PCA Quantify Republican

January 13, 2021

1 Using The First Principal Component of PCA to Quantify How Republican Someone is

Author: V Yang

PCA is a great tool for unsupervised machine learning, which it can dicover the hidden linearly correlated features in our dataset, and extract important informations from a large dataset. However, only using the first principal component of PCA is also a great application of binary classification. In the follow research, I will use the first component of PCA to classify the 100 senators as Rep or Dem by examinating their voting pattern.

The data comes from the 114th congress senator voting record.

1.1 Step 1: Data Cleaning, Setting up PCA, Evaluating Classification

[1]:
```
import pandas as pd
import numpy as np
import scipy as sp
import matplotlib.pyplot as plt
from sklearn.decomposition import PCA
%matplotlib inline

np.random.seed(10)
```

[2]:
```
#in the follow table, each row is a senator, so there are 100 rows
# each column correspond to a specific Amendment/Bill/Act so there are about␣
 ↪330 ish columns
vote_df = pd.read_csv("senate114.csv")
vote_df.head(3)
```

[2]:
```
   Unnamed: 0  Alexander Amdt. No. 2139  Baldwin Amdt. No. 432  \
0           0                         1                      0
1           1                         0                      0
2           2                         0                      1

   Barrasso Amdt. No. 1147  Barrasso Amdt. No. 347  Bennet Amdt. No. 1014  \
0                        0                       1                      0
1                        1                       1                      1
```

```
2                       0                       0                      1

   Blunt Amdt. No. 78 As Modified  Blunt Amdt. No. 928  Booker Amdt. No. 155  \
0                               1                    1                     0
1                               0                    1                     0
2                               0                    0                     1

   Booker Amdt. No. 2169  …  Whitehouse Amdt No. 867  \
0                      0  …                        0
1                      1  …                        0
2                      1  …                        1

   Whitehouse Amdt. No. 148  Whitehouse Amdt. No. 29  \
0                         0                        1
1                         0                        1
2                         1                        1

   Wicker Motion to Instruct Conferees (Re: Combination Length Limitations)  \
0                                                  0
1                                                  0
2                                                  1

   Wyden Admt. No. 1012  Wyden Amdt. No. 2621, As Modified  \
0                     0                                  0
1                     0                                  0
2                     1                                  1

   Wyden Amdt. No. 27  Wyden Amdt. No. 968             name  party
0                   1                    0  Lamar Alexander      R
1                   1                    1     Kelly Ayotte      R
2                   1                    1    Tammy Baldwin      D

[3 rows x 336 columns]
```

[3]:
```
#format the data, so that the name and party are on the most left
all_entry = list(vote_df.columns)
all_entry.insert(0, all_entry.pop())
all_entry.insert(0, all_entry.pop())
vote_df = vote_df.reindex(columns = all_entry)
vote_df = vote_df.drop(["Unnamed: 0"], axis = 1)
vote_df.head(5)
```

[3]:
```
              name party  Alexander Amdt. No. 2139  Baldwin Amdt. No. 432  \
0  Lamar Alexander     R                         1                      0
1     Kelly Ayotte     R                         0                      0
2    Tammy Baldwin     D                         0                      1
3    John Barrasso     R                         1                      0
```

```
4   Michael Bennet     D                        0                      1

   Barrasso Amdt. No. 1147  Barrasso Amdt. No. 347  Bennet Amdt. No. 1014  \
0                        0                       1                      0
1                        1                       1                      1
2                        0                       0                      1
3                        1                       1                      0
4                        0                       0                      1

   Blunt Amdt. No. 78 As Modified  Blunt Amdt. No. 928  Booker Amdt. No. 155  \
0                               1                    1                     0
1                               0                    1                     0
2                               0                    0                     1
3                               1                    1                     0
4                               0                    0                     0

   …  Warren Amdt. No. 652  \
0  …                     0
1  …                     0
2  …                     1
3  …                     0
4  …                     1

   Warren Motion to Instruct (Student Loans) Re: S.Con.Res. 11  \
0                                                  0
1                                                  0
2                                                  1
3                                                  0
4                                                  1

   Whitehouse Amdt No. 867  Whitehouse Amdt. No. 148  Whitehouse Amdt. No. 29  \
0                        0                         0                        1
1                        0                         0                        1
2                        1                         1                        1
3                        0                         0                        1
4                        1                         1                        1

   Wicker Motion to Instruct Conferees (Re: Combination Length Limitations)  \
0                                                  0
1                                                  0
2                                                  1
3                                                  0
4                                                  1

   Wyden Admt. No. 1012  Wyden Amdt. No. 2621, As Modified  \
0                     0                                  0
1                     0                                  0
```

```
2                    1                                  1
3                    0                                  0
4                    1                                  1

   Wyden Amdt. No. 27  Wyden Amdt. No. 968
0                   1                    0
1                   1                    1
2                   1                    1
3                   0                    0
4                   1                    1

[5 rows x 335 columns]
```

[24]:
```
#now everything is good, we have the 100 senators and their party on the left␣
 ↪side of the table
#more data cleaning
# in the orginal data, (1 means Yes), (2 means absent/abstain) (0 means no)
# we want to make it more interpretable (1 means yes) (0 means absent/abstain)␣
 ↪(-1 means no)

vote_df = vote_df.replace(0, -1)
vote_df = vote_df.replace(2, 0)
vote_df.head(5)
```

[24]:
```
              name party  Alexander Amdt. No. 2139  Baldwin Amdt. No. 432  \
0  Lamar Alexander     R                         1                     -1
1     Kelly Ayotte     R                        -1                     -1
2    Tammy Baldwin     D                        -1                      1
3    John Barrasso     R                         1                     -1
4   Michael Bennet     D                        -1                      1

   Barrasso Amdt. No. 1147  Barrasso Amdt. No. 347  Bennet Amdt. No. 1014  \
0                       -1                       1                     -1
1                        1                       1                      1
2                       -1                      -1                      1
3                        1                       1                     -1
4                       -1                      -1                      1

   Blunt Amdt. No. 78 As Modified  Blunt Amdt. No. 928  Booker Amdt. No. 155  \
0                               1                    1                    -1
1                              -1                    1                    -1
2                              -1                   -1                     1
3                               1                    1                    -1
4                              -1                   -1                    -1

   …  Warren Amdt. No. 652  \
0  …                    -1
```

```
1  …                     -1
2  …                      1
3  …                     -1
4  …                      1

   Warren Motion to Instruct (Student Loans) Re: S.Con.Res. 11  \
0                                                 -1
1                                                 -1
2                                                  1
3                                                 -1
4                                                  1

   Whitehouse Amdt No. 867  Whitehouse Amdt. No. 148  Whitehouse Amdt. No. 29  \
0                       -1                        -1                        1
1                       -1                        -1                        1
2                        1                         1                        1
3                       -1                        -1                        1
4                        1                         1                        1

   Wicker Motion to Instruct Conferees (Re: Combination Length Limitations)  \
0                                                 -1
1                                                 -1
2                                                  1
3                                                 -1
4                                                  1

   Wyden Admt. No. 1012  Wyden Amdt. No. 2621, As Modified  \
0                    -1                                 -1
1                    -1                                 -1
2                     1                                  1
3                    -1                                 -1
4                     1                                  1

   Wyden Amdt. No. 27  Wyden Amdt. No. 968
0                   1                   -1
1                   1                    1
2                   1                    1
3                  -1                   -1
4                   1                    1

[5 rows x 335 columns]
```

[25]:
```
# we will first extract the voting record which will be a matrix of R100 by R333
# in order to performe PCA, we must center the data
Data_M = np.array((vote_df.iloc[:, 2:]))
Data_M_OG = Data_M.copy()
Data_M = Data_M - np.mean(Data_M, axis = 0)
```

```
Data_M
```

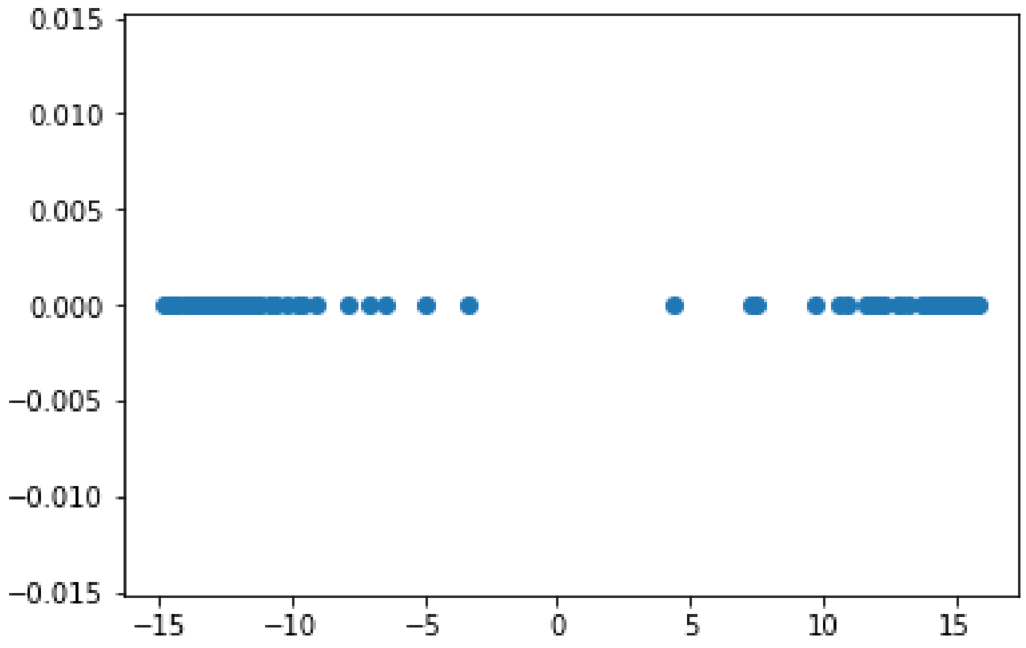
```
[25]: array([[ 1.1 , -0.9 , -0.9 , …, -0.82,  1.  , -1.46],
             [-0.9 , -0.9 ,  1.1 , …, -0.82,  1.  ,  0.54],
             [-0.9 ,  1.1 , -0.9 , …,  1.18,  1.  ,  0.54],
             …,
             [-0.9 ,  1.1 , -0.9 , …, -0.82,  1.  ,  0.54],
             [ 1.1 , -0.9 ,  1.1 , …, -0.82, -1.  , -1.46],
             [-0.9 ,  1.1 , -0.9 , …,  1.18,  1.  ,  0.54]])
```

```
[26]: #now let's fit the first principal component to the data_set
      #which we expect it to be vector in the same dimension as the total number of␣
       ↪bills/act/amendment
      PCA_1 = PCA().fit(Data_M).components_[0]

      PCA_1.shape
```

```
[26]: (333,)
```

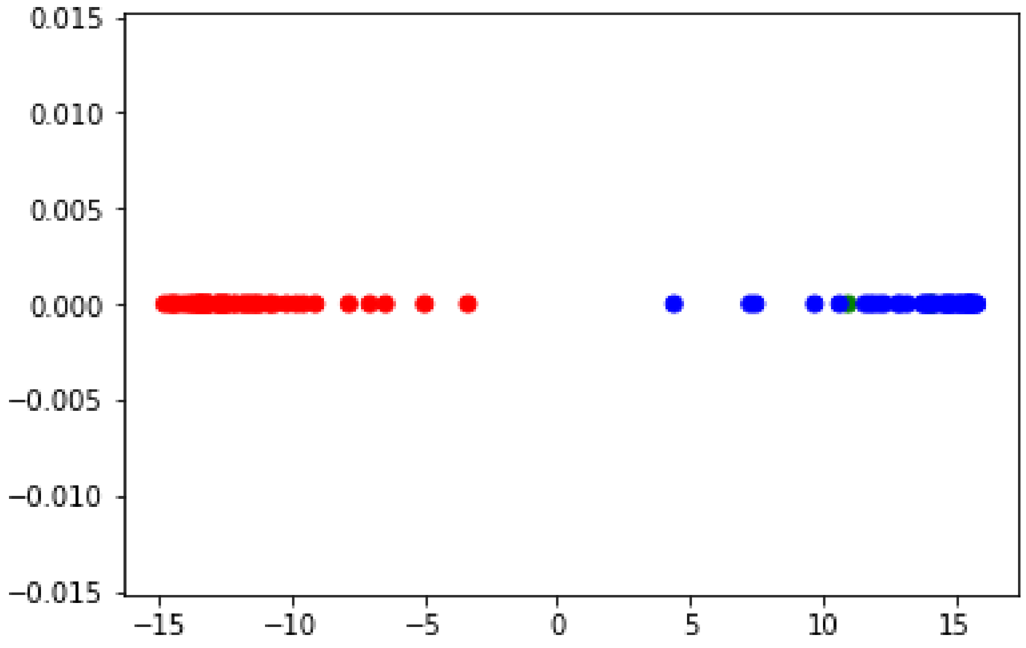
```
[27]: # by performing a matrix muliplication, we project each sentor's voting record␣
       ↪(R333) onto
      # the PCA_1 (R333) to obtain the scalar, and that scalar is quantification of␣
       ↪this senator's republican score
      score = Data_M_OG @ PCA_1
      plt.scatter(score, np.zeros_like(score))
```

```
[27]: <matplotlib.collections.PathCollection at 0x7f16c2f03f10>
```



[28]:
```
# perfect, we see the PCA_1 clearly separate the 100 senators onto two␣
 ↪different groups
# now let's color code it to check if the classification is correct, such that␣
 ↪there is no Dem senators
# in the cluster of Rep senators
# color coding goes as the follow: Dem -> Blue    Rep -> Red    Ind -> Green
def coloring(party):
    if party == "D":
        return "blue"
    if party == "R":
        return "red"
    else:
        return "green"

color_ray = vote_df["party"].apply(coloring)

plt.scatter(score, np.zeros_like(score), c = color_ray)
```

[28]:
```
<matplotlib.collections.PathCollection at 0x7f16c2ed6fa0>
```



[29]:
```
#base on the plot above, the classification is great
# republican are negative scores and Dems are positive scores
```

1.2 Step 2: interpret the weights of PCA_1

[31]:
```
# the weights of PCA_1 carries a significant meaning
# PCA_1 is R333, and total number of bill/amendment/act is also R333
# every index of PCA_1 says about some weight in the corresponding index of␣
 ↪bill/amdt/act
# in our case, the more positive it is , the more democrat poular the bill is
# the more negative it is, the more republican the bill is


# only displaying the first ten wieghts
PCA_1[:10]
```

[31]:
```
array([-0.06638758,  0.07508213, -0.06380683, -0.0677143 ,  0.06999518,
       -0.0718073 , -0.07210122,  0.0692578 ,  0.05918593,  0.06371942])
```

[32]:
```
#now we can look at some of the dem popular bill and popular rep bills
#let's just call bill/amdt/act as props
all_props = vote_df.columns[2:]
weight_prop_pair = list(zip(PCA_1, all_props))
weight_prop_pair.sort()
```

[33]:
```
#let look at the 15 most popular
rep_15 = weight_prop_pair[:15]
for i in rep_15:
    print("weight  " + str(i[0]))
    print("name    " + i[1])
    print("_____________________________________")
```

```
weight  -0.07562953841712632
name    Vitter Amdt. No. 515
_____________________________________
weight  -0.0756295384171263
name    Cornyn Amdt. No. 1114
_____________________________________
weight  -0.0756295384171263
name    Lee Amdt. No. 33
_____________________________________
weight  -0.07505770108274282
name    Motion to Table Manchin Amdt. No. 99
_____________________________________
weight  -0.0747181137813303
name    S.J.Res.8
_____________________________________
weight  -0.07470446318459167
name    Thune Amdt. No. 607
_____________________________________
```

```
weight  -0.0746800297418749
name   H.R. 3762 As Amended
_____________________________________
weight  -0.07456765952540194
name   Upon Reconsideration Motion to Invoke Cloture on the Motion to Proceed to
H.R. 240
_____________________________________
weight  -0.0745102971910762
name   Motion to Table Murray Amdt. No. 2876
_____________________________________
weight  -0.0741705622304681
name   Motion to Table Franken Amdt. No. 17
_____________________________________
weight  -0.07416774539080628
name   Lee Amdt. No. 750 As Modified
_____________________________________
weight  -0.07414331194808947
name   Paul Amdt. No. 2915
_____________________________________
weight  -0.07403671099769953
name   Motion to Proceed to S.J.Res.8
_____________________________________
weight  -0.07389134249205753
name   Moran Amdt. No. 73
_____________________________________
weight  -0.07380372464667687
name   Motion to Invoke Cloture on the Motion to Proceed to H.R. 2685, Upon
Reconsideration
_____________________________________
```

[34]:
```python
#let look at the 15 most popular dem
dem_15 = weight_prop_pair[333-15 :]
for i in dem_15:
    print("weight  " + str(i[0]))
    print("name   " + i[1])
    print("_____________________________________")
```

```
weight  0.0743013391780304
name   Mikulski Amdt. No. 362
_____________________________________
weight  0.07508213452545962
name   Baldwin Amdt. No. 432
_____________________________________
weight  0.07508213452545963
name   Sanders Amdt. No. 323 As Modified
_____________________________________
weight  0.07525186707625842
name   Wyden Admt. No. 1012
```

```
-------------------------------------
weight  0.0756295384171263
name   Durbin Amdt. No. 817

-------------------------------------
weight  0.0756295384171263
name   Franken Amdt. No. 828

-------------------------------------
weight  0.0756295384171263
name   Merkley Amdt. No. 842

-------------------------------------
weight  0.0756295384171263
name   Motion to Waive All Applicable Budgetary Discipline Re: Casey Amdt. No.
2893

-------------------------------------
weight  0.0756295384171263
name   Murray Amdt. No. 801

-------------------------------------
weight  0.0756295384171263
name   Murray Amdt. No. 951

-------------------------------------
weight  0.0756295384171263
name   Reed Amdt. No. 1521

-------------------------------------
weight  0.0756295384171263
name   Stabenow Amdt. No. 1072

-------------------------------------
weight  0.0756295384171263
name   Stabenow Amdt. No. 523

-------------------------------------
weight  0.0756295384171263
name   Warren Amdt. No. 652

-------------------------------------
weight  0.07562953841712632
name   Coons Amdt. No. 966 As Modified

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

1.3 Step 3: Classify the Senators on the smaller data set

[35]:
```
#Many bill are very difficult to understand if you dont have a solid background␣
 ↪of politics
#The following 20 are chosen because their understanding do not require␣
 ↪sufficient political background
#Let's see how good is our classification if we only use those 20 props
```

```
interesting = ['McCain Amdt. No. 1889','Coons Amdt. No. 2243','Wyden Amdt. No.␣
↪968','Brown Amdt. No. 1251','Tester Amdt. No. 2107', 'Gillibrand Amdt. No.␣
↪48','Sanders Amdt. No. 881','Sanders Amdt. No. 2177','Heitkamp Amdt. No.␣
↪133','Booker Amdt. No. 155','Vitter Amdt. No. 515', 'Kirk Amdt. No. 1038',␣
↪'McConnell Amdt. No. 836','Cruz Amdt. No 15','Vitter Amdt. No. 811','Coats␣
↪Amdt. No. 2888','Lee Amdt No. 2162','Paul Amdt. No. 2899','Casey Amdt. No.␣
↪632','Whitehouse Amdt. No. 29']
small_vote_df = vote_df.loc[:, ["name", "party"] + interesting]
small_vote_df.head(5)
```

[35]:

```
               name party  McCain Amdt. No. 1889  Coons Amdt. No. 2243  \
0  Lamar Alexander     R                      1                    -1
1     Kelly Ayotte     R                      1                     1
2    Tammy Baldwin     D                      1                     1
3    John Barrasso     R                     -1                    -1
4   Michael Bennet     D                      1                     1

   Wyden Amdt. No. 968  Brown Amdt. No. 1251  Tester Amdt. No. 2107  \
0                   -1                    -1                     -1
1                    1                     1                     -1
2                    1                     1                      1
3                   -1                    -1                      1
4                    1                     1                      1

   Gillibrand Amdt. No. 48  Sanders Amdt. No. 881  Sanders Amdt. No. 2177  \
0                       -1                     -1                      -1
1                       -1                     -1                      -1
2                        1                      1                       1
3                       -1                     -1                      -1
4                       -1                      1                       1

   …  Vitter Amdt. No. 515  Kirk Amdt. No. 1038  McConnell Amdt. No. 836  \
0  …                     1                    1                        1
1  …                     1                    1                        1
2  …                    -1                   -1                       -1
3  …                     1                    1                        1
4  …                    -1                   -1                       -1

   Cruz Amdt. No 15  Vitter Amdt. No. 811  Coats Amdt. No. 2888  \
0                 1                     1                     1
1                 1                     1                     1
2                -1                    -1                    -1
3                 1                     1                     1
4                -1                     1                    -1

   Lee Amdt No. 2162  Paul Amdt. No. 2899  Casey Amdt. No. 632  \
0                 -1                   -1                    1
```

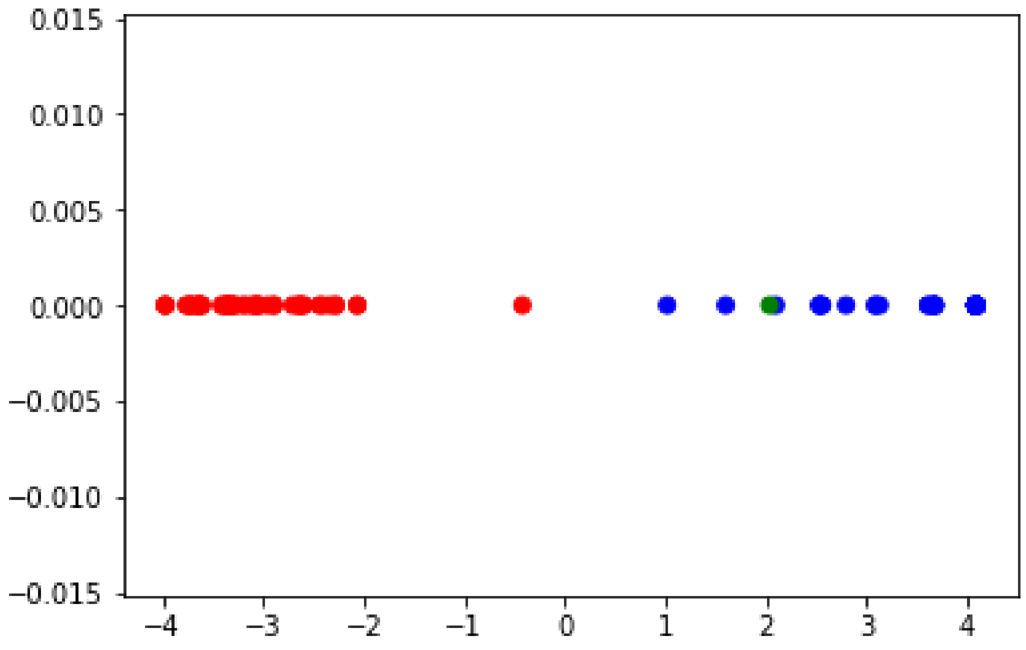
```
1                   1                    -1                     1
2                  -1                    -1                     1
3                   1                     1                     1
4                  -1                    -1                     1

   Whitehouse Amdt. No. 29
0                        1
1                        1
2                        1
3                        1
4                        1

[5 rows x 22 columns]
```

[36]:
```python
#now we have a small matrix to work, 100 by 20 rather than 100 by 333
small_Data_M = np.array((small_vote_df.iloc[:, 2:]))
small_Data_M_OG = small_Data_M.copy()
small_Data_M = small_Data_M - np.mean(small_Data_M, axis = 0)
small_Data_M.shape
```

[36]:
```
(100, 20)
```

[37]:
```python
small_PCA_1 = PCA().fit(small_Data_M).components_[0]
small_score = small_Data_M_OG @ small_PCA_1
plt.scatter(small_score, np.zeros_like(small_score), c = color_ray)
```

[37]:
```
<matplotlib.collections.PathCollection at 0x7f16c2e40850>
```



1.4 Step 4: Score each senator on the scale 0->100

1.4.1 100 means very republican, 0 means very democratic

[38]:
```
#the first thing we have to do here is to normalize the data

offset = np.sum(abs(small_PCA_1))
#linearly transform the score to a number between 0-->100
def convert(n):
    return 100 - (n + offset) * (100/(2 * offset))

senator_score = vote_df.iloc[:, :2]
senator_score["score"] = np.round(convert(small_score))
senator_score = senator_score.sort_values(by = "score")
senator_score.head(50)
```

[38]:
```
                   name party  score
79       Bernie Sanders     I    0.0
44         Mazie Hirono     D    0.0
29       Richard Durbin     D    0.0
52        Amy Klobuchar     D    0.0
54        Patrick Leahy     D    0.0
57       Edward Markey     D    0.0
61      Robert Menendez     D    0.0
62         Jeff Merkley     D    0.0
63     Barbara Mikulski     D    0.0
66  Christopher Murphy     D    0.0
21   Christopher Coons     D    0.0
71          Gary Peters     D    0.0
73            John Reed     D    0.0
37  Kirsten Gillibrand     D    0.0
16            Bob Casey     D    0.0
35            Al Franken     D    0.0
14           Ben Cardin     D    0.0
82      Charles Schumer     D    0.0
85       Jeanne Shaheen     D    0.0
97   Sheldon Whitehouse     D    0.0
10       Sherrod  Brown     D    0.0
9         Barbara Boxer     D    0.0
2         Tammy Baldwin     D    0.0
7           Cory Booker     D    0.0
81         Brian Schatz     D    0.0
5    Richard Blumenthal     D    0.0
96     Elizabeth Warren     D    0.0
32     Dianne Feinstein     D    5.0
```

```
99           Ron Wyden     D    5.0
87     Debbie Stabenow     D    5.0
67        Patty Murray     D    5.0
12      Maria Cantwell     D    5.0
41     Martin Heinrich     D    6.0
93           Tom Udall     D    6.0
89          Jon Tester     D   12.0
49       Timothy Kaine     D   12.0
15      Thomas Carper      D   12.0
68         Bill Nelson     D   16.0
95         Mark Warner     D   19.0
4       Michael Bennet     D   19.0
74          Harry Reid     D   19.0
28        Joe Donnelly     D   19.0
42      Heidi Heitkamp     D   24.0
50          Angus King     I   25.0
59    Claire McCaskill     D   30.0
56         Joe Manchin     D   38.0
20       Susan Collins     R   55.0
72         Rob Portman     R   75.0
65      Lisa Murkowski     R   75.0
88        Dan Sullivan     R   78.0
```

[39]:
```
senator_score.iloc[50:, :].head(50)
```

[39]:
```
                  name party  score
36        Cory Gardner     R   78.0
78         Marco Rubio     R   79.0
64         Jerry Moran     R   79.0
39      Chuck Grassley     R   80.0
51           Mark Kirk     R   82.0
1         Kelly Ayotte     R   82.0
90          John Thune     R   82.0
45         John Hoeven     R   82.0
58         John McCain     R   82.0
13      Shelley Capito     R   83.0
25          Mike Crapo     R   85.0
75        James  Risch     R   85.0
27        Steve Daines     R   86.0
8         John Boozman     R   87.0
48         Ron Johnson     R   87.0
22          Bob Corker     R   87.0
92          Pat Toomey     R   87.0
18           Dan Coats     R   88.0
86      Richard Shelby     R   88.0
38      Lindsey Graham     R   89.0
26            Ted Cruz     R   89.0
```

```
76         Pat Roberts     R   90.0
60      Mitch McConnell    R   90.0
43          Dean Heller    R   90.0
83        Timothy Scott    R   91.0
11         Richard Burr    R   91.0
84  Jefferson Sessions     R   91.0
91        Thomas Tillis    R   91.0
80            Ben Sasse    R   91.0
0       Lamar Alexander    R   91.0
47       Johnny Isakson    R   91.0
77          Mike Rounds    R   91.0
34           Jeff Flake    R   91.0
69            Rand Paul    R   92.0
30            Mike Enzi    R   92.0
40          Orrin Hatch    R   94.0
23          John Cornyn    R   94.0
19         Thad Cochran    R   94.0
3         John Barrasso    R   95.0
46           Jim Inhofe    R   95.0
6             Roy Blunt    R   95.0
17         Bill Cassidy    R   95.0
70         David Perdue    R   95.0
98         Roger Wicker    R   95.0
24           Tom Cotton    R   95.0
94         David Vitter    R   96.0
55             Mike Lee    R   96.0
31           Joni Ernst    R   99.0
33          Deb Fischer    R   99.0
53       James Lankford    R   99.0
```

[ ]:

[ ]:

[ ]:

[ ]:

[ ]: