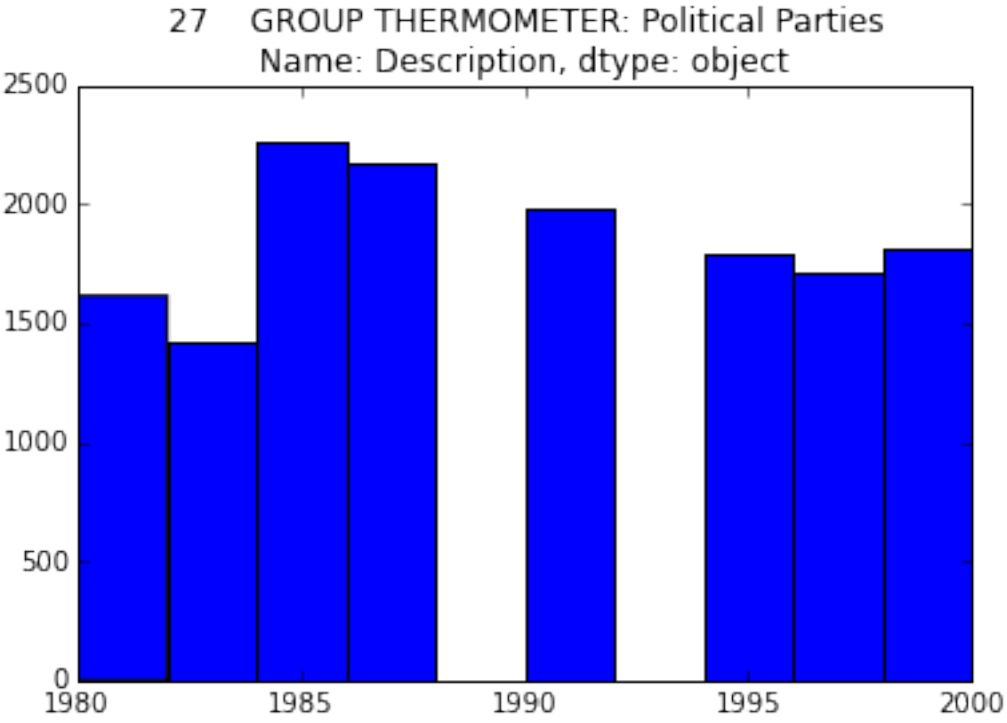
24 GROUP THERMOMETER: Middle Class People
Name: Description, dtype: object

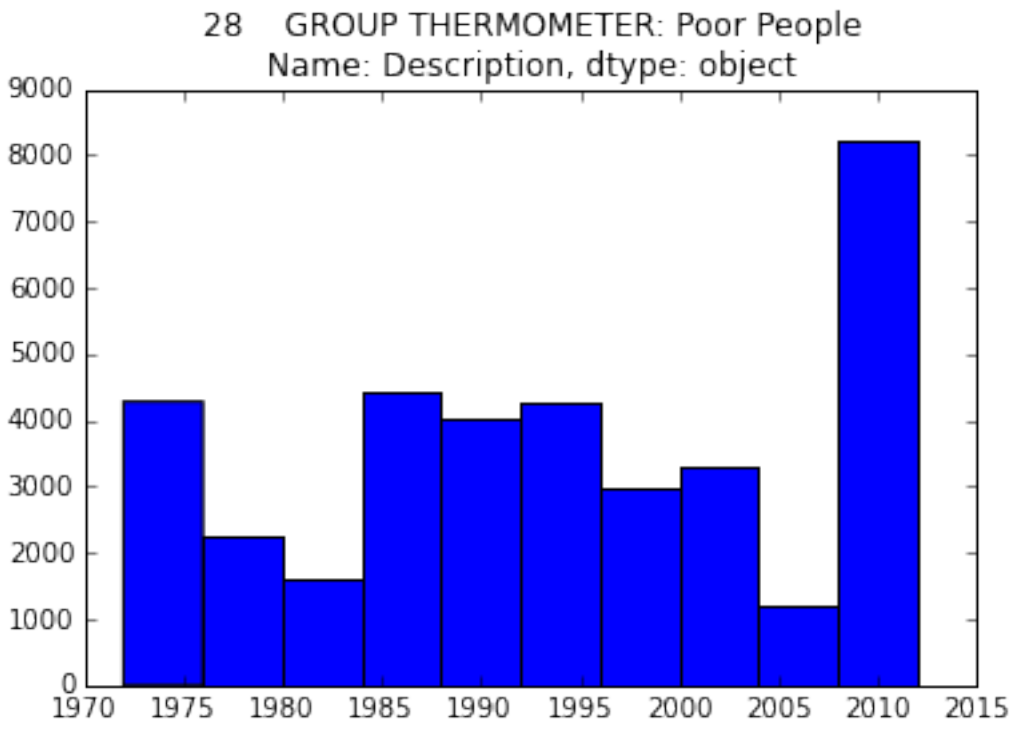


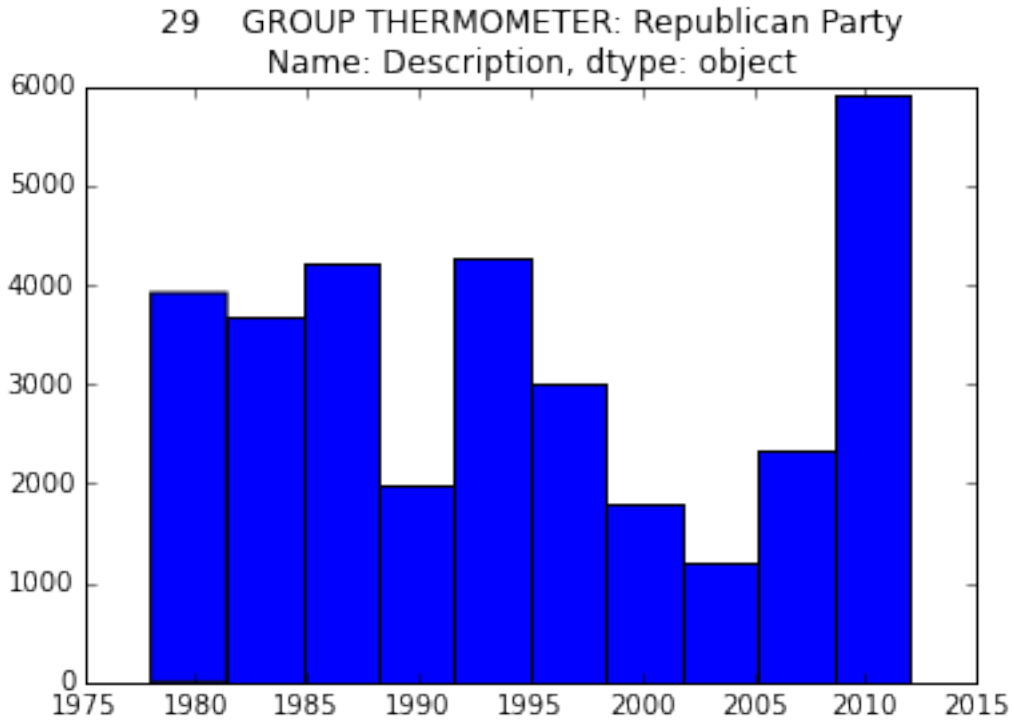
25 GROUP THERMOMETER: People on Welfare
Name: Description, dtype: object

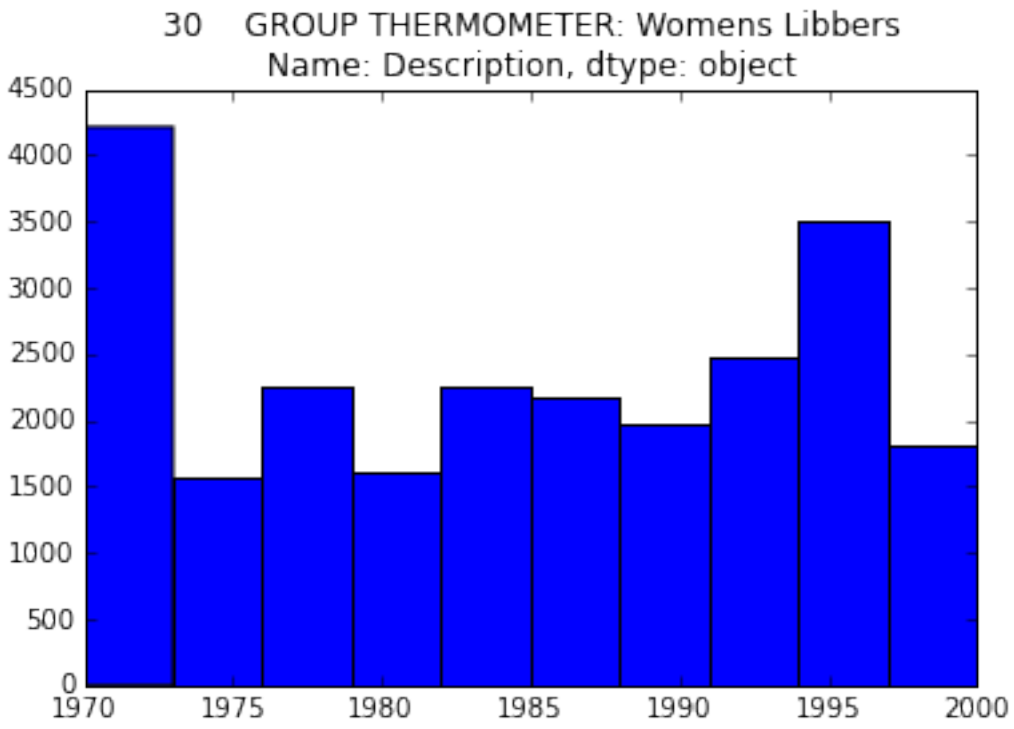


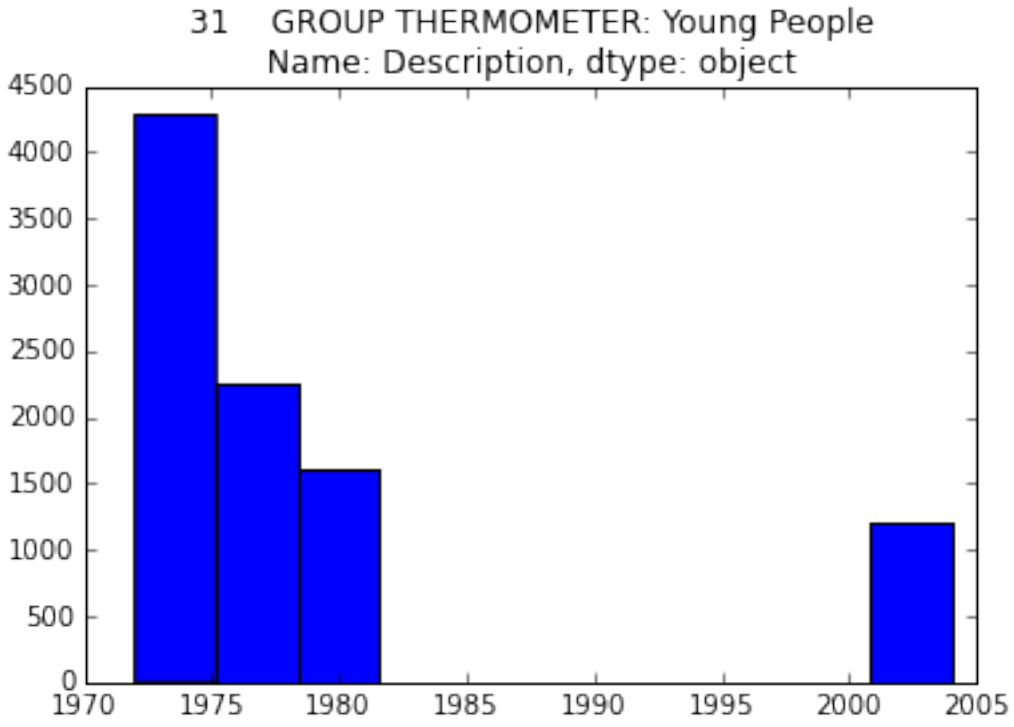

26 GROUP THERMOMETER: Political Independents
Name: Description, dtype: object
2500
2000
1500
1000
500
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
+1.98e3


27 GROUP THERMOMETER: Political Parties
Name: Description, dtype: object
2500
2000
1500
1000
500
0
1980
1985
1990
1995
2000

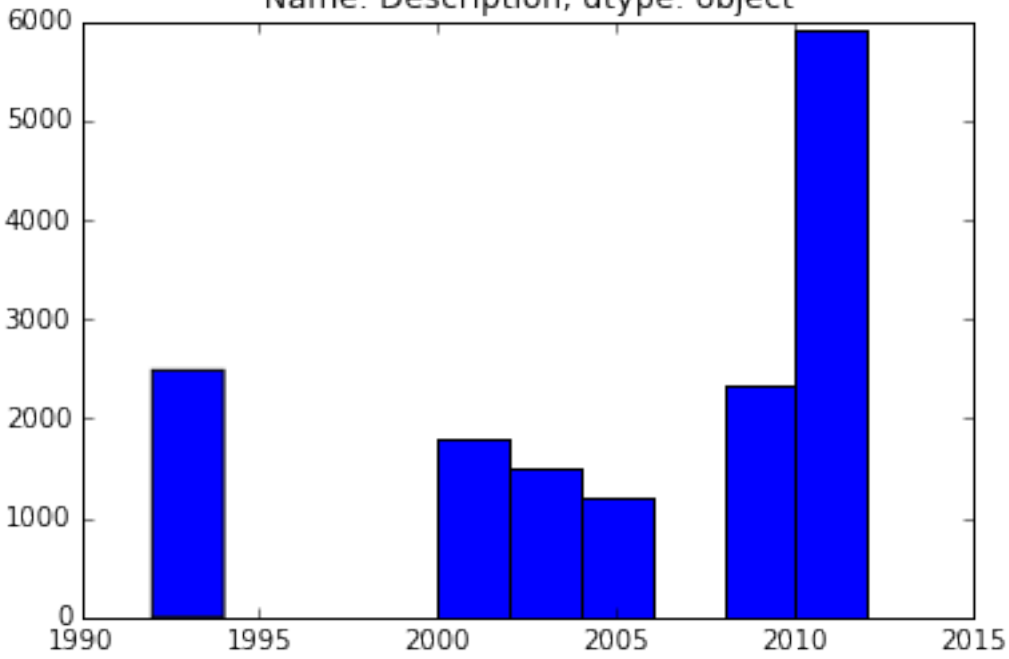

28 GROUP THERMOMETER: Poor People
Name: Description, dtype: object
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015


29 GROUP THERMOMETER: Republican Party
Name: Description, dtype: object
6000
5000
4000
3000
2000
1000
0
1975
1980
1985
1990
1995
2000
2005
2010
2015

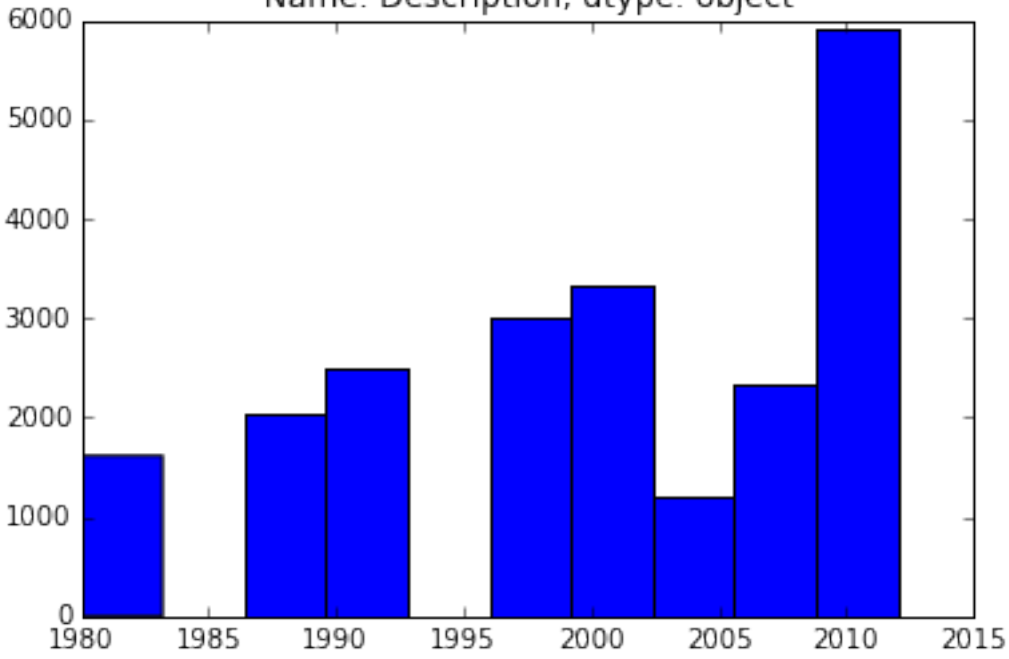

30 GROUP THERMOMETER: Womens Libbers
Name: Description, dtype: object
4500
4000
3500
3000
2500
2000
1500
1000
500
0
1970
1975
1980
1985
1990
1995
2000

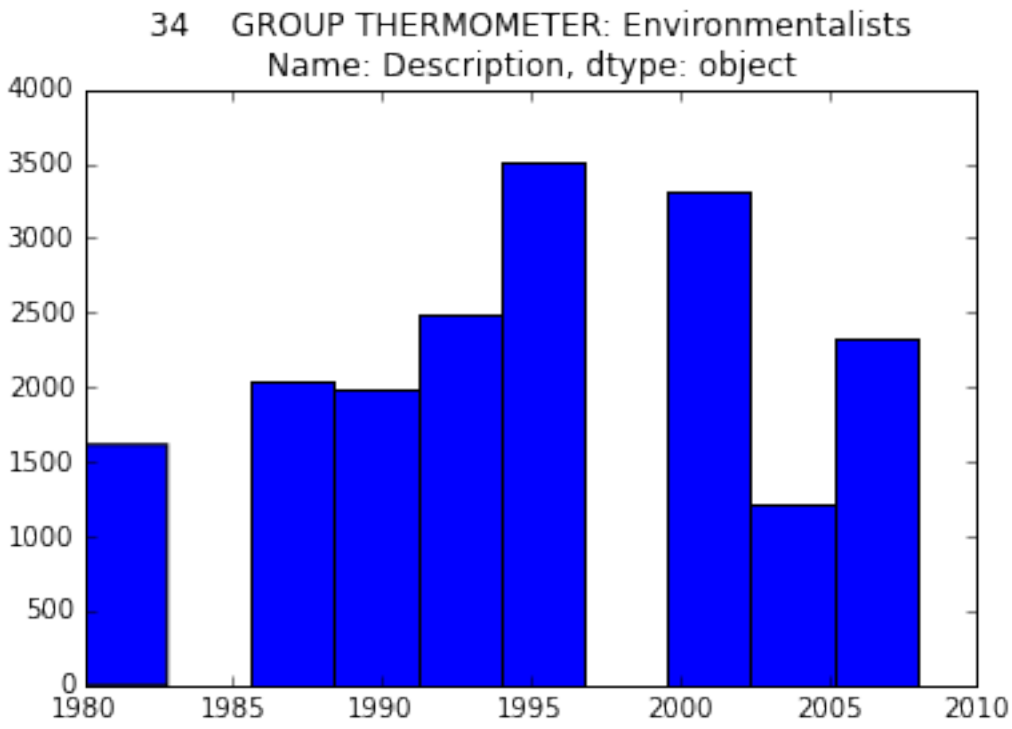

31 GROUP THERMOMETER: Young People
Name: Description, dtype: object
4500
4000
3500
3000
2500
2000
1500
1000
500
0
1970
1975
1980
1985
1990
1995
2000
2005

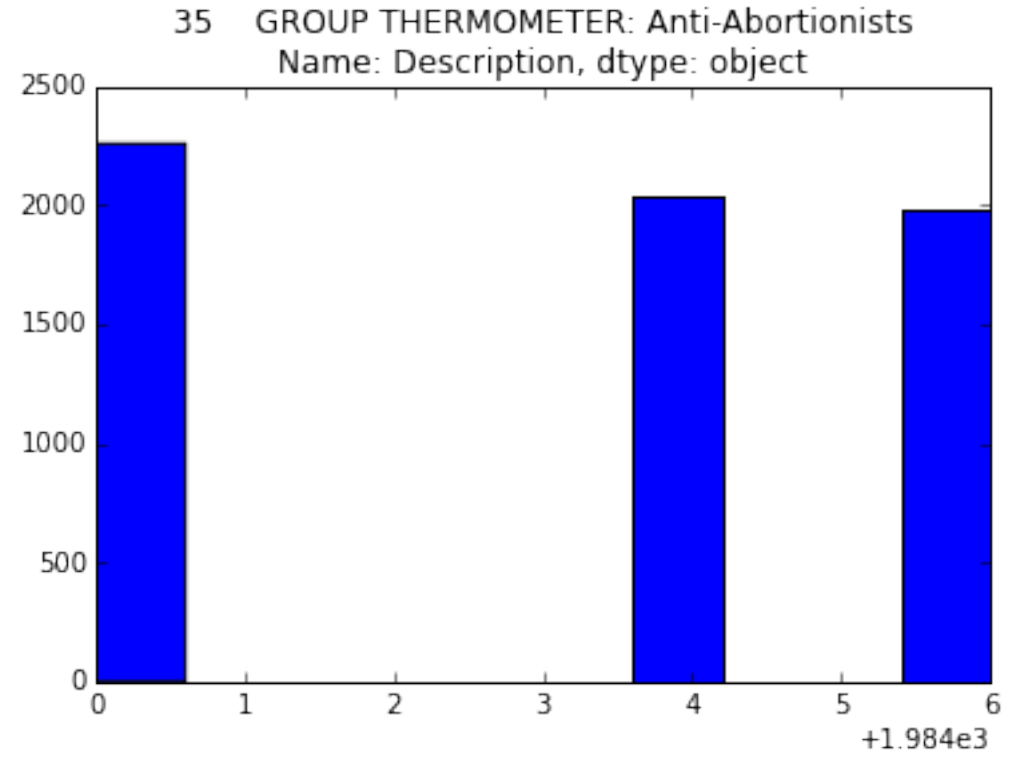
32 GROUP THERMOMETER: Asian-Americans
Name: Description, dtype: object

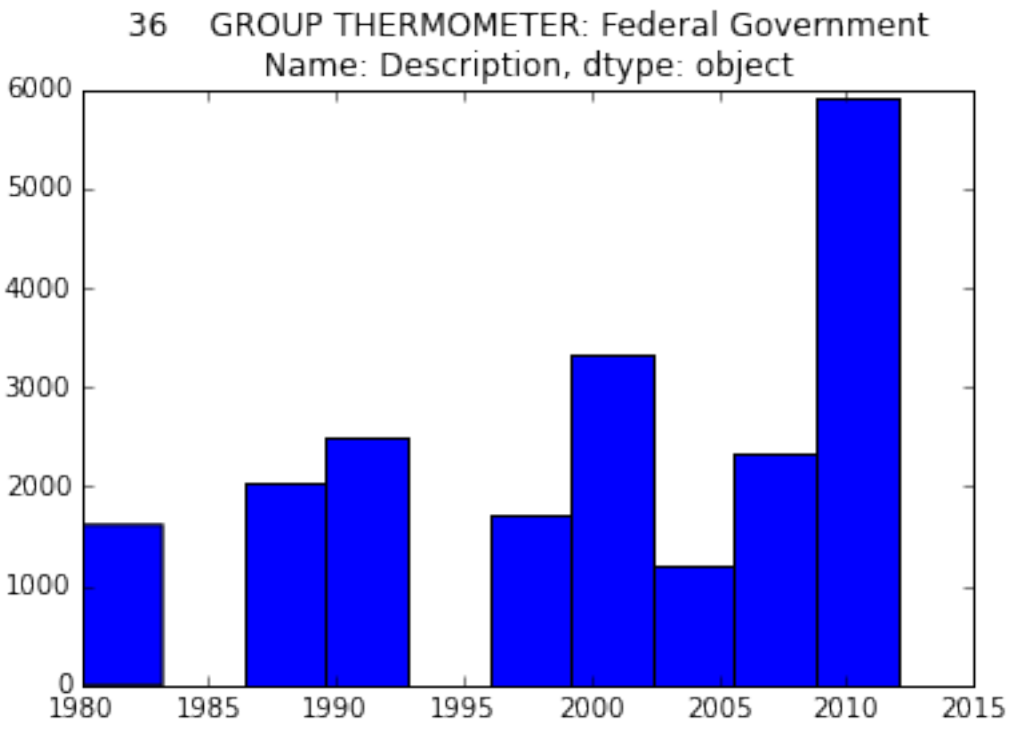


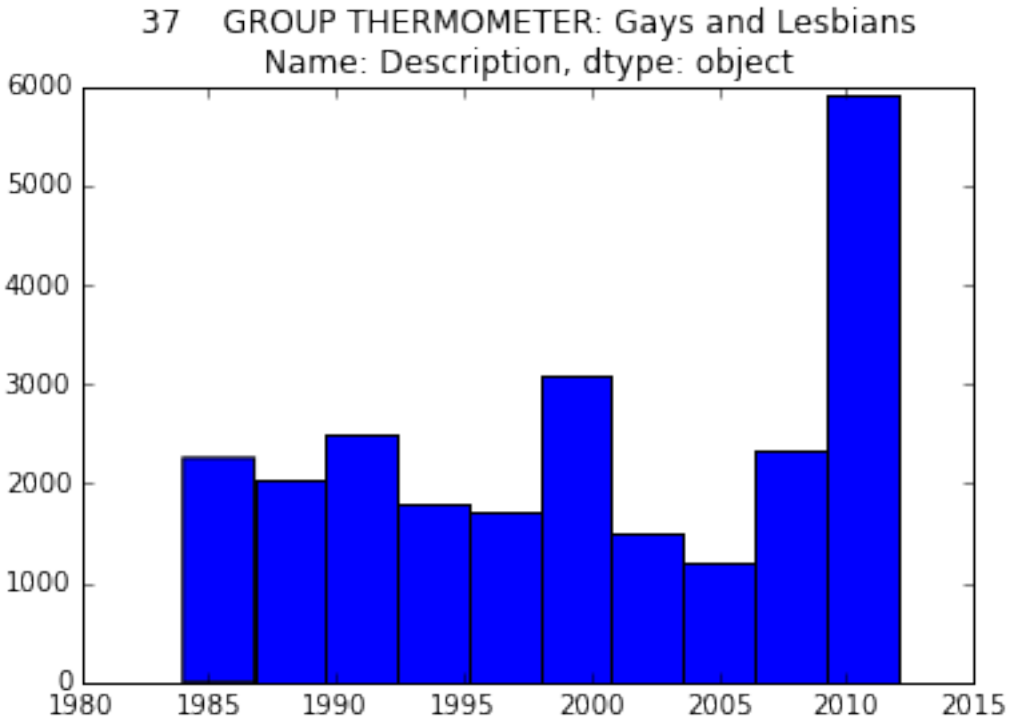
33 GROUP THERMOMETER: Congress
Name: Description, dtype: object

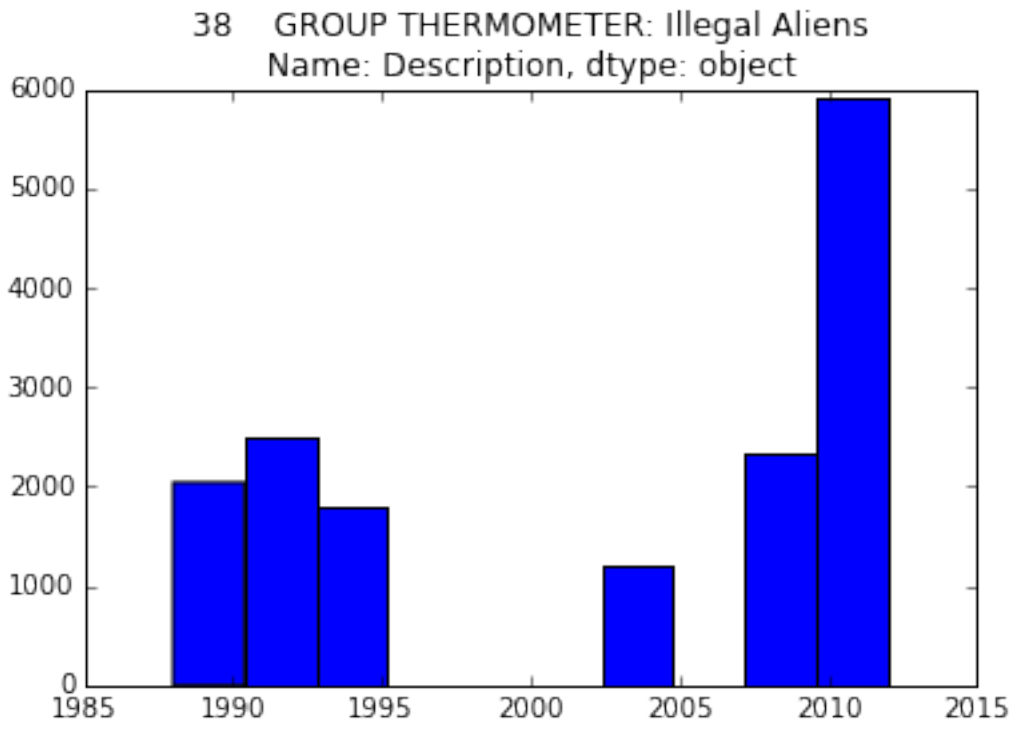


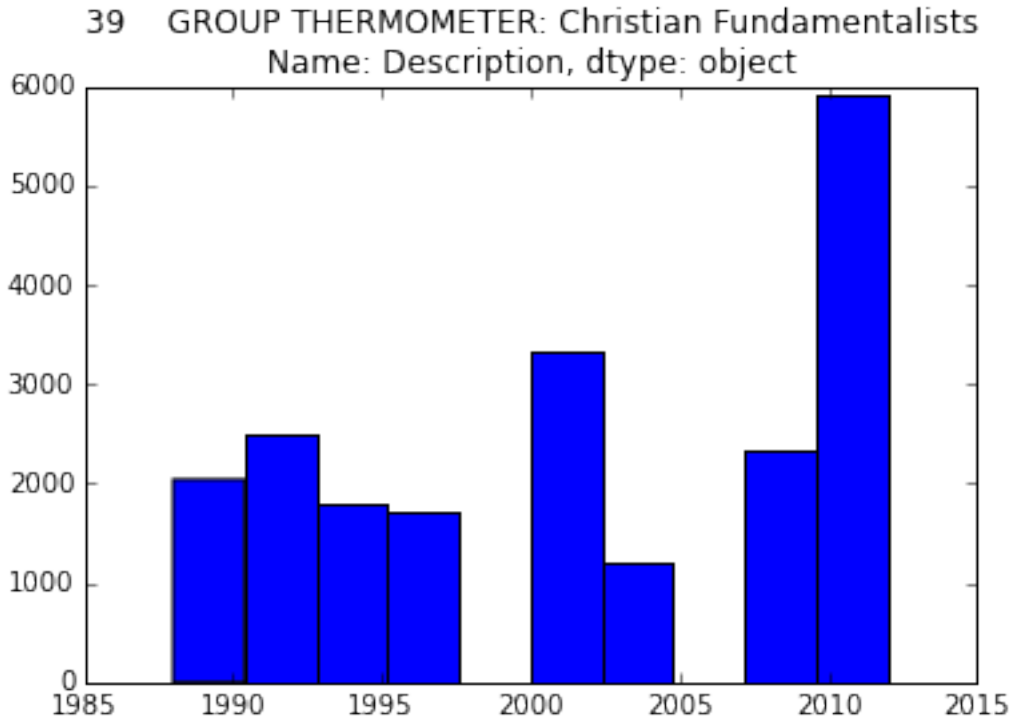

34 GROUP THERMOMETER: Environmentalists
Name: Description, dtype: object
4000
3500
3000
2500
2000
1500
1000
500
0
1980
1985
1990
1995
2000
2005
2010


35 GROUP THERMOMETER: Anti-Abortionists
Name: Description, dtype: object
2500
2000
1500
1000
500
0
0
1
2
3
4
5
6
+1.984e3

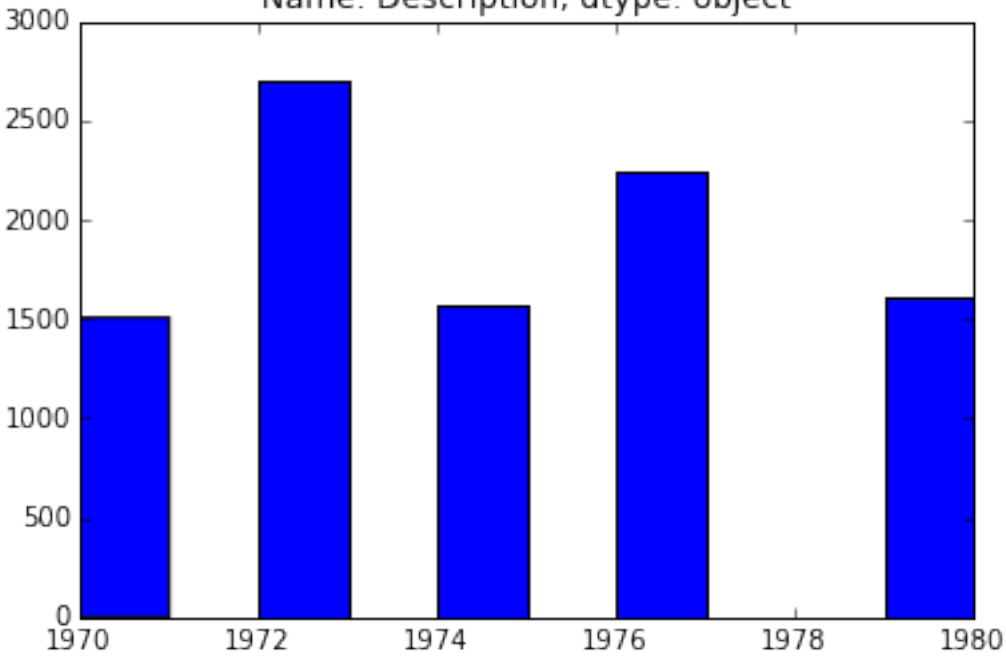

36 GROUP THERMOMETER: Federal Government
Name: Description, dtype: object
6000
5000
4000
3000
2000
1000
0
1980
1985
1990
1995
2000
2005
2010
2015


37 GROUP THERMOMETER: Gays and Lesbians
Name: Description, dtype: object
6000
5000
4000
3000
2000
1000
0
1980
1985
1990
1995
2000
2005
2010
2015

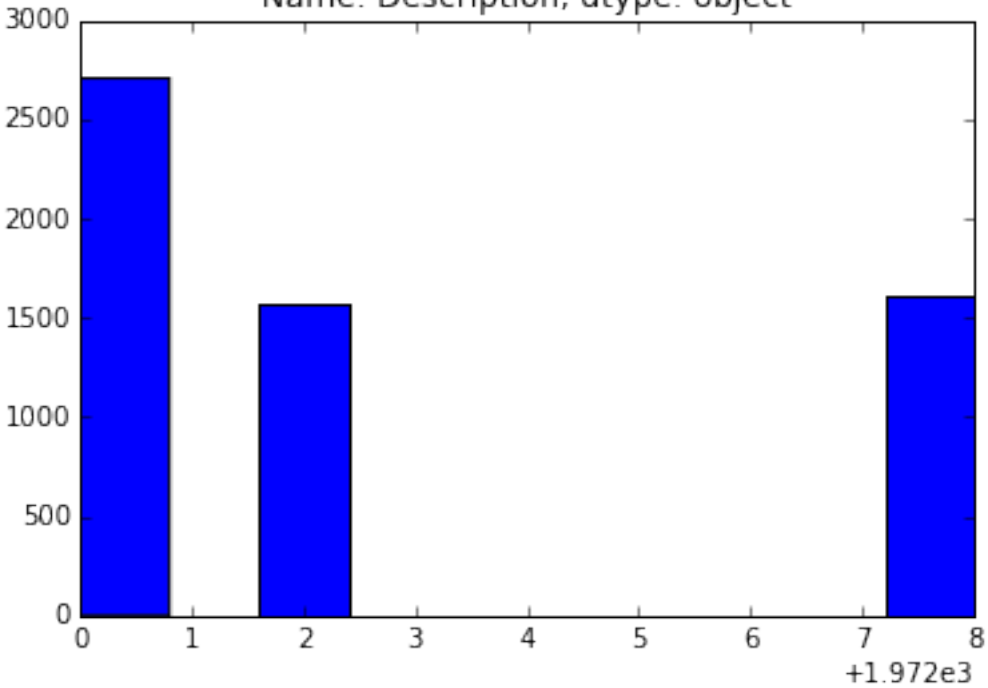

38 GROUP THERMOMETER: Illegal Aliens
Name: Description, dtype: object
6000
5000
4000
3000
2000
1000
0
1985
1990
1995
2000
2005
2010
2015

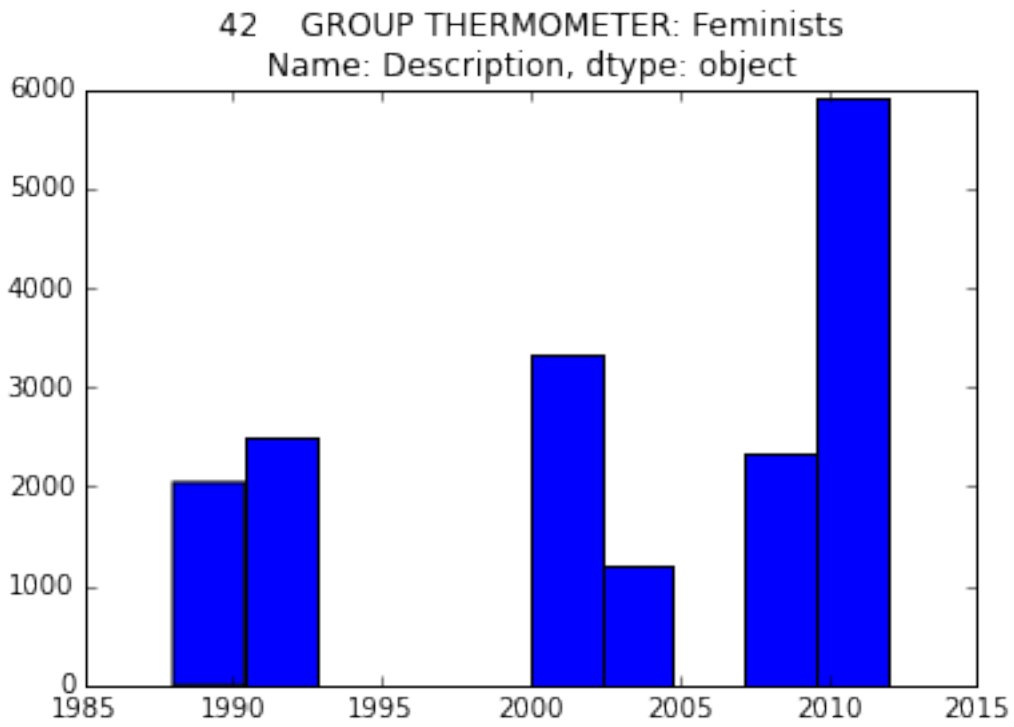

39 GROUP THERMOMETER: Christian Fundamentalists
Name: Description, dtype: object
6000
5000
4000
3000
2000
1000
0
1985
1990
1995
2000
2005
2010
2015

40 GROUP THERMOMETER: Radical Students
Name: Description, dtype: object



41 GROUP THERMOMETER: Farmers
Name: Description, dtype: object

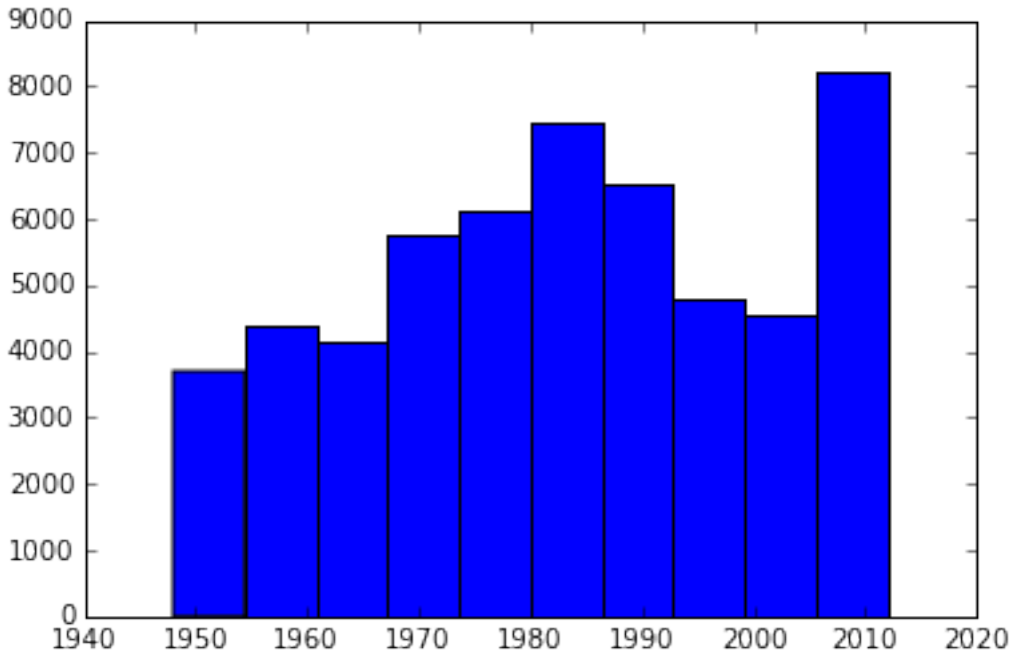


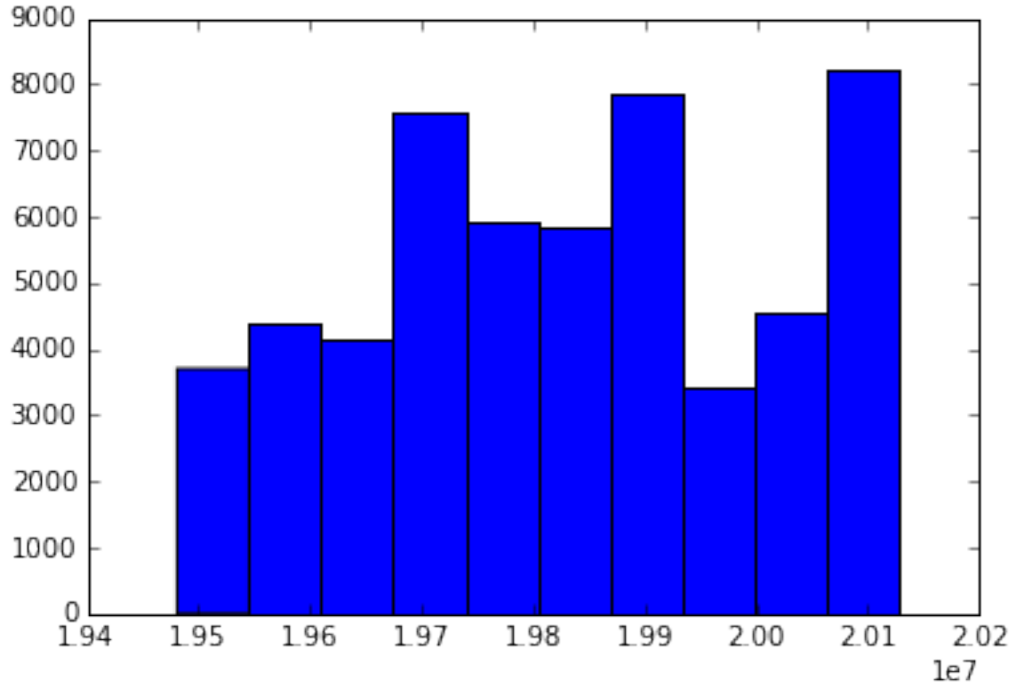


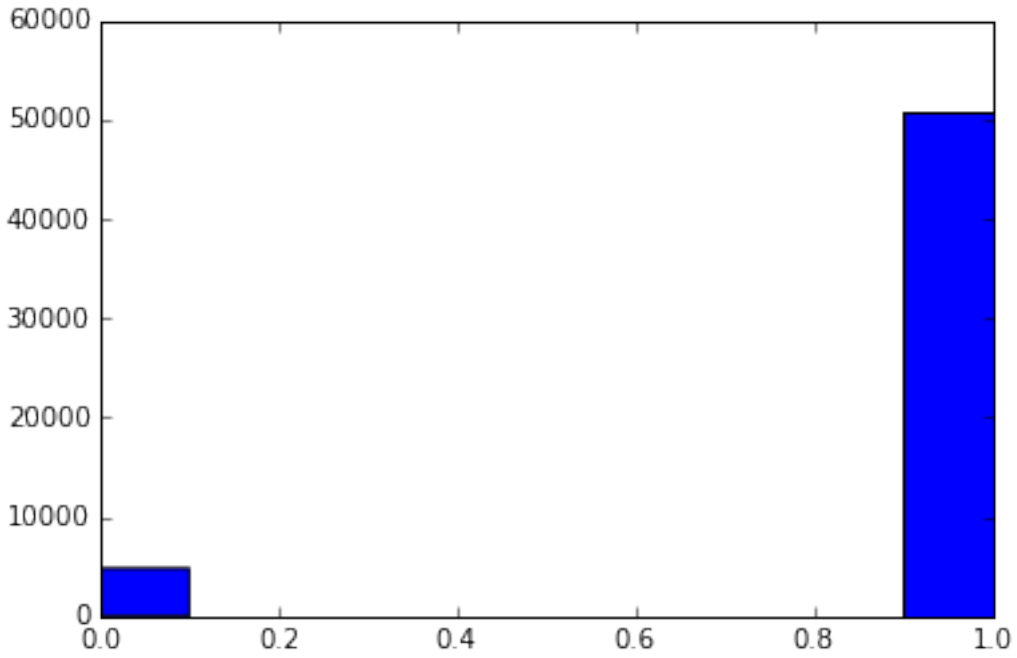
1.3 Variable Histograms

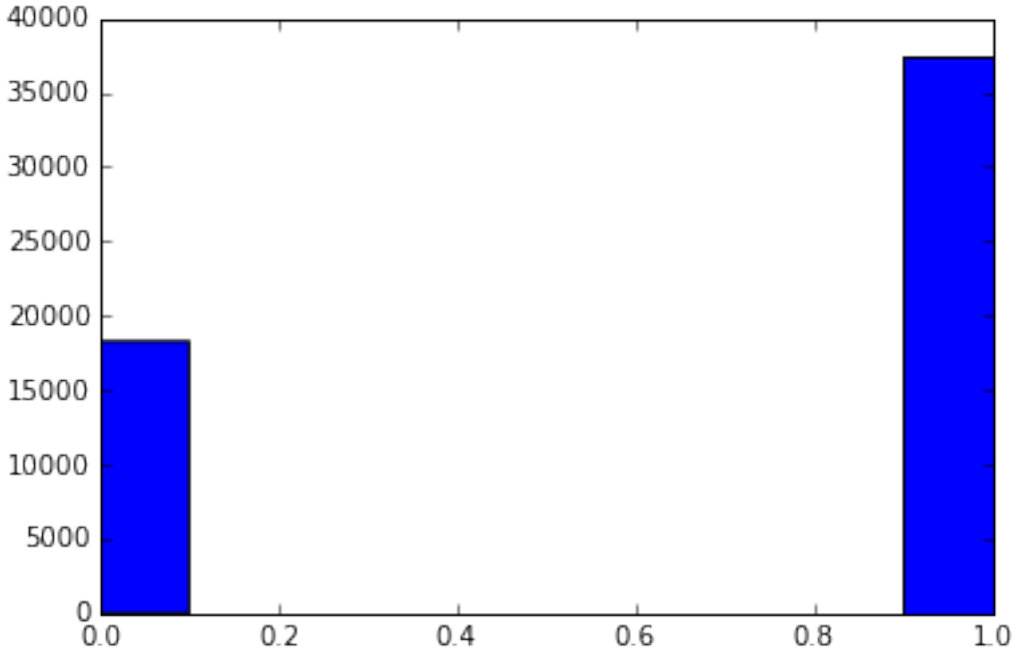
Shows histograms for each survey question.

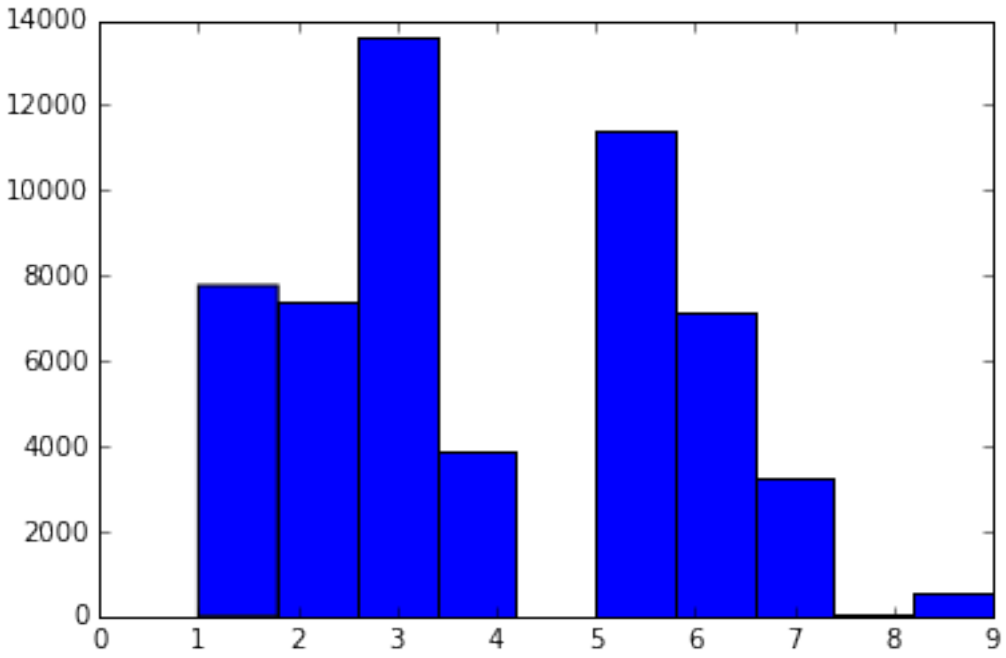
```
In [25]: # generate histograms for each survey question
         for var in dfinfo['VarName']:
             data = df[var][np.logical_not(pd.isnull(df[var]))].values
             plt.hist(data)
             plt.show()
```

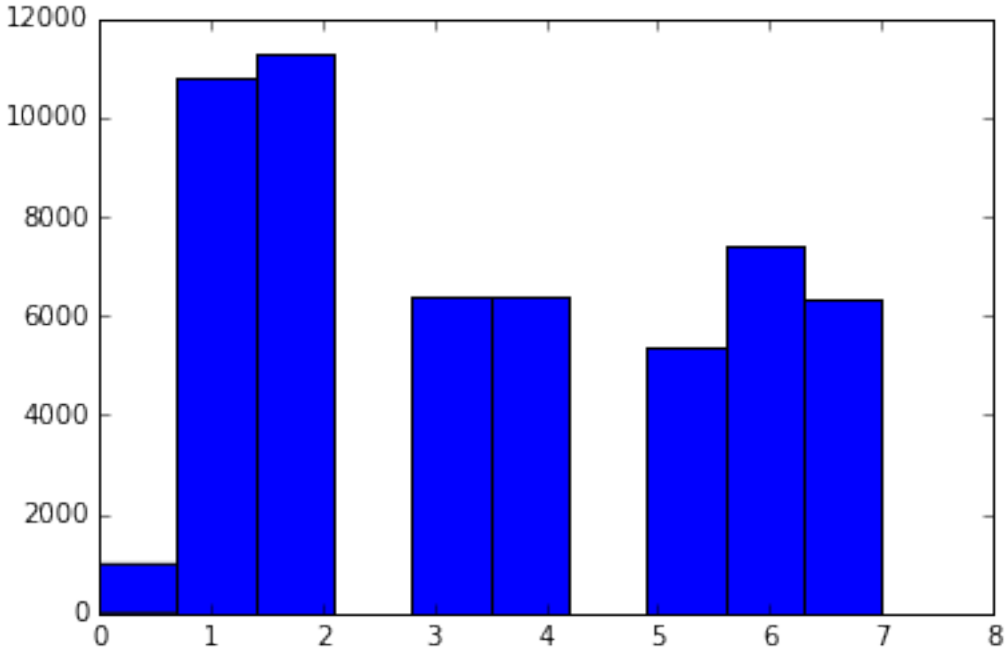

9000
8000
7000
6000
5000
4000
3000
2000
1000
0
1940
1950
1960
1970
1980
1990
2000
2010
2020

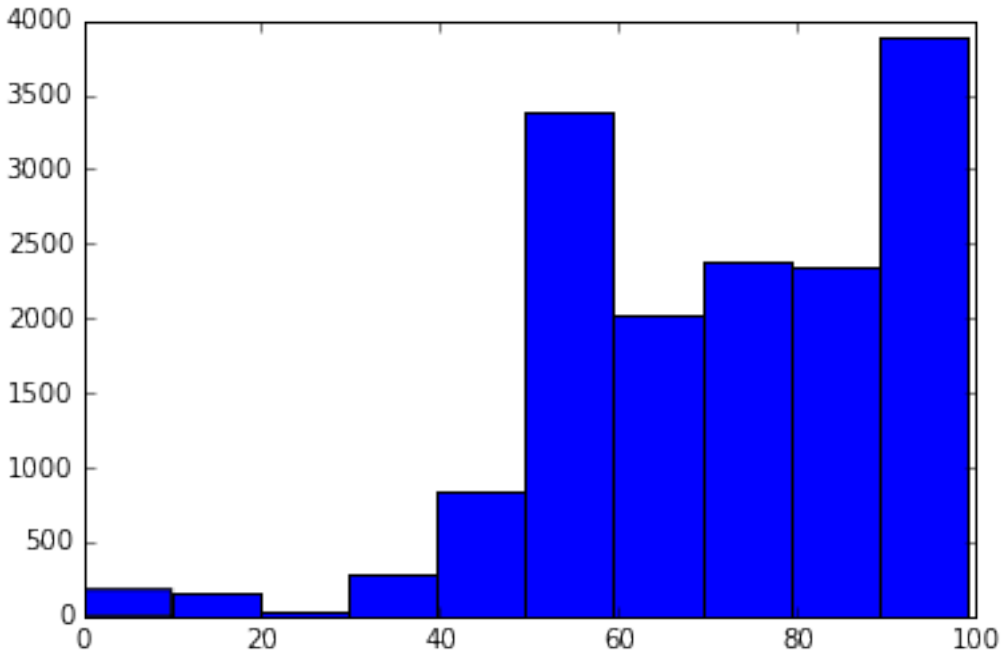

9000
8000
7000
6000
5000
4000
3000
2000
1000
0
1.94
1.95
1.96
1.97
1.98
1.99
2.00
2.01
2.02
1e7

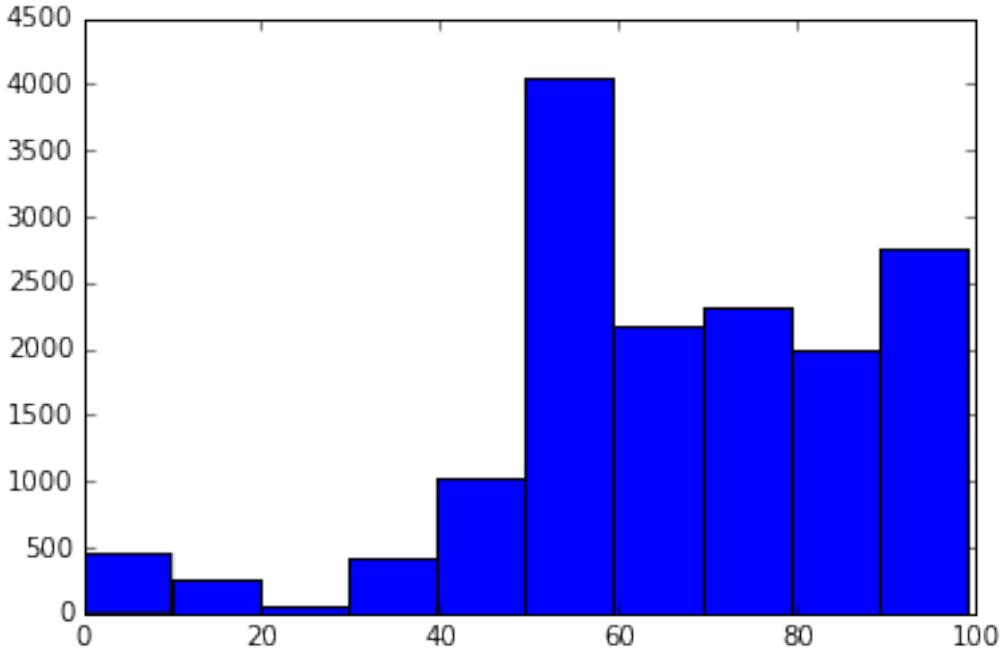

60000
50000
40000
30000
20000
10000
0
0.0
0.2
0.4
0.6
0.8
1.0

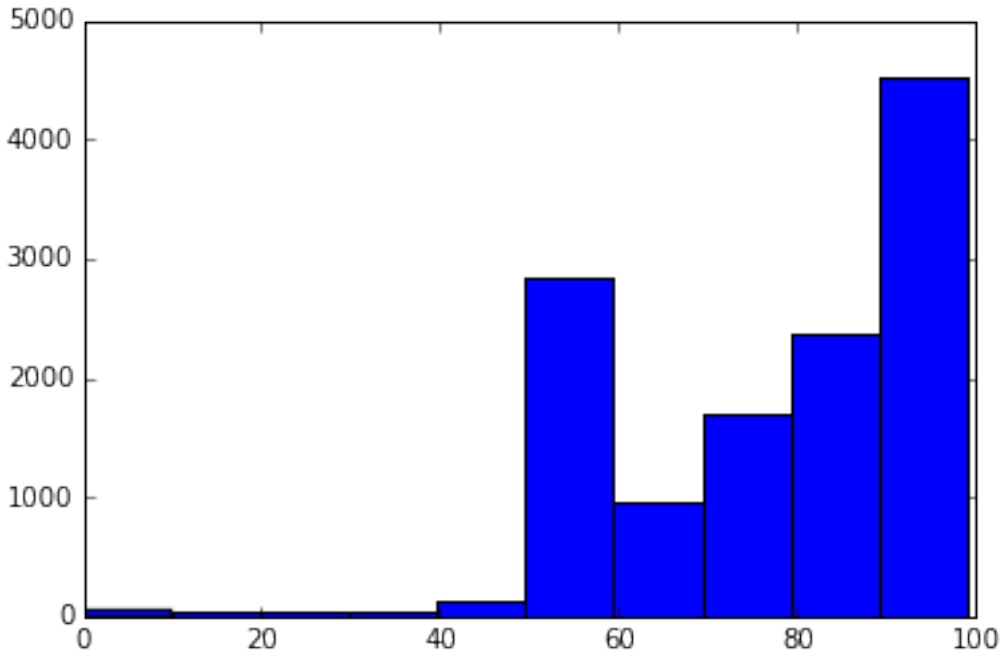

40000
35000
30000
25000
20000
15000
10000
5000
0
0.0
0.2
0.4
0.6
0.8
1.0

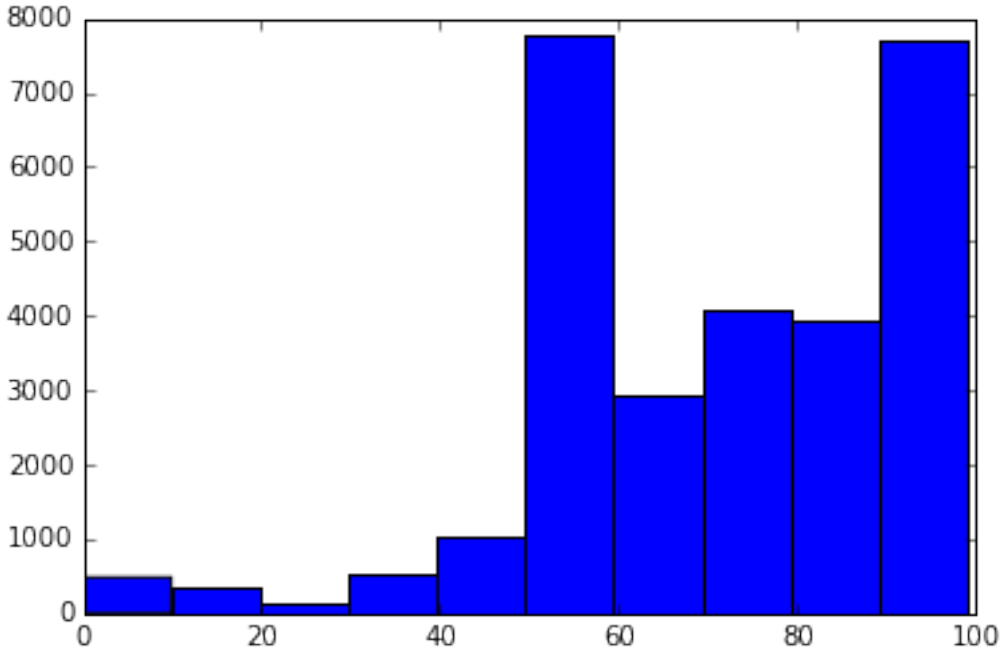

14000
12000
10000
8000
6000
4000
2000
0
0
1
2
3
4
5
6
7
8
9

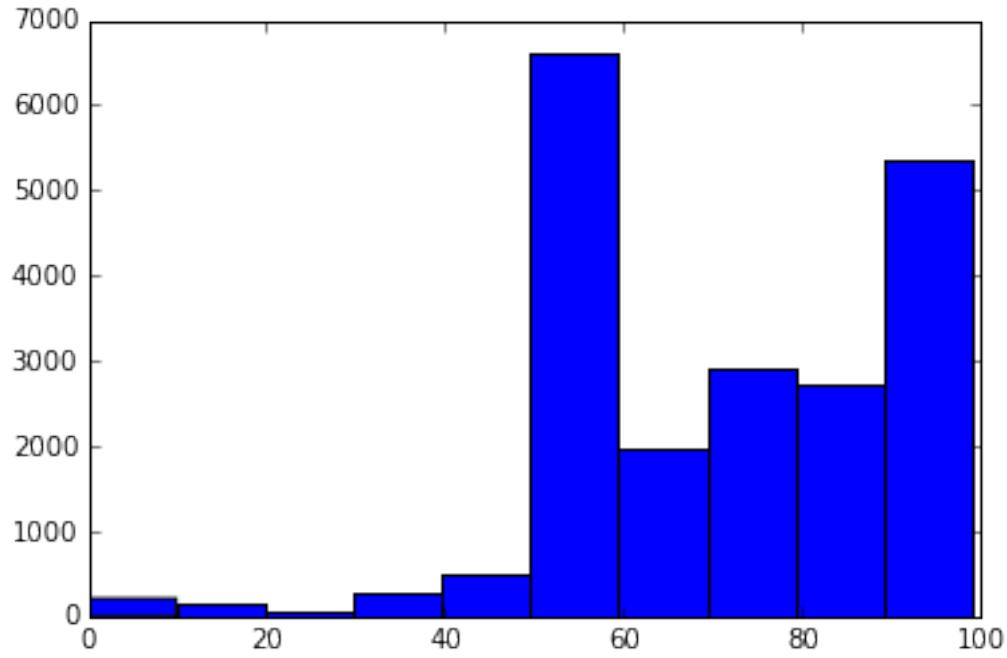

12000
10000
8000
6000
4000
2000
0
0
1
2
3
4
5
6
7
8

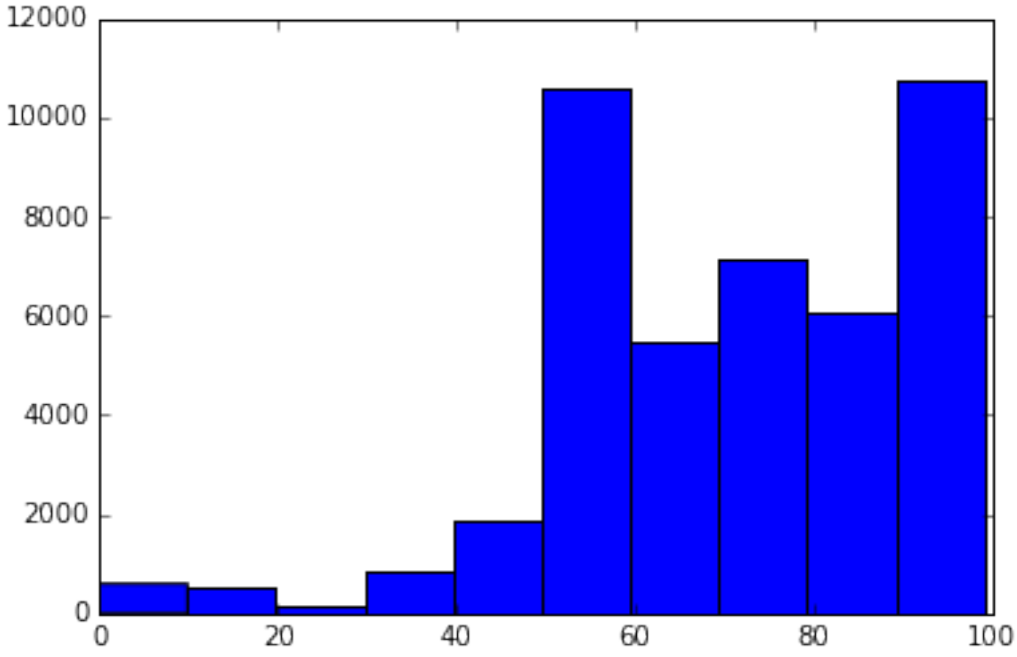

4000
3500
3000
2500
2000
1500
1000
500
0
0
20
40
60
80
100

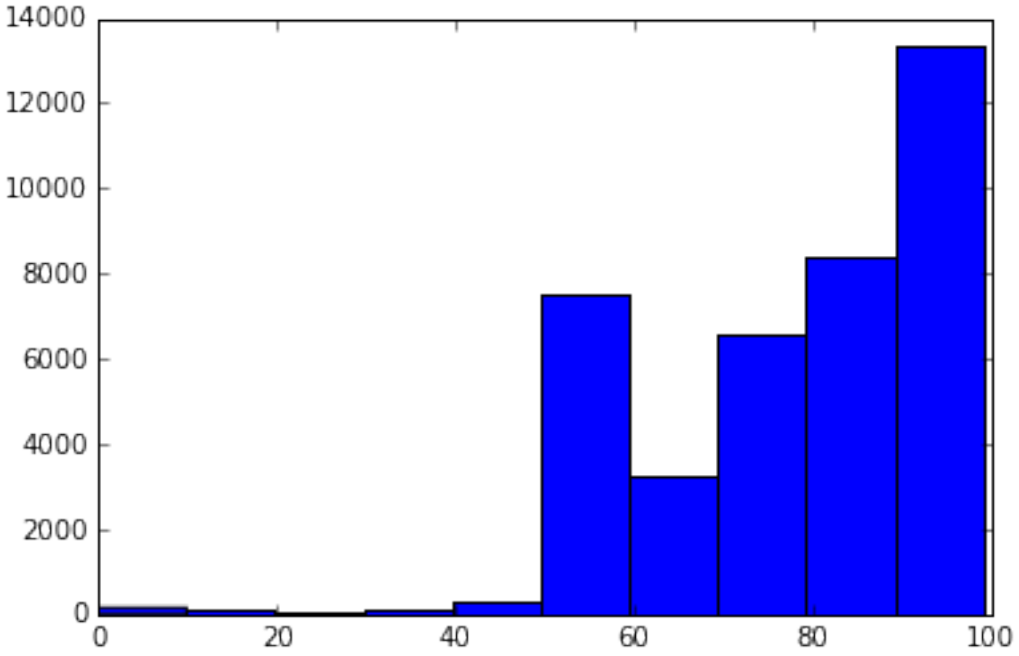

4500
4000
3500
3000
2500
2000
1500
1000
500
0
0
20
40
60
80
100

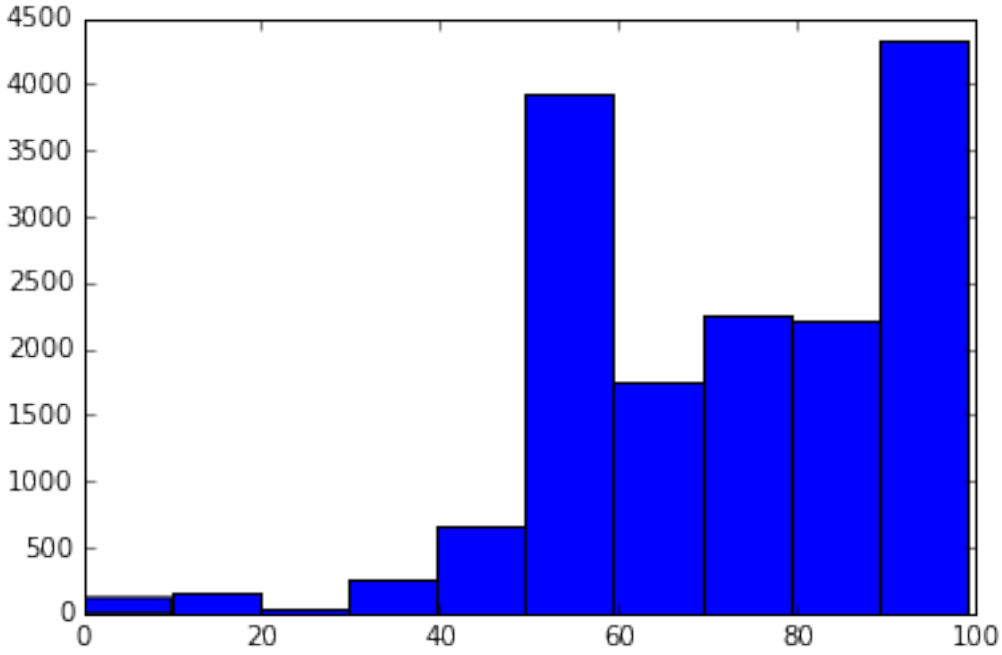

5000
4000
3000
2000
1000
0
0
20
40
60
80
100

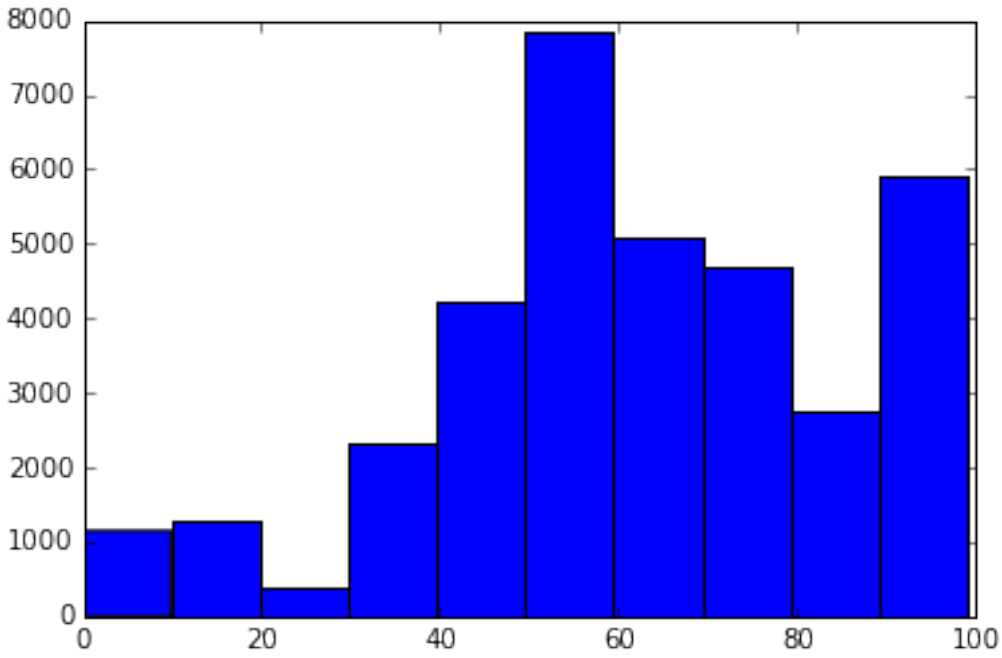

8000
7000
6000
5000
4000
3000
2000
1000
0
0
20
40
60
80
100

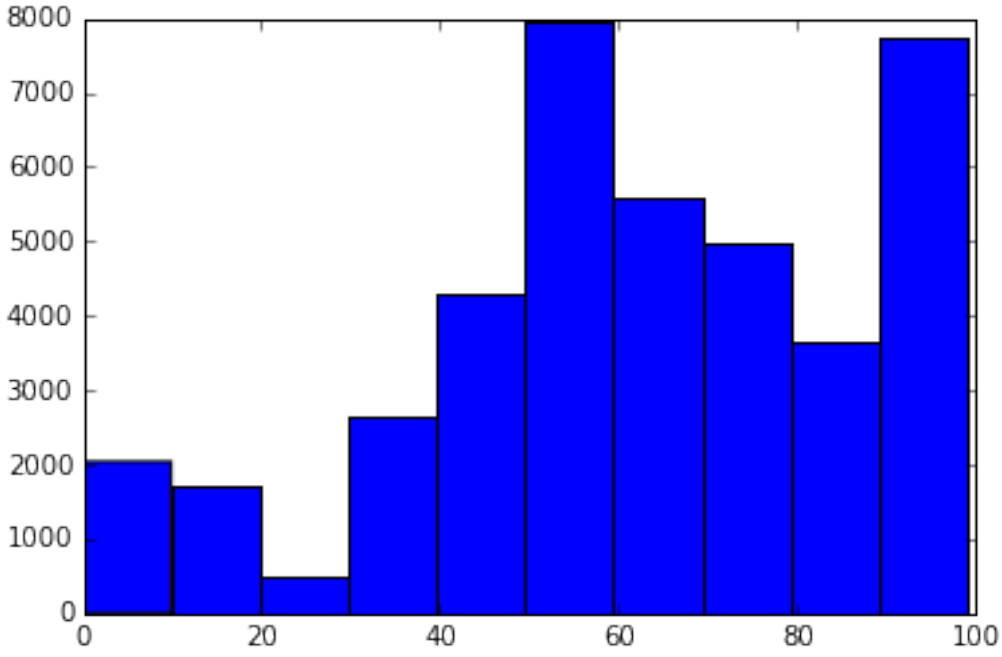

7000
6000
5000
4000
3000
2000
1000
0
0
20
40
60
80
100

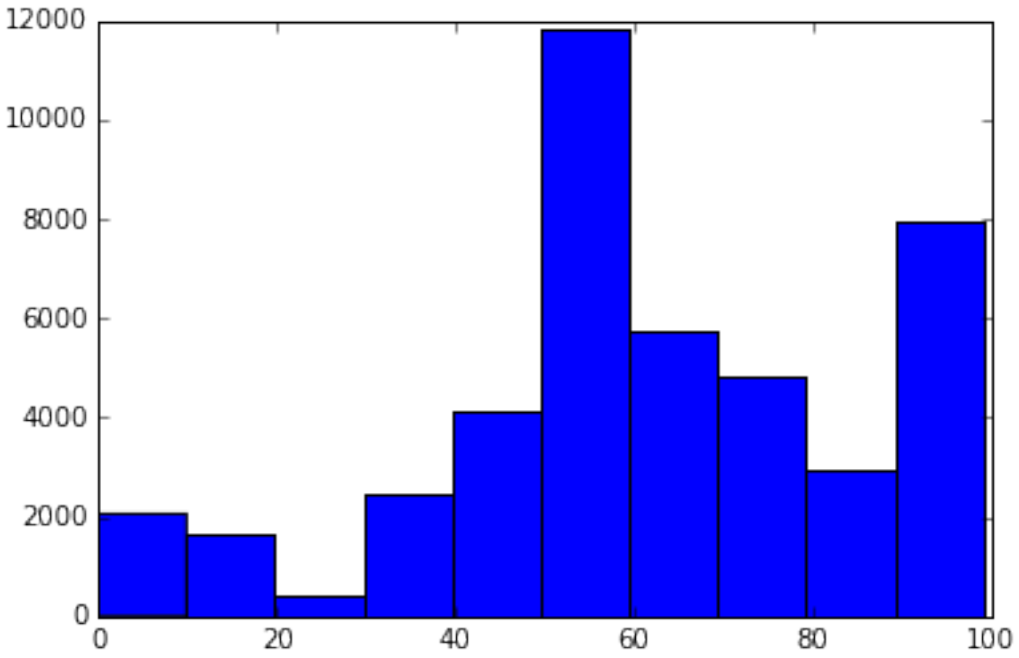

12000
10000
8000
6000
4000
2000
0
0
20
40
60
80
100

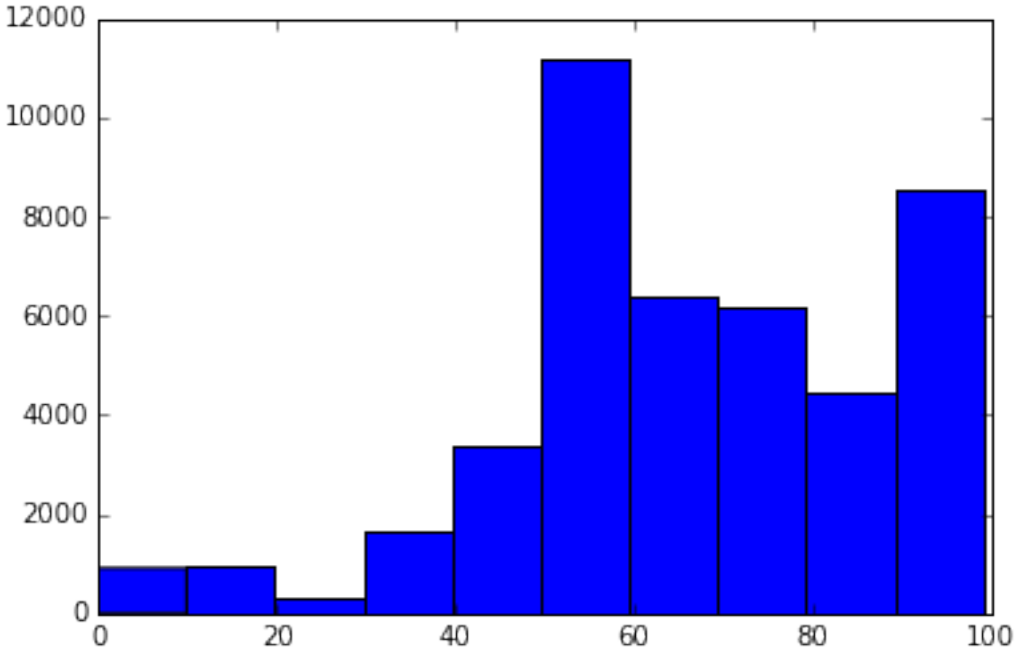

14000
12000
10000
8000
6000
4000
2000
0
0
20
40
60
80
100

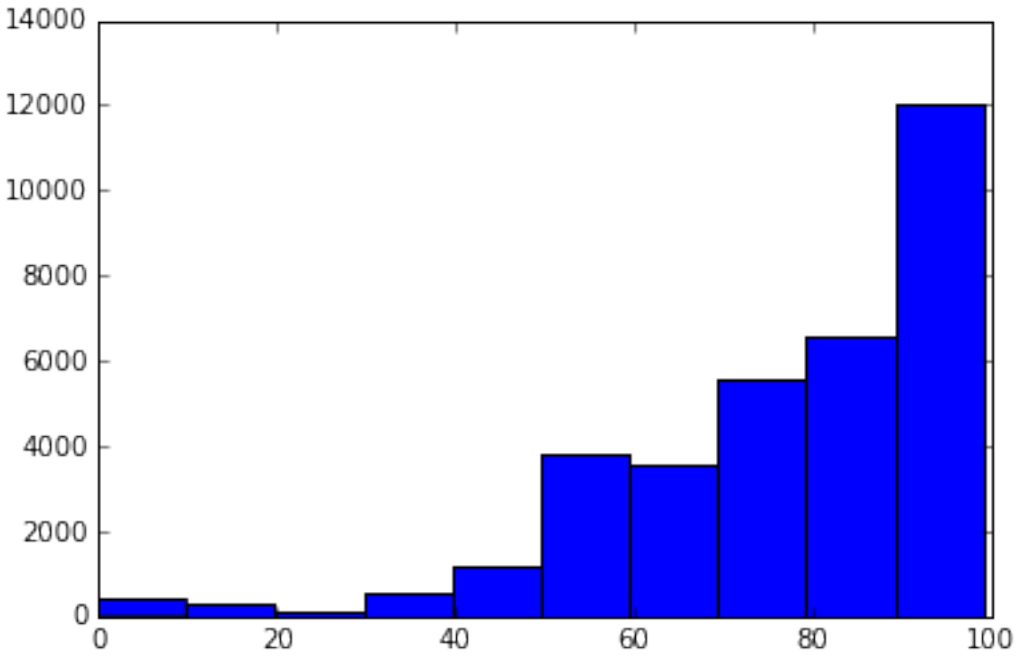

4500
4000
3500
3000
2500
2000
1500
1000
500
0
0
20
40
60
80
100

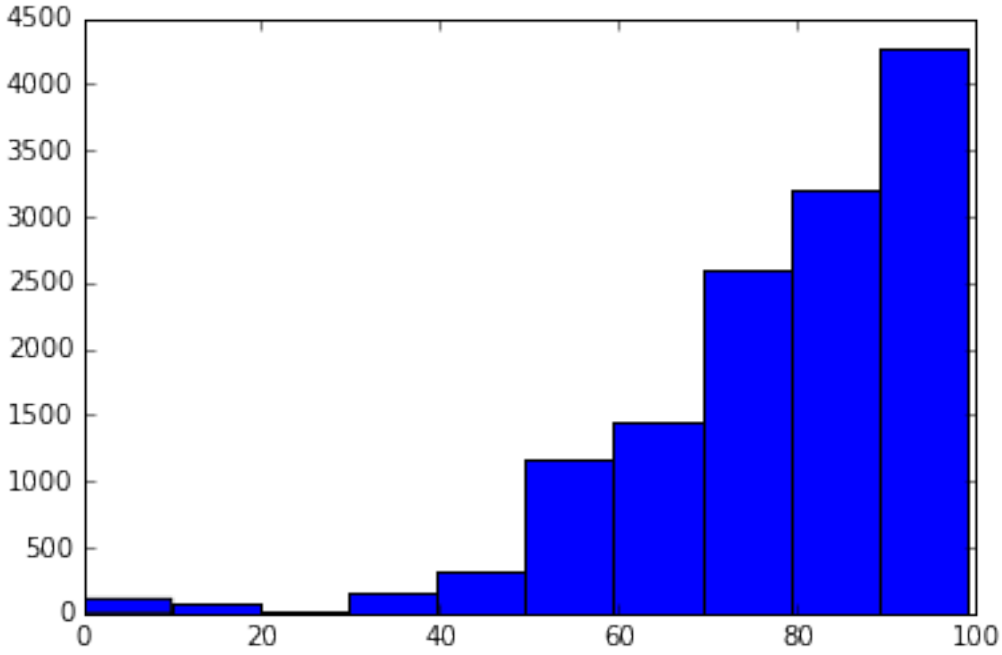

8000
7000
6000
5000
4000
3000
2000
1000
0
0
20
40
60
80
100

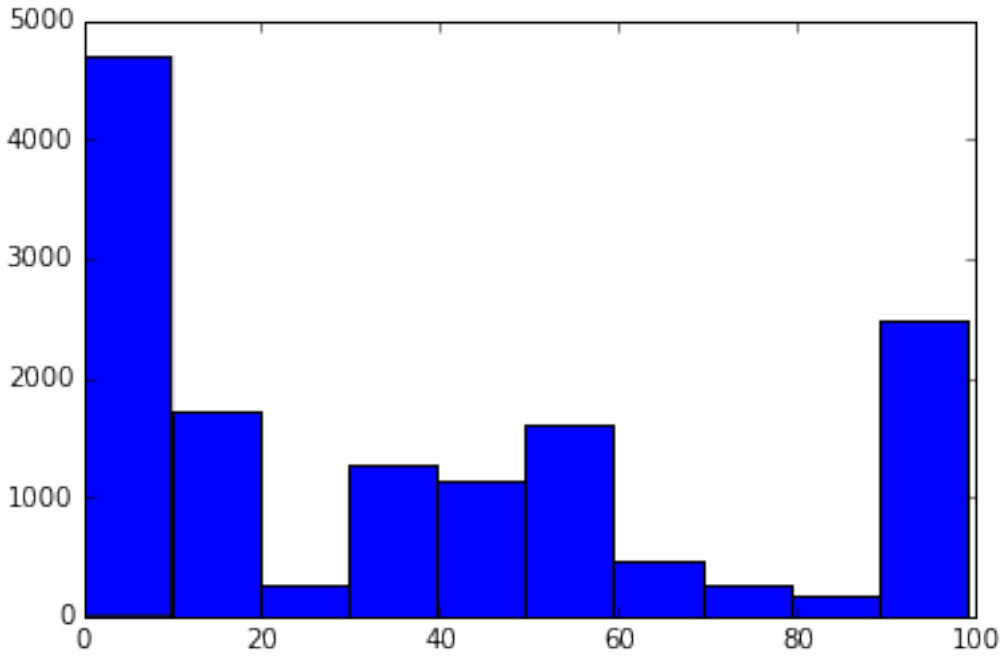

8000
7000
6000
5000
4000
3000
2000
1000
0
0
20
40
60
80
100

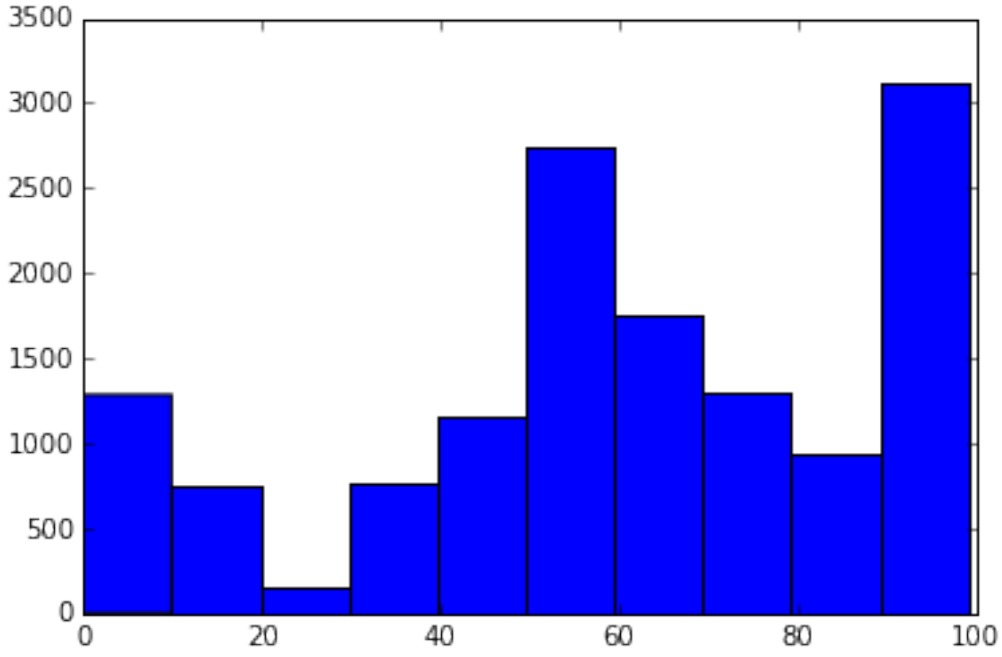

12000
10000
8000
6000
4000
2000
0
0
20
40
60
80
100

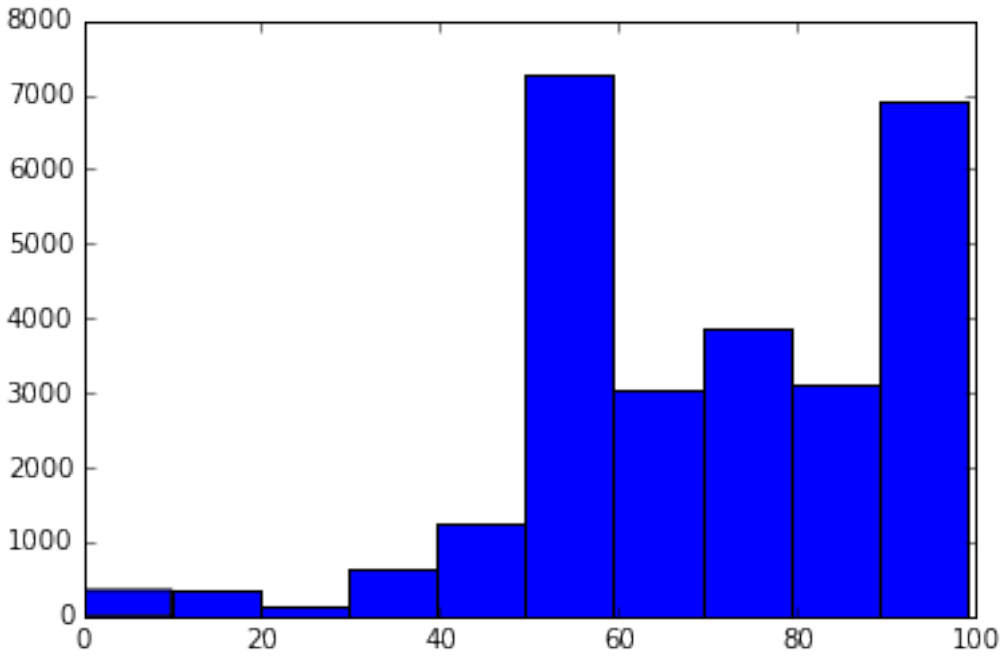

12000
10000
8000
6000
4000
2000
0
0
20
40
60
80
100

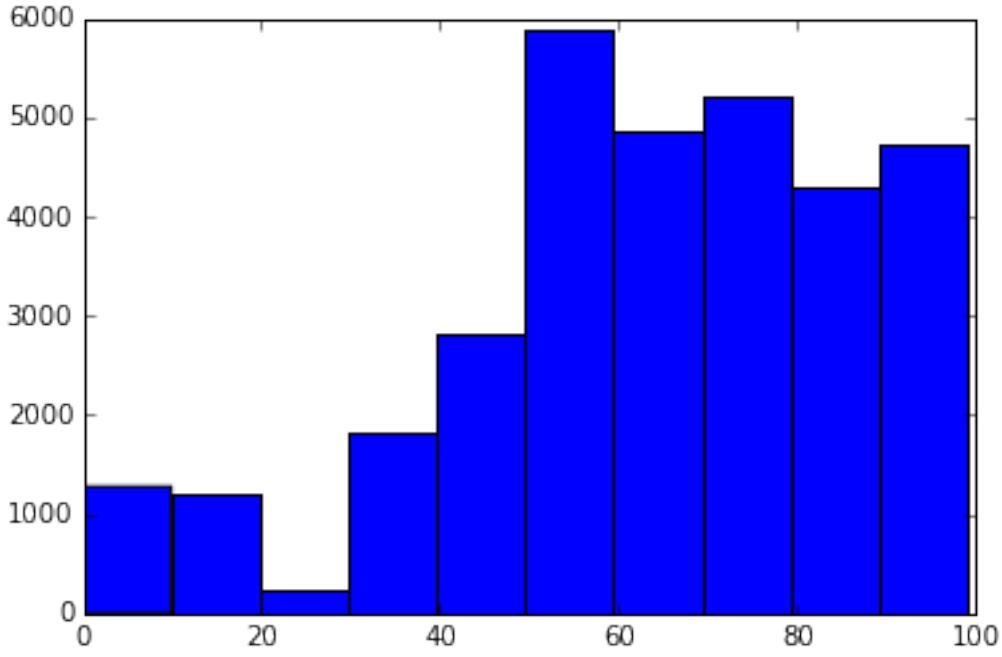

14000
12000
10000
8000
6000
4000
2000
0
0
20
40
60
80
100

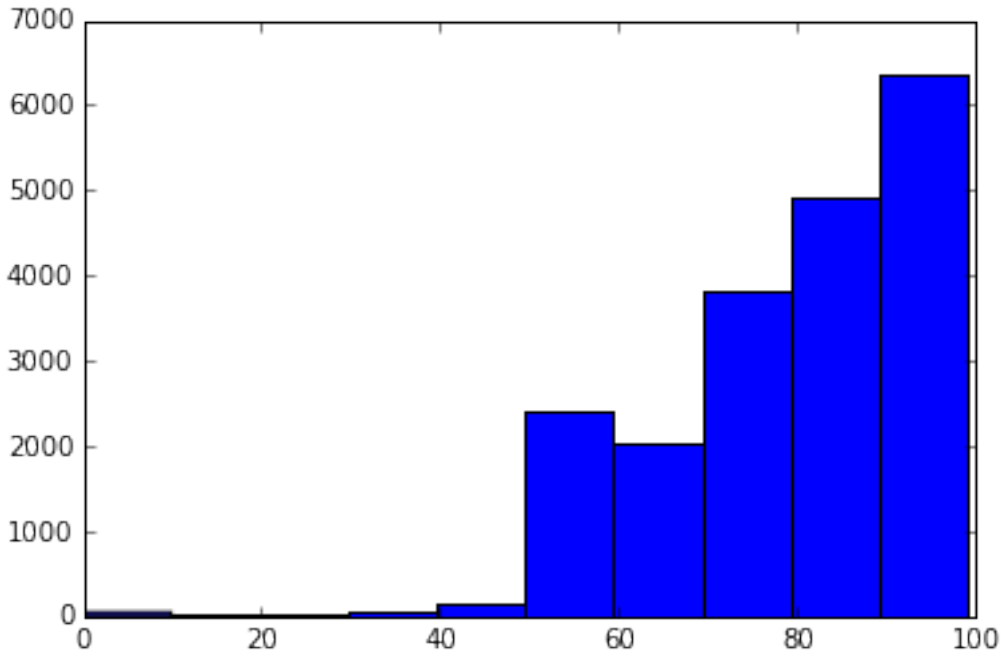

4500
4000
3500
3000
2500
2000
1500
1000
500
0
0
20
40
60
80
100

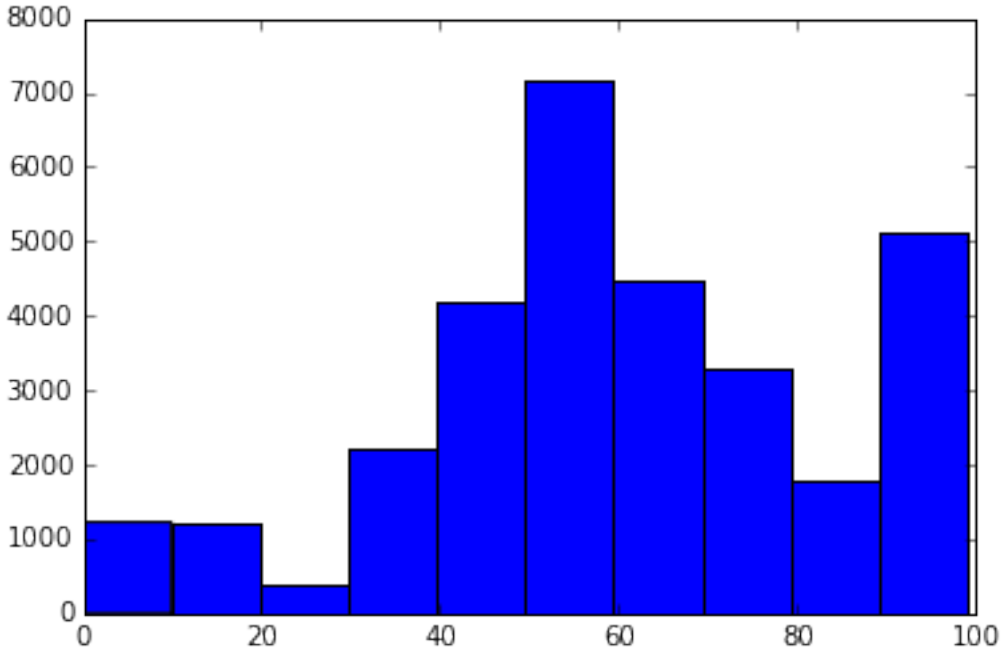

5000
4000
3000
2000
1000
0
0
20
40
60
80
100

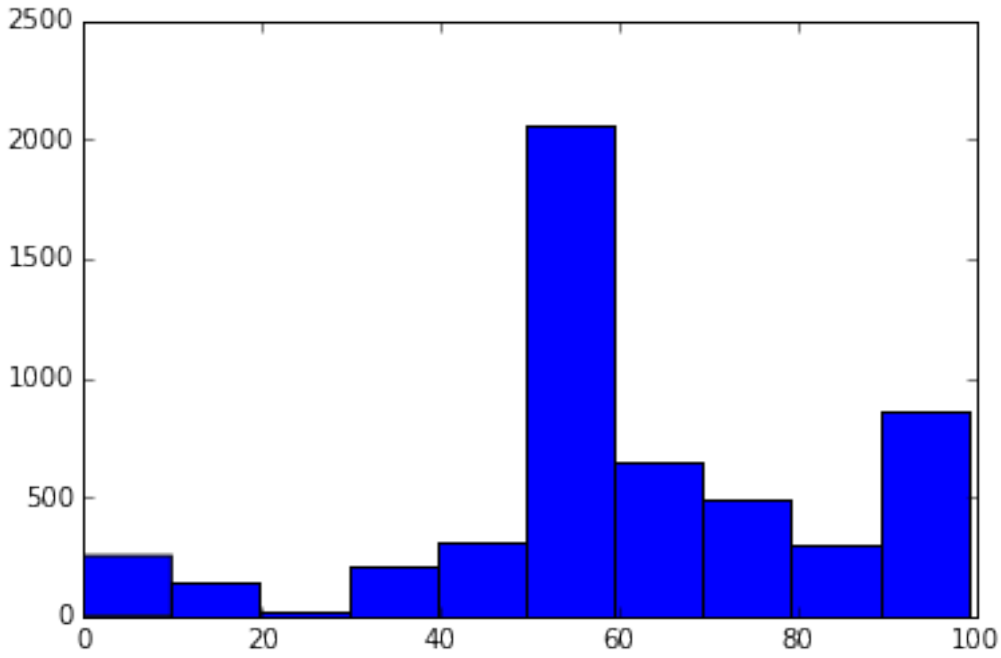

3500
3000
2500
2000
1500
1000
500
0
0
20
40
60
80
100

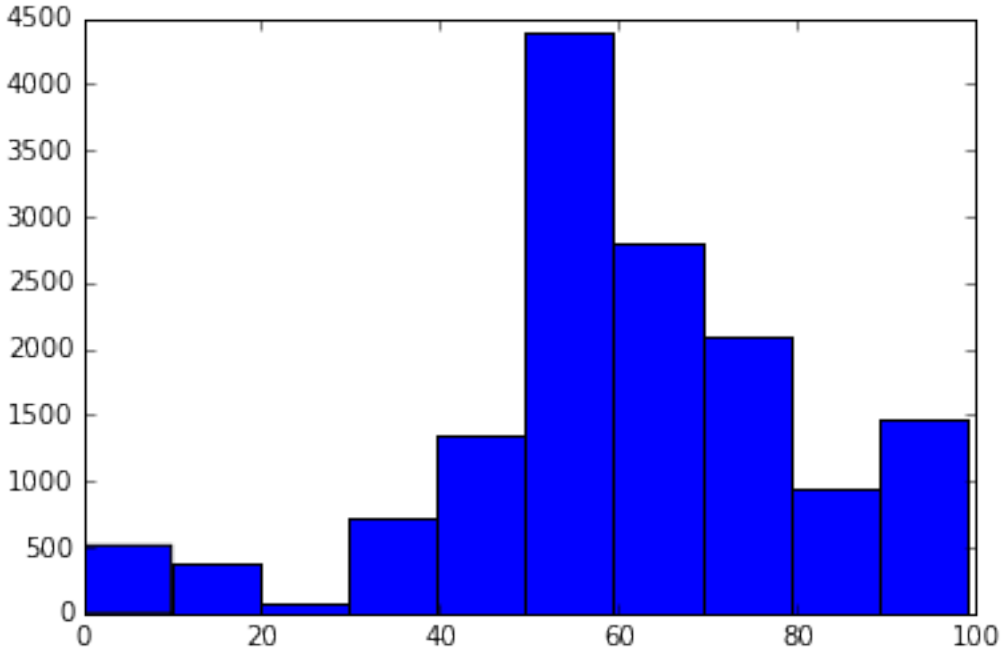

8000
7000
6000
5000
4000
3000
2000
1000
0
0
20
40
60
80
100

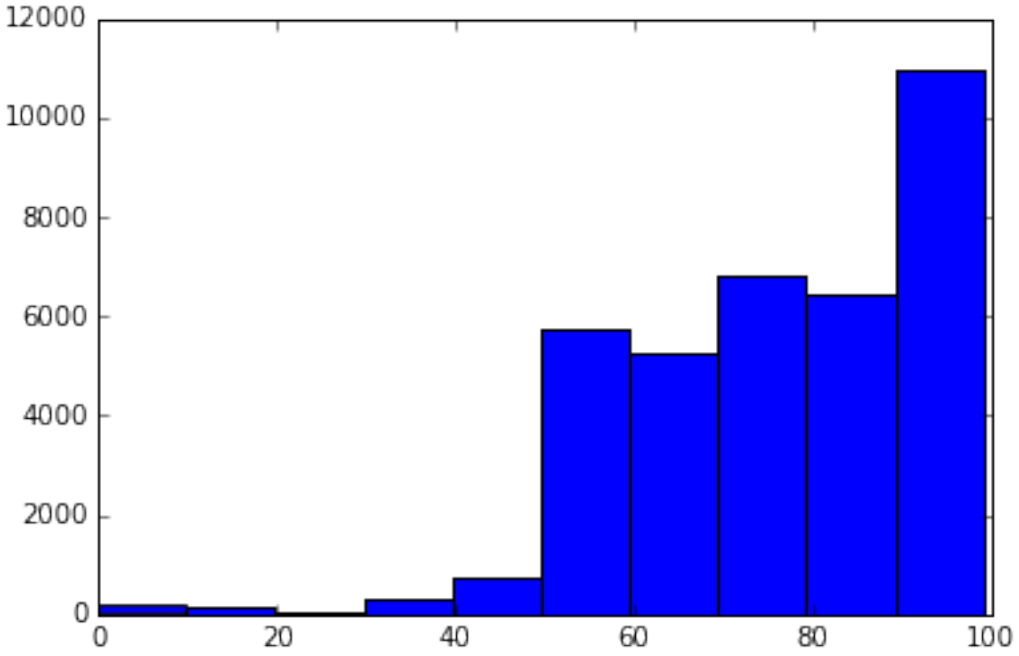

6000
5000
4000
3000
2000
1000
0
0
20
40
60
80
100

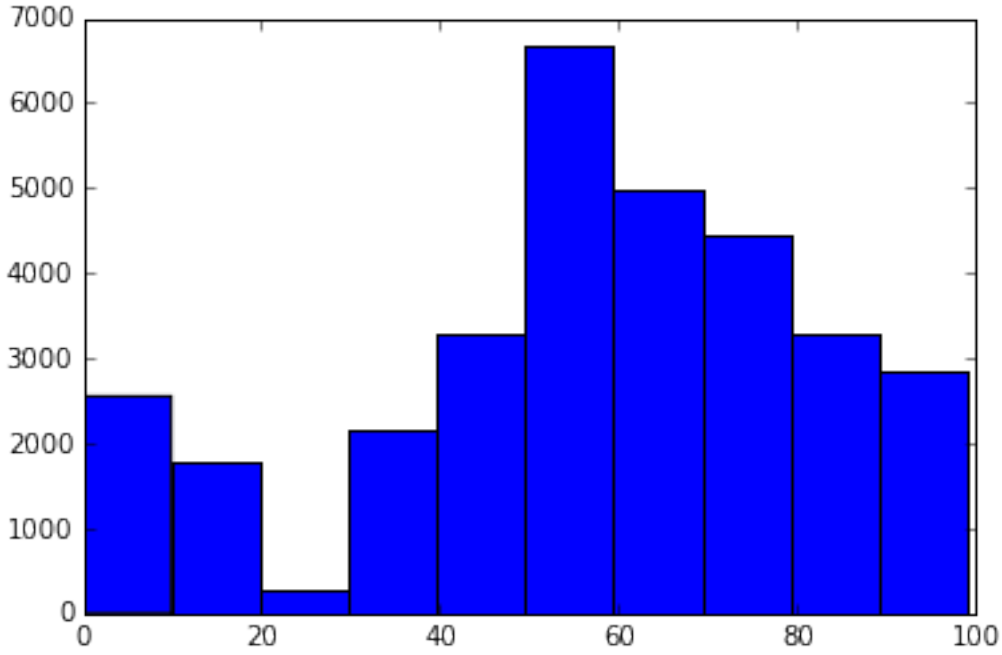

7000
6000
5000
4000
3000
2000
1000
0
0
20
40
60
80
100

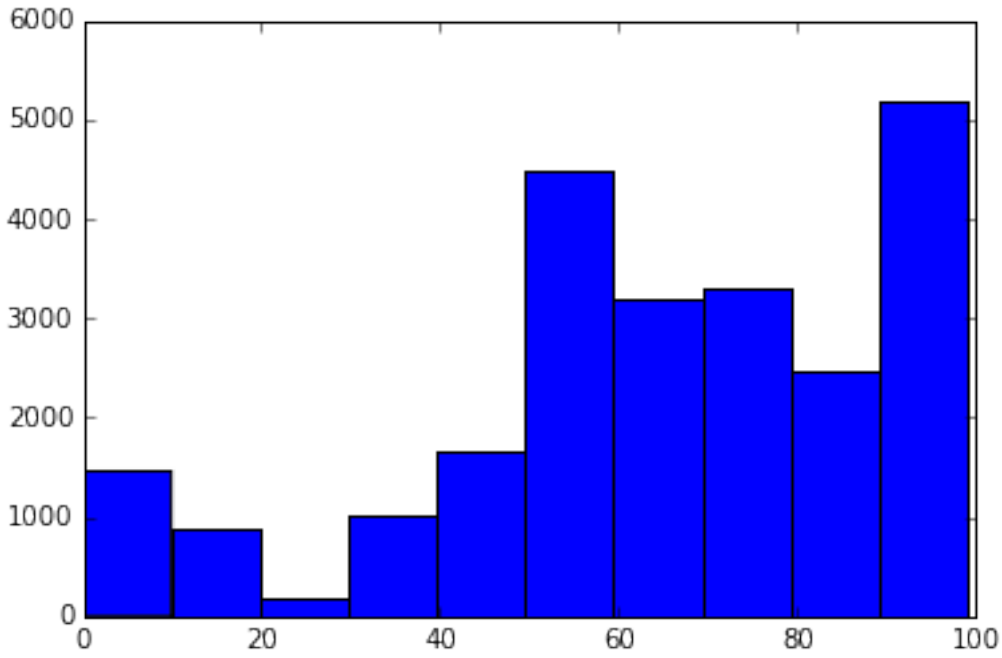

8000
7000
6000
5000
4000
3000
2000
1000
0
0
20
40
60
80
100

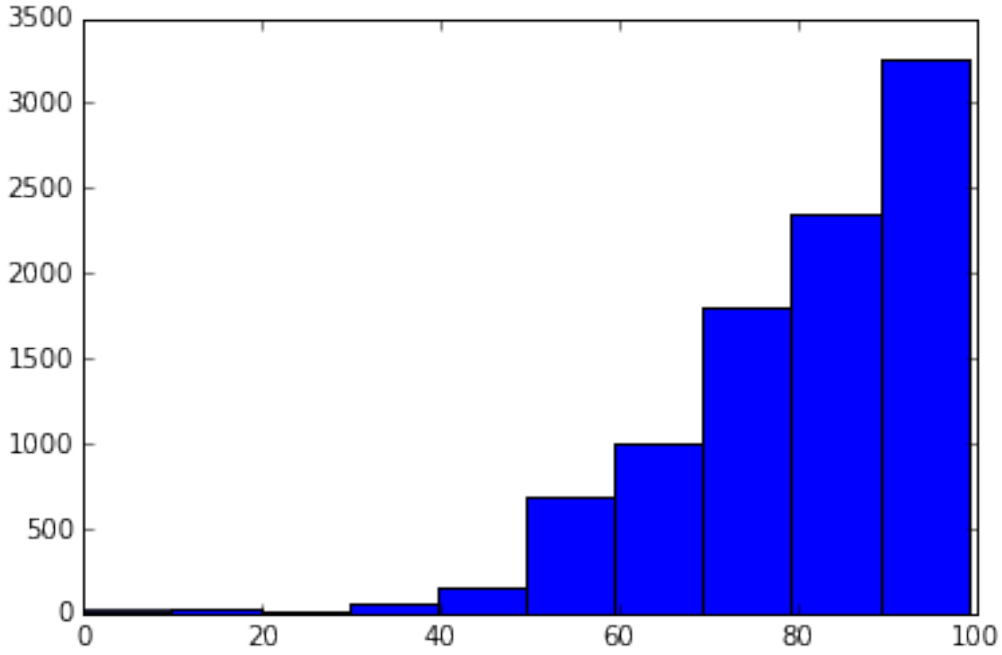

2500
2000
1500
1000
500
0
0
20
40
60
80
100

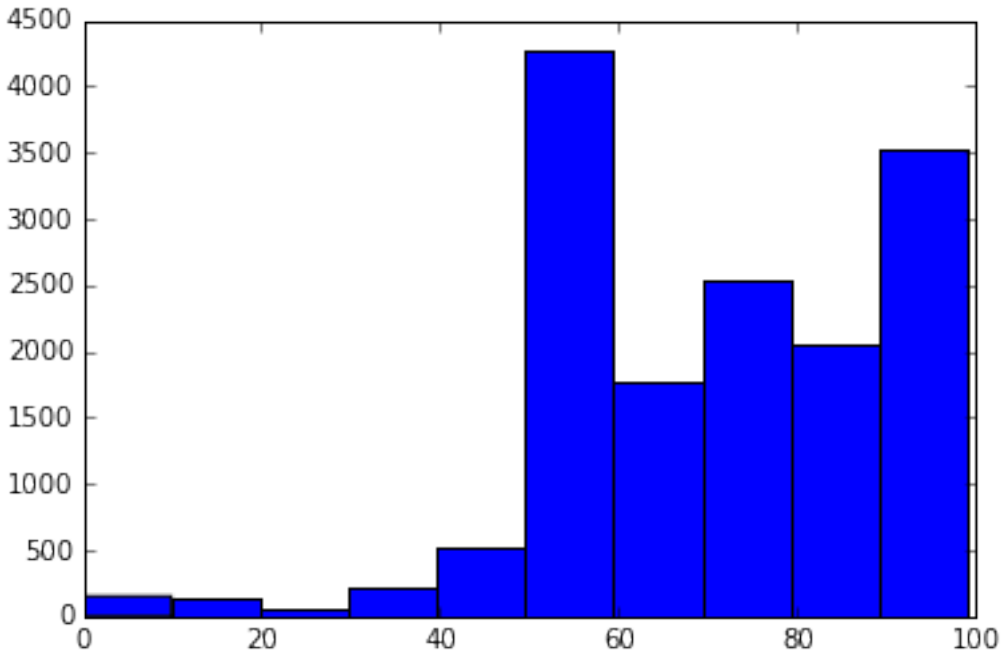

4500
4000
3500
3000
2500
2000
1500
1000
500
0
0
20
40
60
80
100

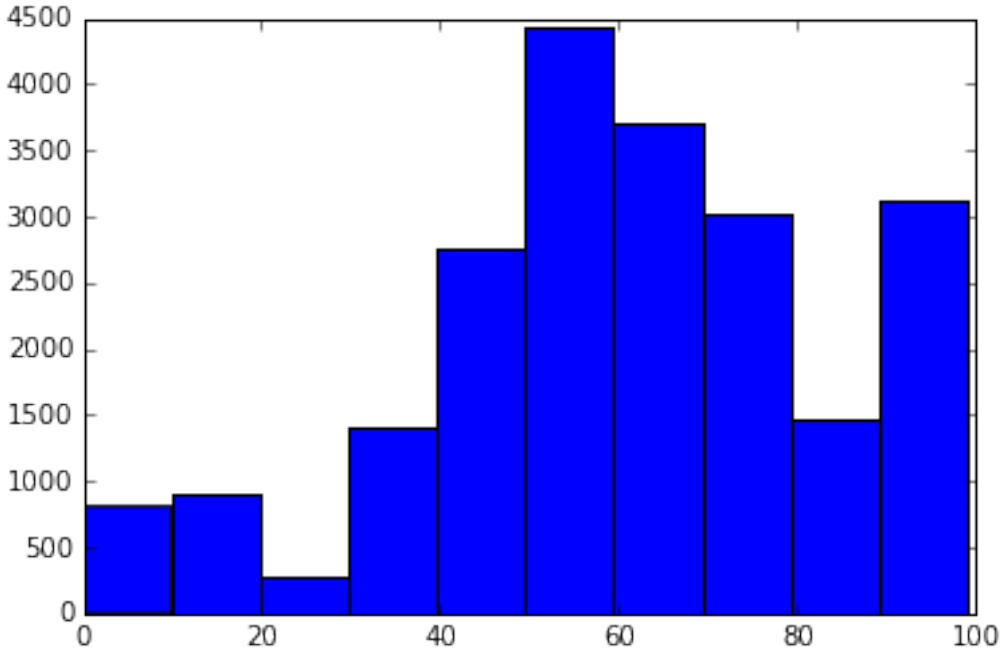

12000
10000
8000
6000
4000
2000
0
0
20
40
60
80
100

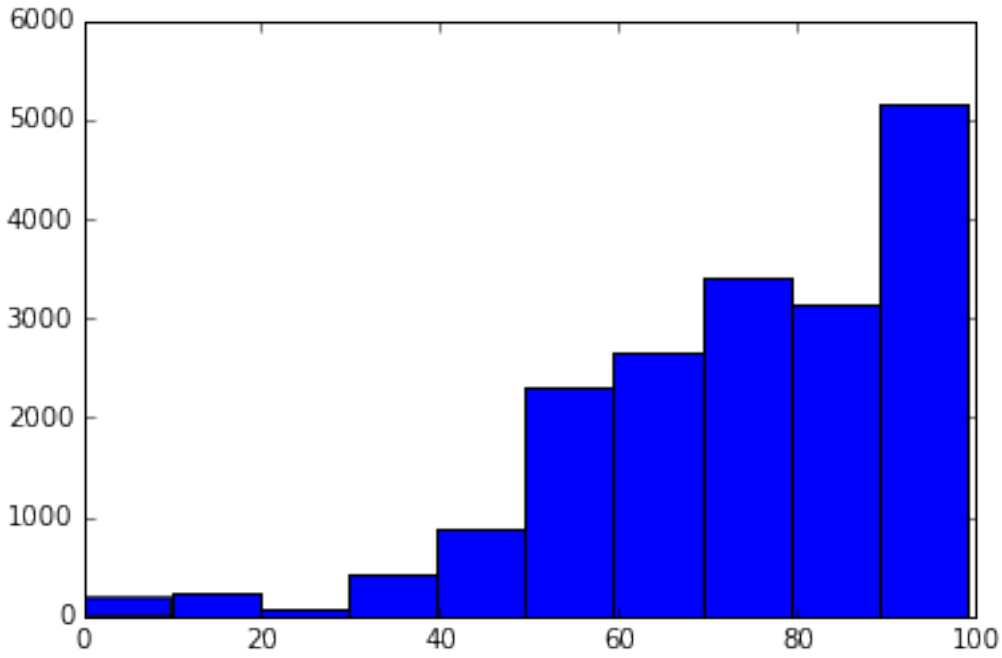

7000
6000
5000
4000
3000
2000
1000
0
0
20
40
60
80
100

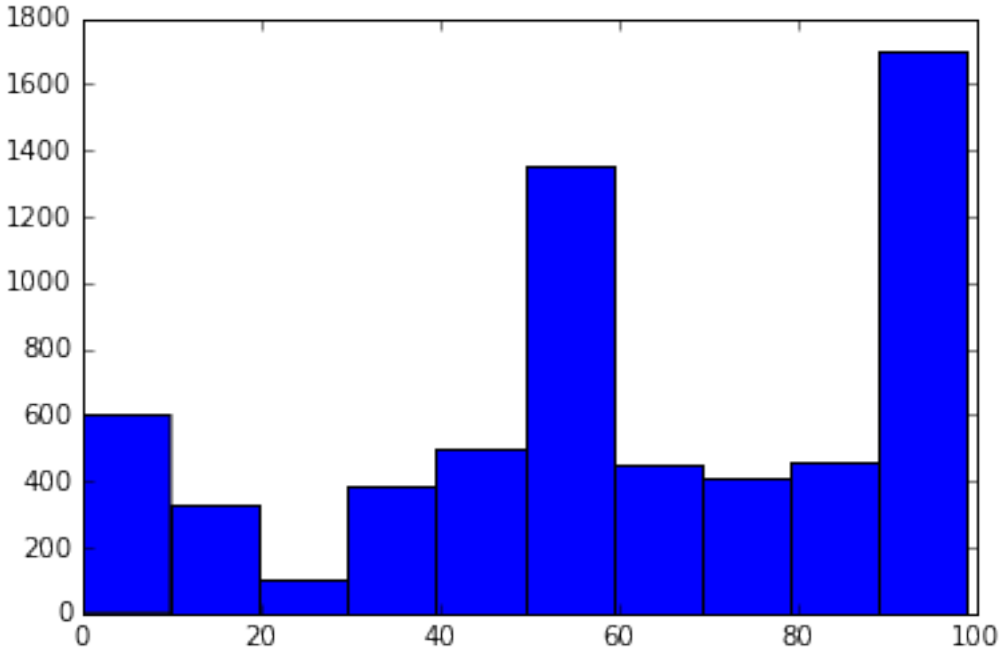

6000
5000
4000
3000
2000
1000
0
0
20
40
60
80
100

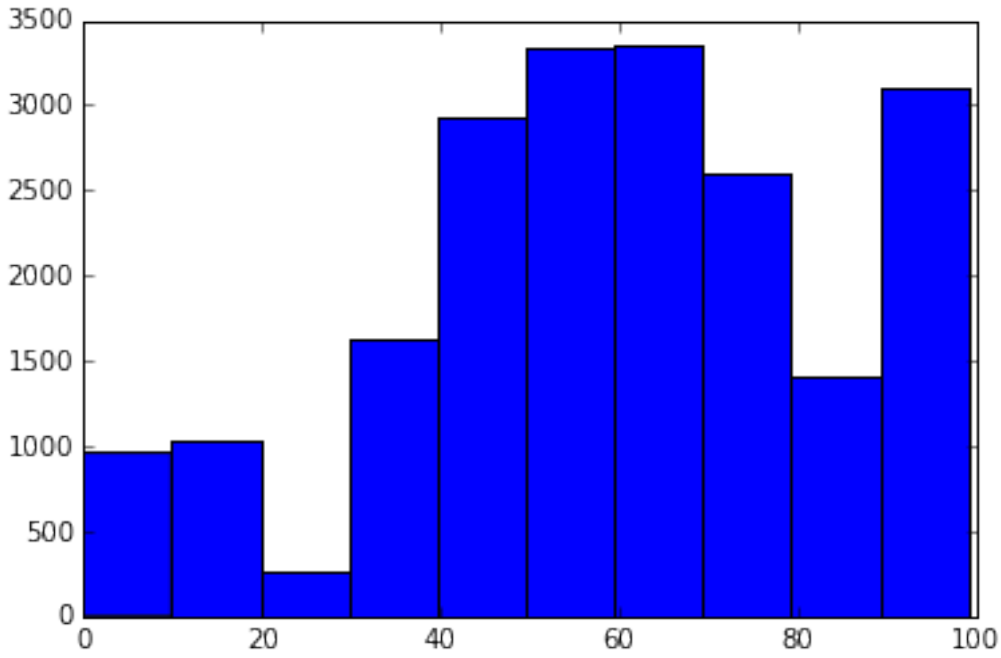

3500
3000
2500
2000
1500
1000
500
0
0
20
40
60
80
100

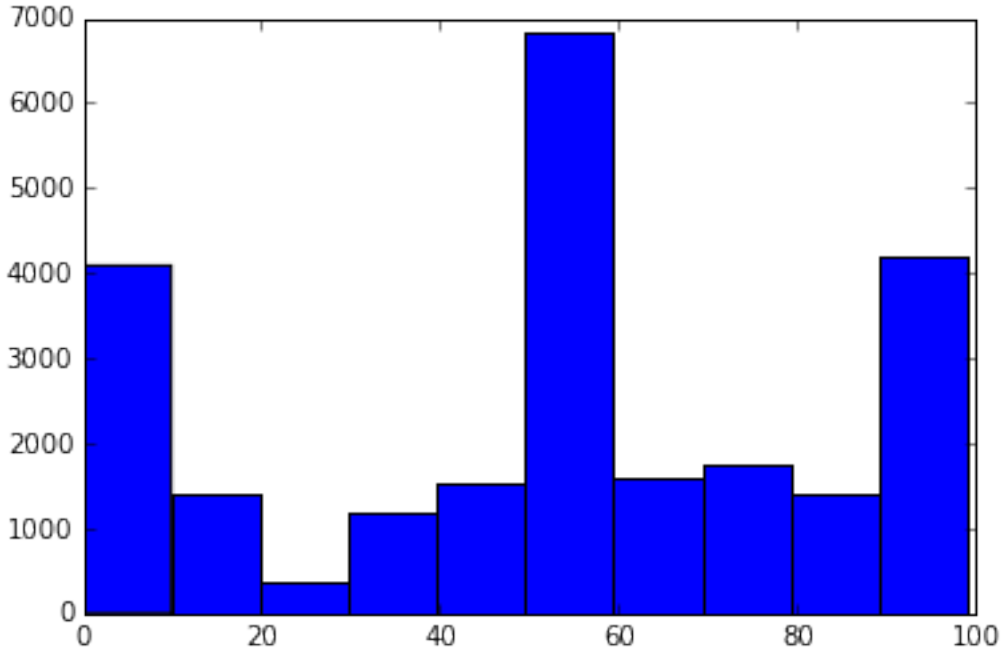

4500
4000
3500
3000
2500
2000
1500
1000
500
0
0
20
40
60
80
100

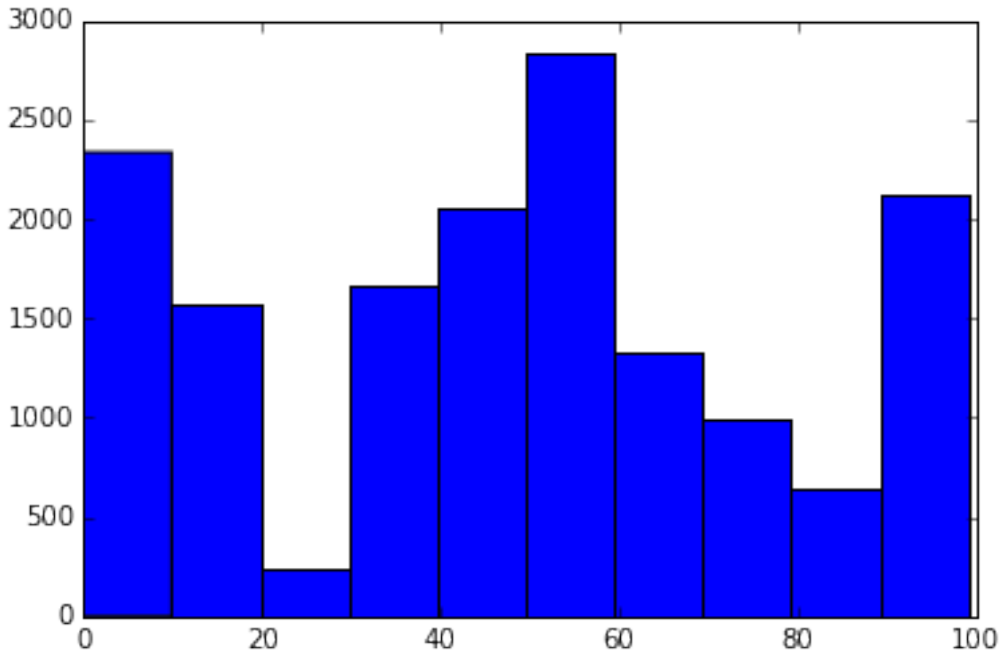

4500
4000
3500
3000
2500
2000
1500
1000
500
0
0
20
40
60
80
100

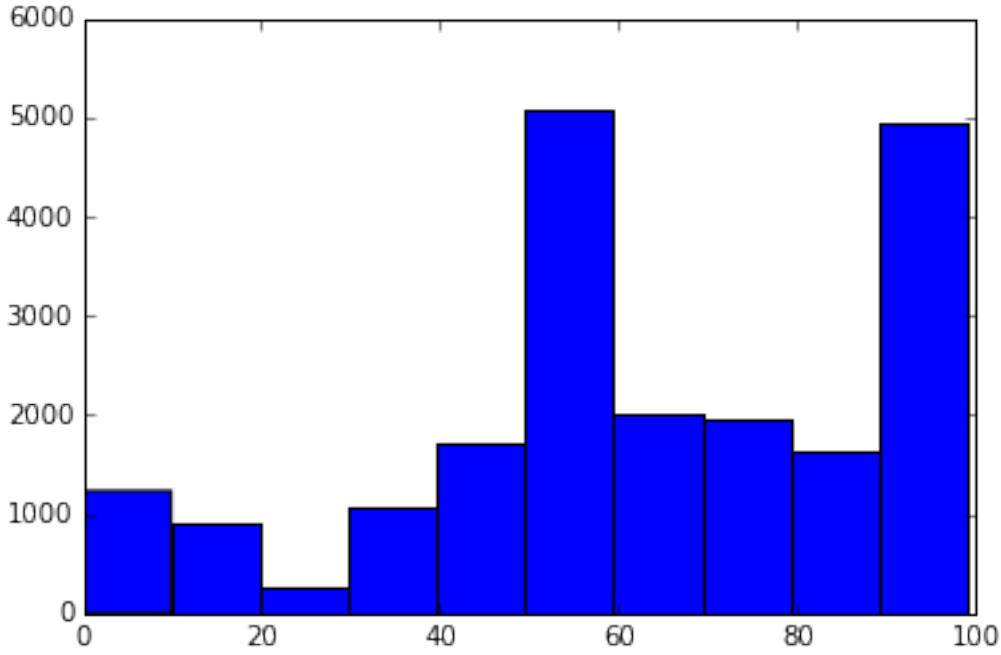

6000
5000
4000
3000
2000
1000
0
0
20
40
60
80
100

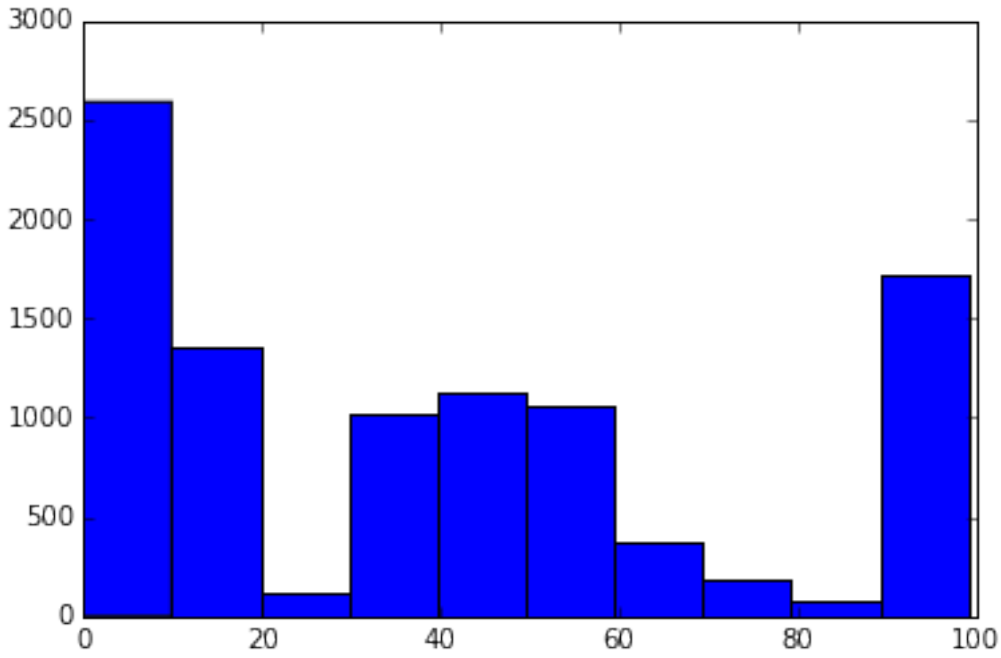

1800
1600
1400
1200
1000
800
600
400
200
0
0
20
40
60
80
100

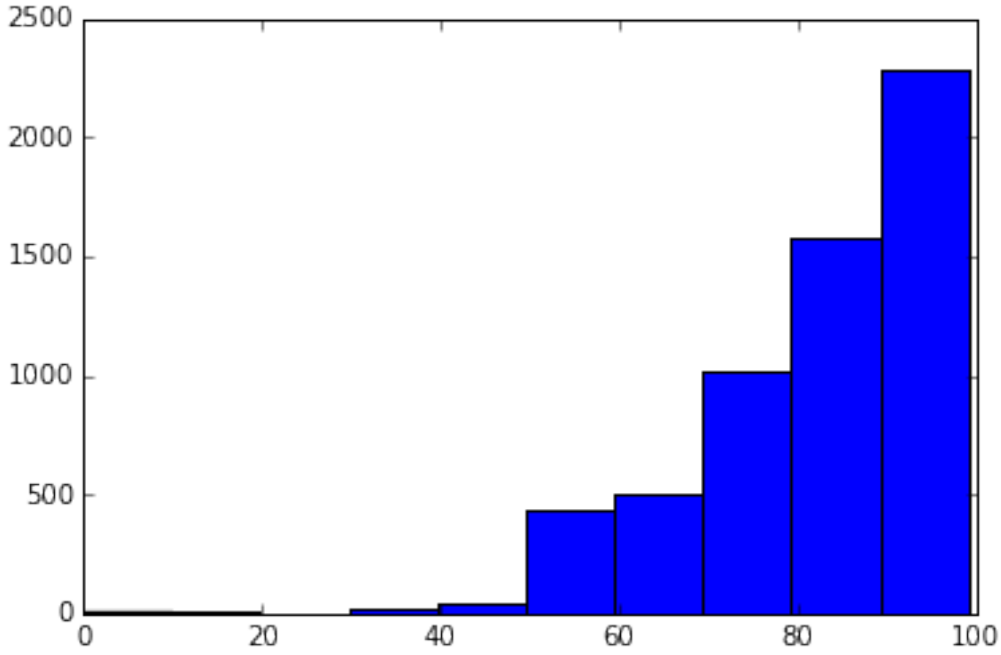

3500
3000
2500
2000
1500
1000
500
0
0
20
40
60
80
100

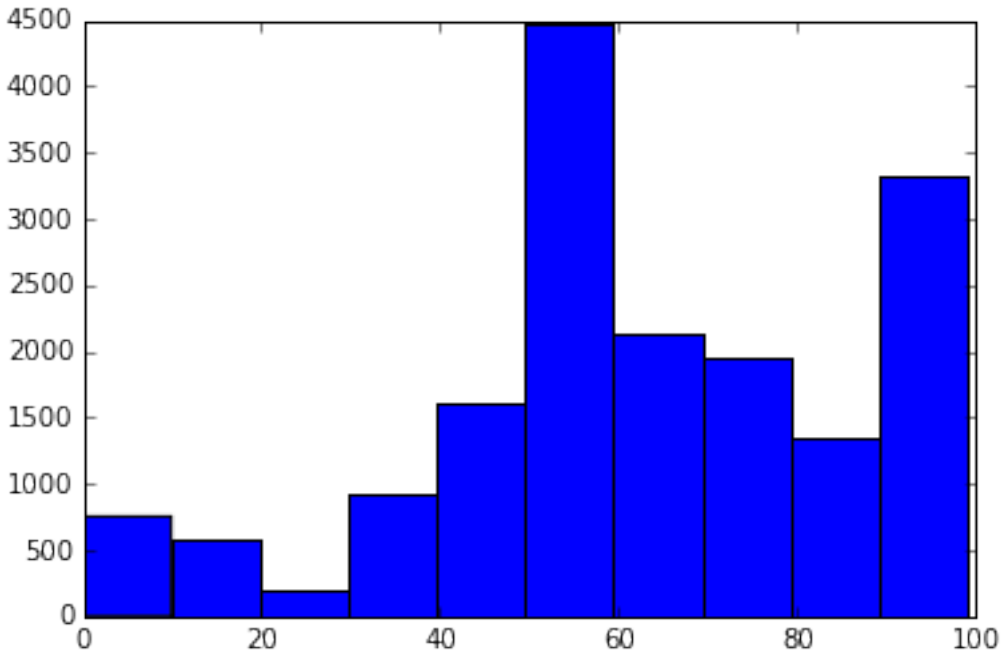

7000
6000
5000
4000
3000
2000
1000
0
0
20
40
60
80
100


3000
2500
2000
1500
1000
500
0
0
20
40
60
80
100


6000
5000
4000
3000
2000
1000
0
0
20
40
60
80
100


3000
2500
2000
1500
1000
500
0
0
20
40
60
80
100


2500
2000
1500
1000
500
0
0
20
40
60
80
100


4500
4000
3500
3000
2500
2000
1500
1000
500
0
0
20
40
60
80
100