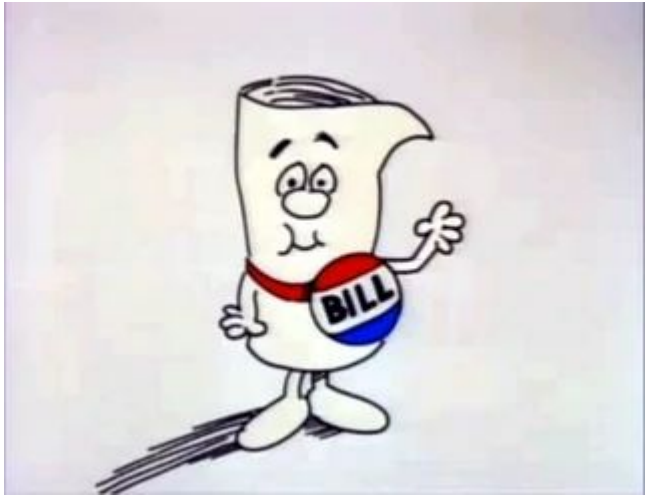
Congressional Vote Clustering

Group 2: Le Nguyen, Greeshma Ganji, Tolulope Olatunbosun, Nandhini Lakshman

Table of Contents

1. Problem Description
2. Data Mining
3. Data Cleaning
4. Data Analysis
5. Creating a model
6. Results
7. Conclusion
8. References



Problem Description

- Current American politics is very divided along party lines
- We want to see if this political divide shows up in the data
- Can members of congress be clustered into the two parties based on their voting record?
    - Do members of the same party tend to vote with each other?
    - Are some members willing to vote against their own party?
    - Are there some members that are hard to classify based on their voting recording?

Data Mining: Web Scraping

Web Scraping: collected relevant data from a given website

- For every bill, take the congressperson and their vote on said bill
- Collect this information on all bills to build a data set

We haven't done this in class, so let's take a look at what this looks like!

govtrack Home Bills & Votes Legislators Using GovTrack About Us S Log In Search

Congress / Votes

Voting Records

Each year the U.S. Senate and House of Representatives take thousands of votes, some to pass bills, resolutions, nominations, and treaties, and others on procedural matters such as on cloture and other motions. Not all votes are recorded, such as when there is no one opposed. This page shows the outcome of all recorded votes on the Senate floor and House floor. It does not include votes in committee.

TRACK VOTES

Get an email every time Congress votes on a bill or other matter.

Filter Votes

session

2021 (117th Congress)

Note: Even-year sessions extend a few days into the next year.

chamber

- [x] All
- [ ] Senate (473 items)
- [ ] House (386 items)

category

- [x] All
- [ ] Amendment (220 items)
- [ ] Cloture (132 items)
- [ ] Procedural (132 items)
- [ ] Nomination (124 items)
- [ ] Passage under Suspension (121 items)
- [ ] Passage (116 items)
- [ ] Unknown Category (12 items)

859 items matched your search.

sort by Date (Latest First)

Motion to Invoke Cloture: Re: Reed Amdt. No. 3867 As Modified to H.R. 4350
Senate Vote #473 Cloture Motion Rejected 45/51
Nov 29, 2021 5:32 p.m.

Table Motion to Reconsider: H R 5376 Build Back Better Act
House Vote #386 Passed 217/105
Nov 19, 2021 10:17 a.m.

H.R. 5376: Build Back Better Act
House Vote #385 Passed 220/213
Nov 19, 2021 9:44 a.m.
On Passage of the Bill in the House

On Motion to Recommit: H.R. 5376: Build Back Better Act
House Vote #384 Failed 208/220
Nov 19, 2021 9:08 a.m.

H.Res. 803: Providing for further consideration of the bill (H.R. 5376) to provide for reconciliation pursuant to title II of S. Con. Res. 14.
House Vote #383 Passed 220/211
Nov 18, 2021 8:11 p.m.

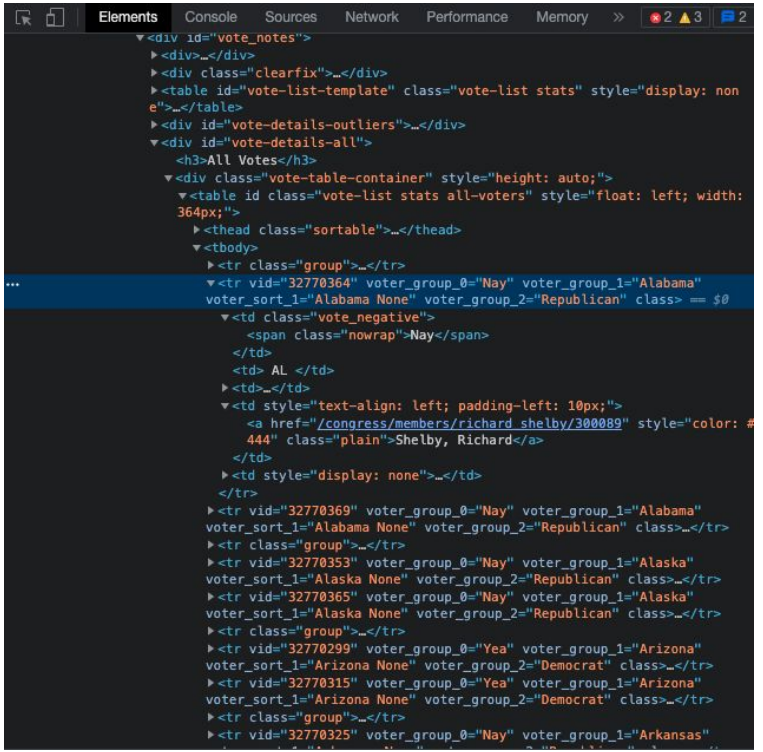
All Votes

Vote	State	Party	Representative
			Alabama
Nay	AL	R	Shelby, Richard
Nay	AL	R	Tuberville, Tommy
			Alaska
Nay	AK	R	Murkowski, Lisa
Nay	AK	R	Sullivan, Dan
			Arizona
Yea	AZ	D	Kelly, Mark
Yea	AZ	D	Sinema, Kyrsten
			Arkansas
Nay	AR	R	Boozman, John
Nay	AR	R	Cotton, Tom
			California
Yea	CA	D	Feinstein, Dianne
Yea	CA	D	Padilla, Alex
			Colorado
Yea	CO	D	Bennet, Michael
Yea	CO	D	Hickenlooper, John
			Connecticut
Yea	CT	D	Blumenthal, Richard
Yea	CT	D	Murphy, Christopher
			Delaware
Yea	DE	D	Carper, Thomas
Yea	DE	D	Coons, Christopher
			Florida
Nay	FL	R	Rubio, Marco

Vote	State	Party	Representative
Nay	KY	R	Paul, Rand
			Louisiana
No Vote	LA	R	Cassidy, Bill
Nay	LA	R	Kennedy, John
			Maine
Yea	ME	R	Collins, Susan
Yea	ME	I	King, Angus
			Maryland
Yea	MD	D	Cardin, Benjamin
Yea	MD	D	Van Hollen, Chris
			Massachusetts
Nay	MA	D	Markey, Ed
Nay	MA	D	Warren, Elizabeth
			Michigan
Yea	MI	D	Peters, Gary
Yea	MI	D	Stabenow, Debbie
			Minnesota
Yea	MN	D	Klobuchar, Amy
Yea	MN	D	Smith, Tina
			Mississippi
Nay	MS	R	Hyde-Smith, Cindy
Nay	MS	R	Wicker, Roger
			Missouri
No Vote	MO	R	Blunt, Roy
Nay	MO	R	Hawley, Josh
			Montana

Vote	State	Party	Representative
Nay	ND	R	Cramer, Kevin
Nay	ND	R	Hoeven, John
			Ohio
Yea	OH	D	Brown, Sherrod
Nay	OH	R	Portman, Rob
			Oklahoma
Nay	OK		[illegible]
Nay	OK	R	[illegible]
			Oregon
Nay	OR	D	Merkley, Jeff
Nay	OR	D	Wyden, Ron
			Pennsylvania
Yea	PA	D	Casey, Bob
Nay	PA	R	Toomey, Pat
			Rhode Island
Yea	RI	D	Reed, John
Yea	RI	D	Whitehouse, Sheldon
			South Carolina
Nay	SC	R	Graham, Lindsey
Nay	SC	R	Scott, Tim
			South Dakota
Nay	SD	R	Rounds, Mike
Nay	SD	R	Thune, John
			Tennessee
Nay	TN	R	Blackburn, Marsha
Nay	TN	R	Hagerty, Bill

Inspect

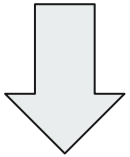


Data Cleaning

Problem	Solution
Two word states listed in multiple indices (list[0] = “New”, list[1] = “York”)	Append indices together where this happened
Not every member of congress votes on every bill, ragged data set	Create a list of all congress people, add missing people into bills they didn’t vote on. Assume they would vote with their party.
Data in string format “Yea” or “Nay”	Encode string values “Nay” = 0, “Yea” = 1

Clean Data

	Name	State	Party	Bill_1	Bill_2	Bill_3	Bill_4	Bill_5	Bill_6	Bill_7
0	a_dutch_ruppersberger	Maryland	Democrat	1	1	1	1	0	1	1
1	abigail_spanberger	Virginia	Democrat	1	1	1	1	0	1	1
2	adam_kinzinger	Illinois	Republican	1	0	0	0	1	0	1
3	adam_schiff	California	Democrat	1	1	1	1	0	1	1
4	adam_smith	Washington	Democrat	1	1	1	1	0	1	1



Transpose

Name	a_dutch_ruppersberger	abigail_spanberger	adam_kinzinger	adam_schiff	adam_smith	adrian_smith	adriano_espaillat	al_green
Bill_1	1	1	1	1	1	1	1	1
Bill_2	1	1	0	1	1	0	1	1
Bill_3	1	1	0	1	1	0	1	1
Bill_4	1	1	0	1	1	0	1	1
Bill_5	0	0	1	0	0	1	0	0

Data Analysis

Finding the mean count of democrats and republicans who voted against their party for all 311 bills

Gives us an idea about voting behaviour and certain outliers

```
# On average, 3 democrats voted against their party
np.mean(demAgainst)
```

```
3.867741935483871
```

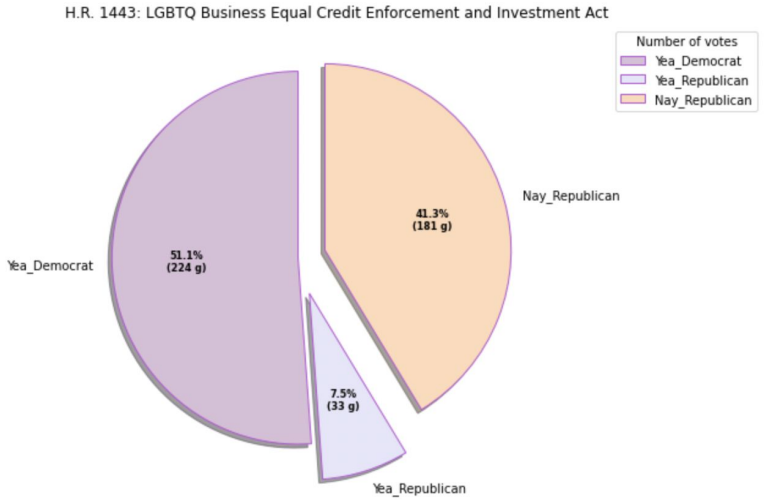
```
# On average, 12 republicans voted against their own party
np.mean(repAgainst)
```

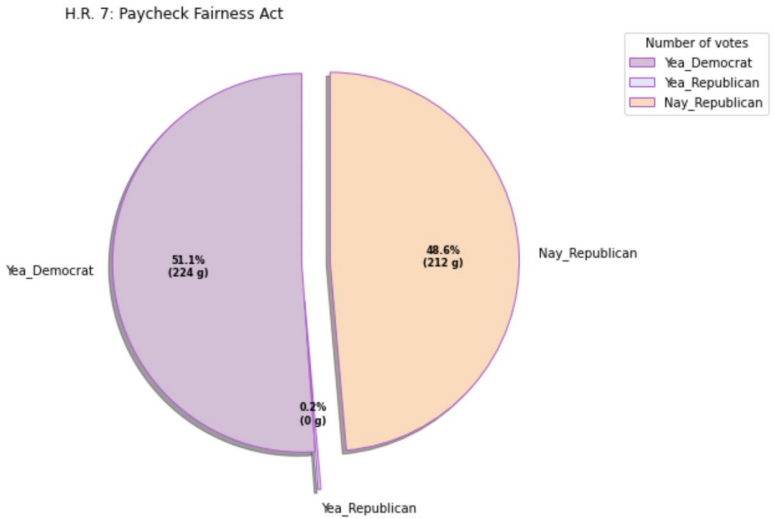
```
12.993548387096775
```

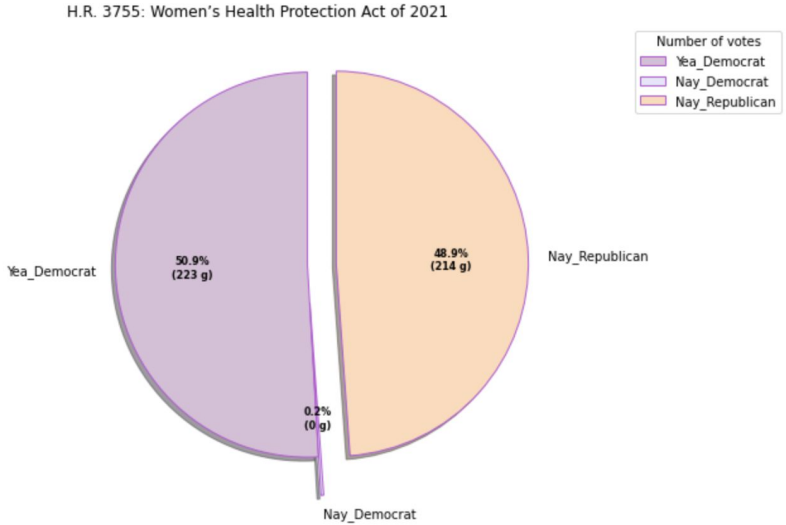
Data Visualization

```
# LGBTQ Business Equal Credit Enforcement and Investment Act
df.groupby("Bill_180")["Party"].value_counts()
```

```
Bill_180  Party
Nay       Republican    181
Yea       Democrat      224
          Republican     33
Name: Party, dtype: int64
```

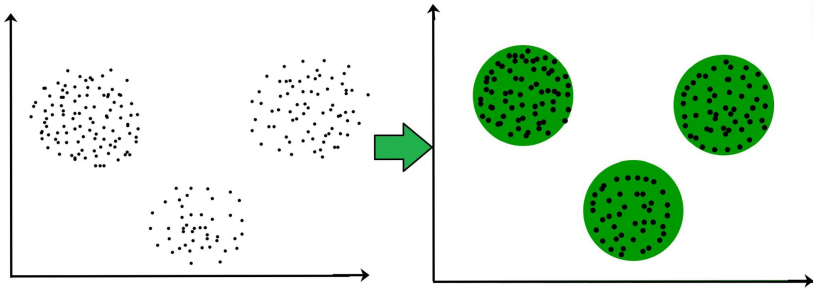



H.R. 7: Paycheck Fairness Act
Number of votes
Yea_Democrat
Yea_Republican
Nay_Republican
51.1%
(224 g)
48.6%
(212 g)
0.2%
(0 g)
Yea_Democrat
Nay_Republican
Yea_Republican


H.R. 3755: Women's Health Protection Act of 2021
Number of votes
Yea_Democrat
Nay_Democrat
Nay_Republican
50.9%
(223 g)
48.9%
(214 g)
0.2%
(0 g)
Yea_Democrat
Nay_Republican
Nay_Democrat

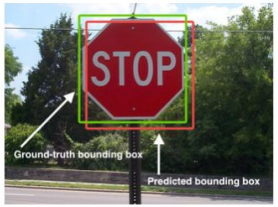
Model Creation

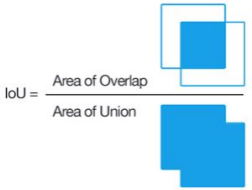
- We want some algorithm that can look at how each person voted and classify them
- First need to define some distance metric to determine how close one congressperson is to another — many such metrics exist: Manhattan, Euclidean, Cosine Similarity, Jaccard, and more
- Use an algorithm that can take in the distances to make clusters
    - Will everyone cluster into their own party?
    - Will there be some people who are difficult to cluster?

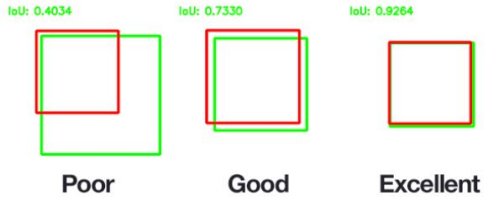


Jaccard Distance

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

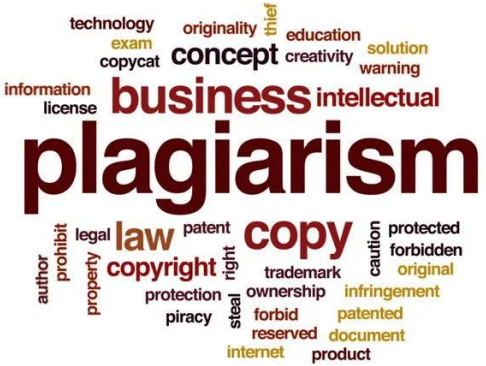




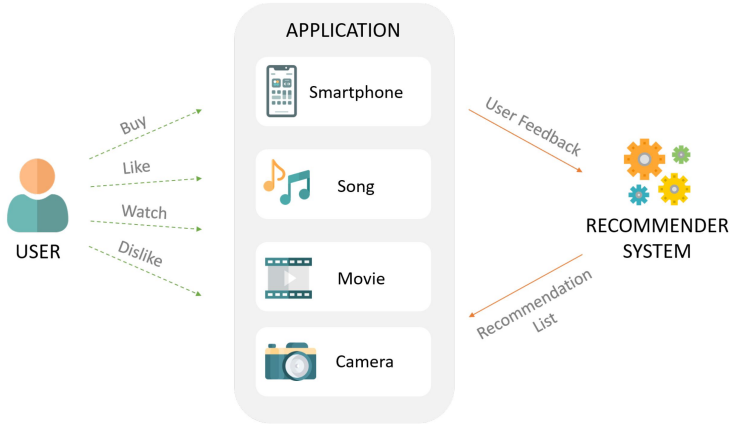


- Generally speaking, the Jaccard index is used when there are binary situations. This is fitting in our context as we manipulated the data to go from 3 voting options (Yea, Nay, Not Voting) to 2 options (Yea, Nay) with data imputation — based on mode of data set since such small percentage voted against their party.
- Aided in binarizing the choices.

Jaccard Distance Use Cases



- Recommendation Systems
- E-Commerce
- In Natural language processing, Jaccard distance is often used to measure text similarity analysis — how much word choice overlap exists between documents. This pattern analysis is often used for plagiarism detection.



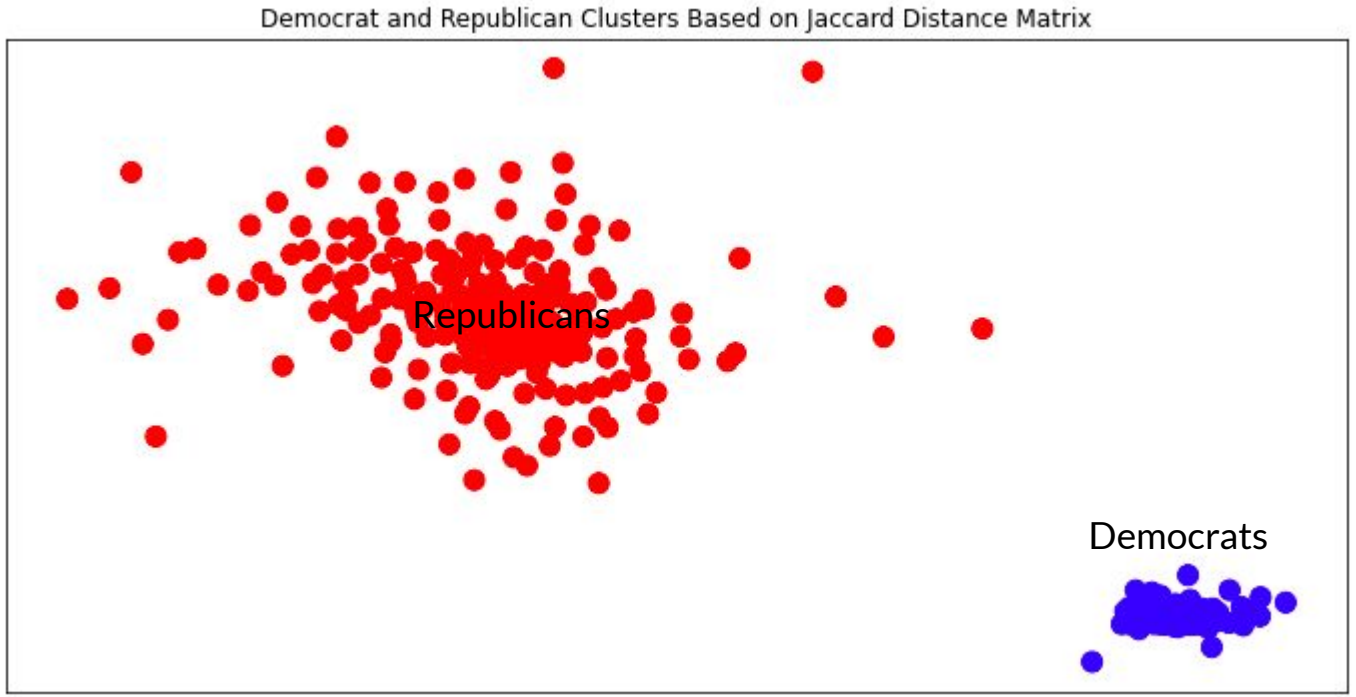
Jaccard Distance Results

abigail_spanberger	adam_kinzinger
1	1
1	0
1	0
1	0
0	1
1	0
1	1
0	0
0	0
1	0

- Often used to measure similarity and diversity of data sets.
- Jaccard Distance = intersection of A&B/Union of A&B
  - Elements of A in B/All elements in A and B
- Where both people vote together/All votes
- J(Spanberger,Kinzinger) = 0.61. This is a relatively high distance between politicians, suggesting perhaps they are polar and that they disagree on most issues.
- We can then produce a distance matrix- distance from one person to everyone else

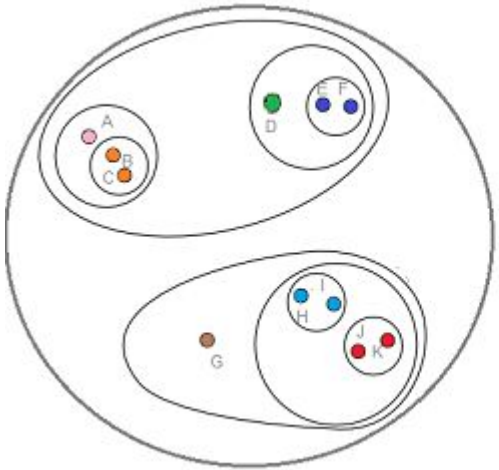
```
[0.    , 0.0444, 0.6327, 0.0242, 0.0283]
[0.0444, 0.    , 0.6103, 0.0593, 0.0558]
[0.6327, 0.6103, 0.    , 0.64  , 0.6409]
[0.0242, 0.0593, 0.64  , 0.    , 0.0121]
[0.0283, 0.0558, 0.6409, 0.0121, 0.    ]
```

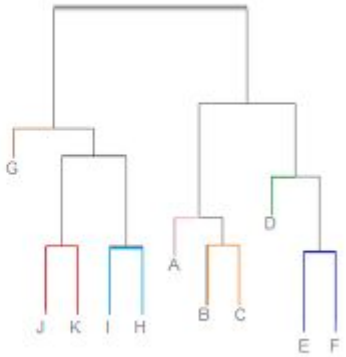
Plotted Distance Matrix



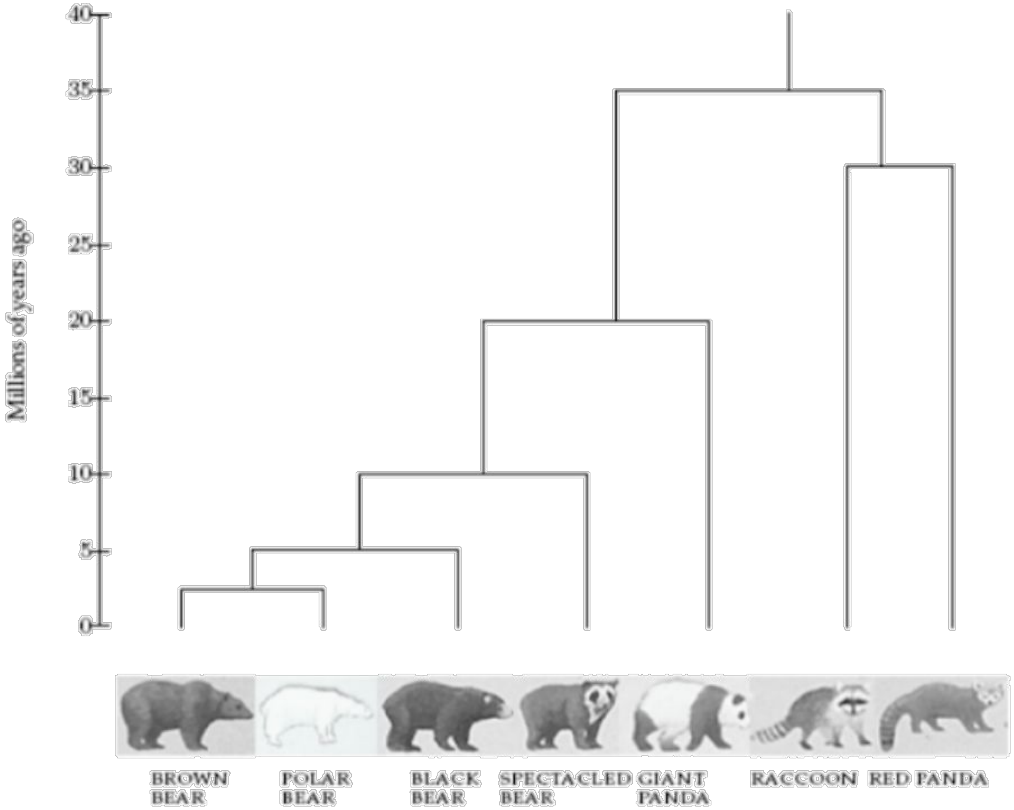
Hierarchical Clustering

- Given the distances between each point we can use Hierarchical Clustering
- This will group all points closest to each other then group all clusters closest to each other

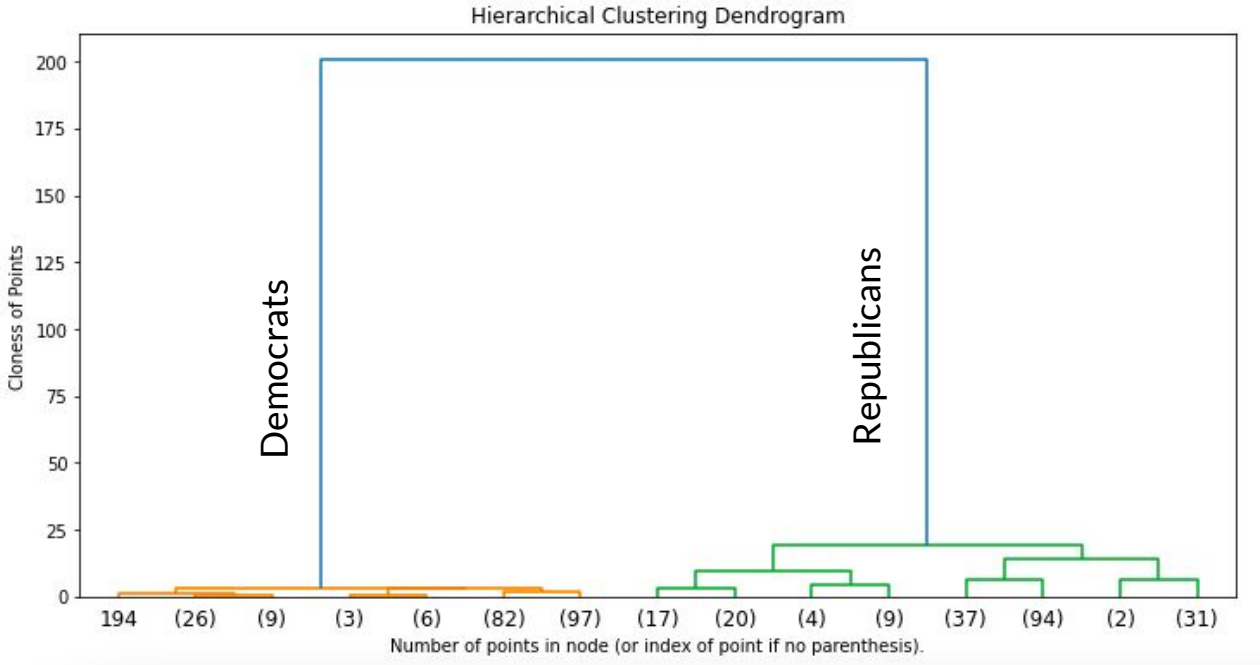




Hierarchical Clustering Example



Results



Accuracy

(Y_pred == Y)/len(Y) *100

= 100%

Conclusion

- Initial analysis in the data reflected the political division in congress
- This was further shown by using a distance metric to measure see how close/far people voted in their own party
- The vast difference in voting between the parties made clustering them accurately easy

References

Congressional Votes Database

https://www.govtrack.us/congress/votes

Related Article

https://www.forbes.com/sites/ericaswallow/2013/11/17/senate-voting-relationships-data/?sh=427aac744031

Sklearn Algomeratrive Clustering

https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html

Plotting Distance Matrix

https://stackoverflow.com/questions/36339865/generating-graph-from-distance-matrix-using-networkx-inconsistency-python