Political data

Political data

- Spoken opinion

Political data

- Spoken opinion
- Votes cast

Political data

- Spoken opinion
- Votes cast

Political data

- Spoken opinion
- Votes cast

Works both ways:

- citizen

Political data

- Spoken opinion
- Votes cast

Works both ways:

- citizen
  - polls

Political data

- Spoken opinion
- Votes cast

Works both ways:

- citizen
  - polls
  - elections

Political data

- Spoken opinion
- Votes cast

Works both ways:

- citizen
  - polls
  - elections
- representative

Political data

- Spoken opinion
- Votes cast

Works both ways:

- citizen
  - polls
  - elections
- representative
  - debates, press statements, ...

Political data

- Spoken opinion
- Votes cast

Works both ways:

- citizen
  - polls
  - elections
- representative
  - debates, press statements, ...
  - rollcalls

Political data

- Spoken opinion
- Votes cast

Works both ways:

- citizen
  - polls
  - elections
- representative
  - debates, press statements, ...
  - rollcalls

Political data

- Spoken opinion
- Votes cast

Works both ways:

- citizen
  - polls
  - elections
- representative
  - debates, press statements, ...
  - rollcalls

Analyzing representatives helps hold them responsible to their voters. Spoken opinion is hard, domain-specific and not easy to do mathematically.

Where do we get the data?

Both the Senate and the House of Representatives release their vote records as XML files on the web.

Where do we get the data?

This gives us a sequence of rollcalls.

For each rollcall, we get a sequence of votes:

⋮	⋮
Lieberman, ID,CT,	Aye
Casey, D,PA,	Aye
Biden, D,DE,	Aye
Ensign, R,NV,	Nay
Stabenow, D,MI,	Aye
Merkley, D,OR,	Aye
Kennedy, D,MA,	Aye
Kerry, D,MA,	Aye
⋮	⋮

Where do we get the data?

This gives us a sequence of rollcalls.

Taking them all together, we get a matrix of votes:

Lieberman, ID,CT,	Aye	Aye	Nay	Aye
Casey, D,PA,	Aye	Aye	Nay	Nay
Biden, D,DE,	Aye	Aye	Nay	Nay
Ensign, R,NV,	Nay	Aye	Nay	Aye
Stabenow, D,MI,	Aye	Aye	Nay	Nay
Merkley, D,OR,	Aye	Aye	Nay	Nay
Kennedy, D,MA,	Aye	Aye	Nay	Nay
Kerry, D,MA,	Aye	Aye	Nay	Nay

Where do we get the data?

This gives us a sequence of rollcalls.

Replacing "Aye" with 1 and "Nay" with -1 we get a numeric matrix:

⋮	⋮	⋮	⋮	⋮
Lieberman, ID,CT,	+1	+1	-1	+1
Casey, D,PA,	+1	+1	-1	-1
Biden, D,DE,	+1	+1	-1	-1
Ensign, R,NV,	-1	+1	-1	+1
Stabenow, D,MI,	+1	+1	-1	-1
Merkley, D,OR,	+1	+1	-1	-1
Kennedy, D,MA,	+1	+1	-1	-1
Kerry, D,MA,	+1	+1	-1	-1
⋮	⋮	⋮	⋮	⋮

Where do we get the data?

This gives us a sequence of rollcalls.

So we can view each congressman as a vector...

Lieberman, ID,CT,	+1	+1	-1	+1
Casey, D,PA,	+1	+1	-1	-1
Biden, D,DE,	+1	+1	-1	-1
Ensign, R,NV,	-1	+1	-1	+1
Stabenow, D,MI,	+1	+1	-1	-1
Merkley, D,OR,	+1	+1	-1	-1
Kennedy, D,MA,	+1	+1	-1	-1
Kerry, D,MA,	+1	+1	-1	-1

Where do we get the data?

This gives us a sequence of rollcalls.

Or each rollcall as a vector...

⋮	⋮	⋮	⋮	⋮
Lieberman, ID,CT,	+1	+1	-1	+1
Casey, D,PA,	+1	+1	-1	-1
Biden, D,DE,	+1	+1	-1	-1
Ensign, R,NV,	-1	+1	-1	+1
Stabenow, D,MI,	+1	+1	-1	-1
Merkley, D,OR,	+1	+1	-1	-1
Kennedy, D,MA,	+1	+1	-1	-1
Kerry, D,MA,	+1	+1	-1	-1
⋮	⋮	⋮	⋮	⋮

Where do we get the data?

So now it's all just numbers!

⋮	⋮	⋮	⋮
+1	+1	-1	+1
+1	+1	-1	-1
+1	+1	-1	-1
-1	+1	-1	+1
+1	+1	-1	-1
+1	+1	-1	-1
+1	+1	-1	-1
+1	+1	-1	-1
⋮	⋮	⋮	⋮

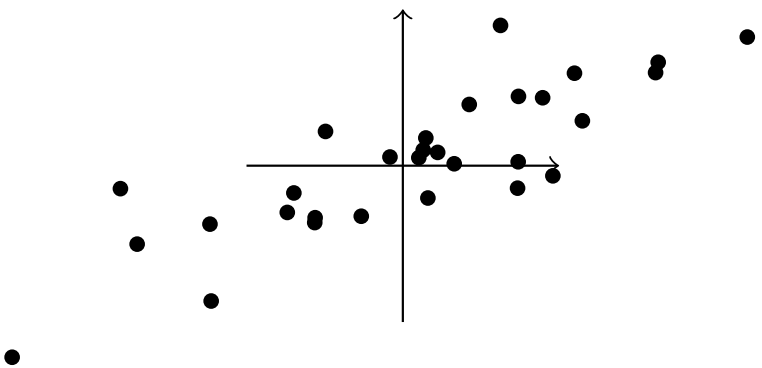
⋮	⋮	⋮	⋮
+1	+1	-1	+1
+1	+1	-1	-1
+1	+1	-1	-1
-1	+1	-1	+1
+1	+1	-1	-1
+1	+1	-1	-1
+1	+1	-1	-1
+1	+1	-1	-1
⋮	⋮	⋮	⋮

Data Analysis Paradigms

Now that we have numbers, we can view them as high-dimensional vectors, and use geometric data analysis methods on them!

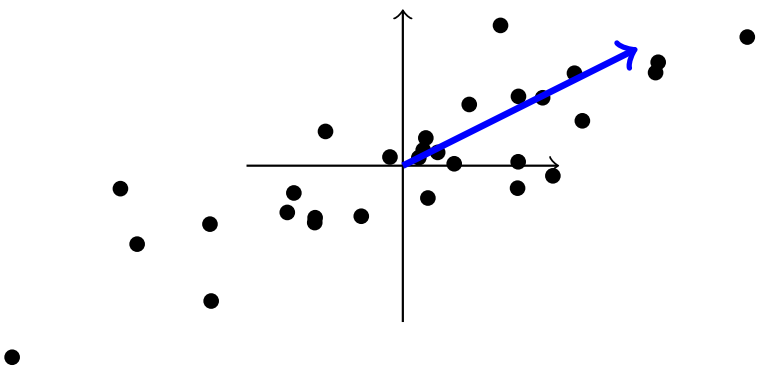
Data Analysis Paradigms

Now that we have numbers, we can view them as high-dimensional vectors, and use geometric data analysis methods on them!
Principal Component Analysis



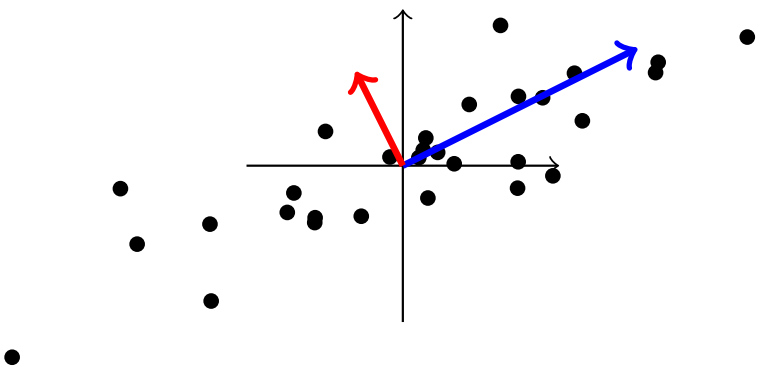
Data Analysis Paradigms

Now that we have numbers, we can view them as high-dimensional vectors, and use geometric data analysis methods on them!
Principal Component Analysis



Data Analysis Paradigms

Now that we have numbers, we can view them as high-dimensional vectors, and use geometric data analysis methods on them!
Principal Component Analysis

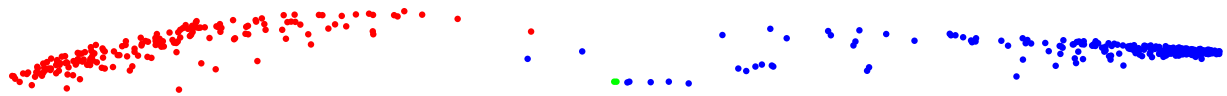


So what about the congress?

As it turns out? Remarkably easy to analyze. PCA axes come with weights. For both congress and senate members, one axis dominates.
We look at data from the House of Representatives from 2009:

So what about the congress?

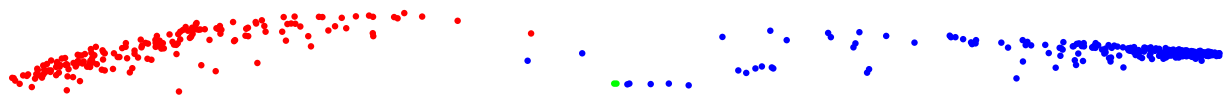
As it turns out? Remarkably easy to analyze. PCA axes come with weights. For both congress and senate members, one axis dominates.
We look at data from the House of Representatives from 2009:



And at data from the Senate from the same year:

So what about the congress?

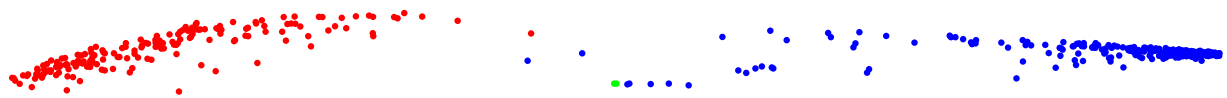
As it turns out? Remarkably easy to analyze. PCA axes come with weights. For both congress and senate members, one axis dominates.
We look at data from the House of Representatives from 2009:



And at data from the Senate from the same year:

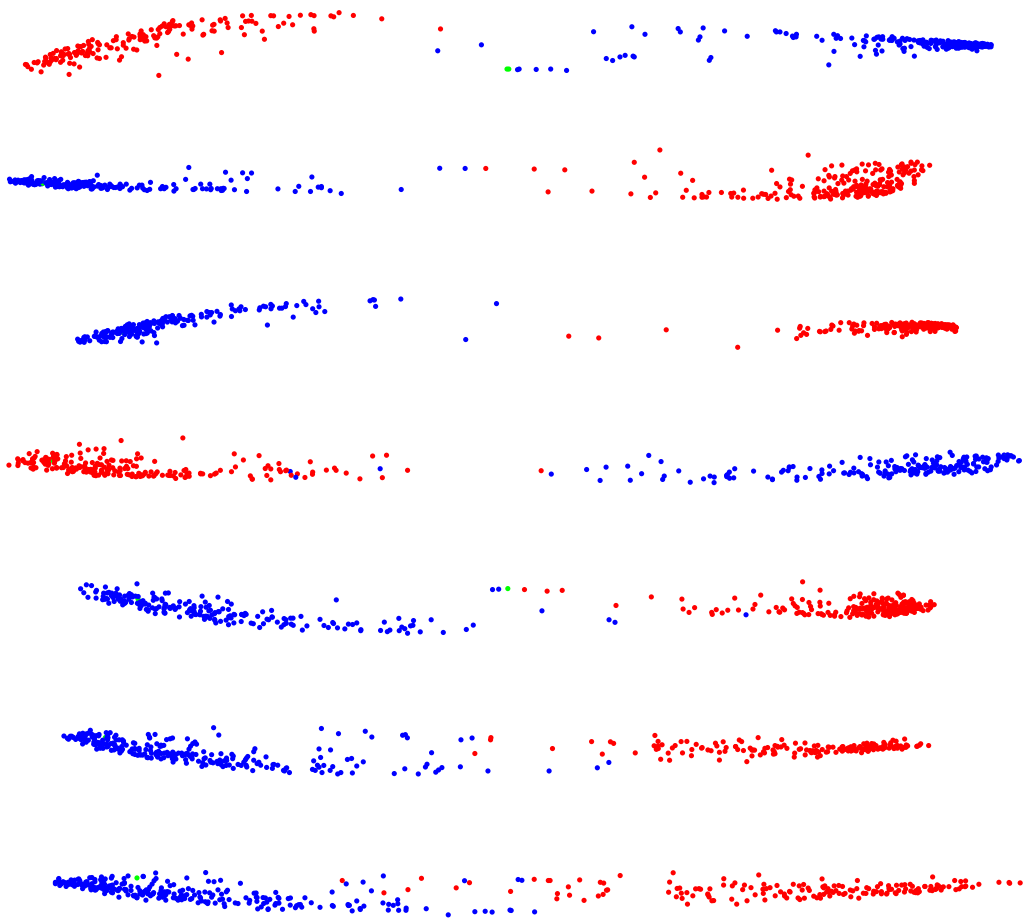
So what about the congress?

As it turns out? Remarkably easy to analyze. PCA axes come with weights. For both congress and senate members, one axis dominates.
We look at data from the House of Representatives from 2009:



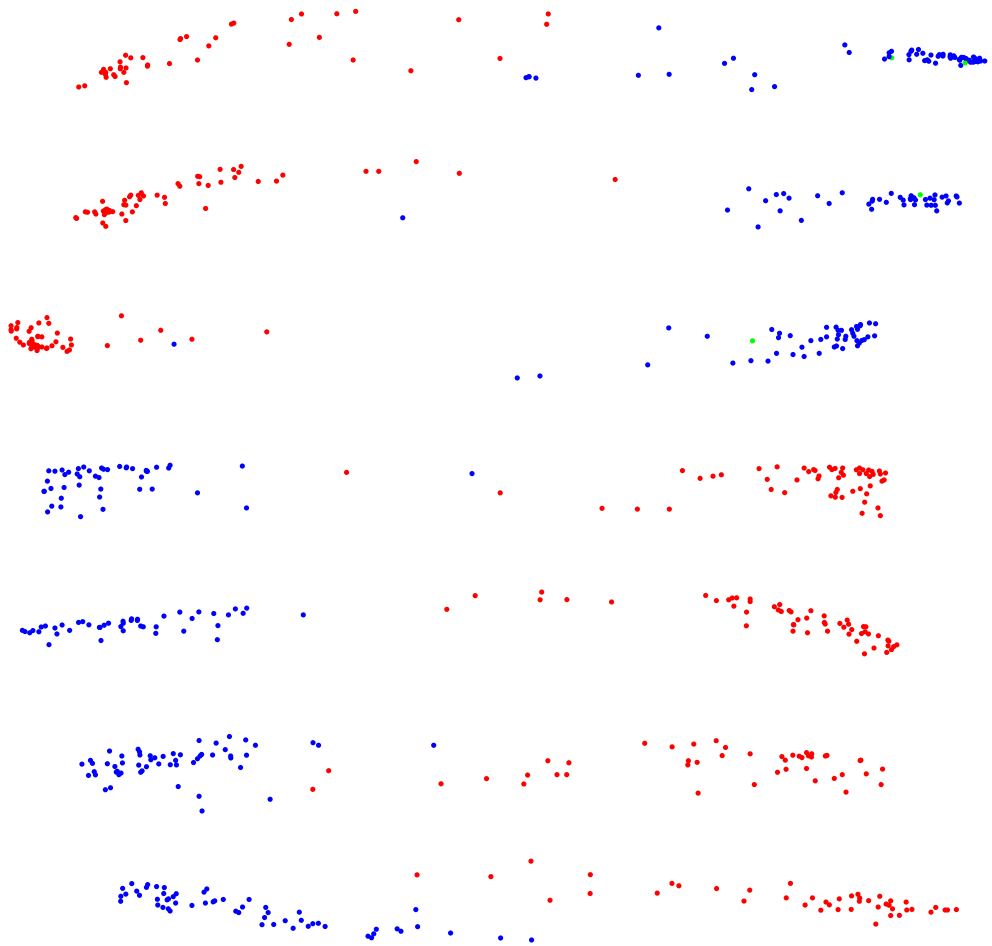
And at data from the Senate from the same year:

The pattern persists



From top to bottom: house members for the years 2009, 2006, 2003, 2000, 1997, 1994, 1991.

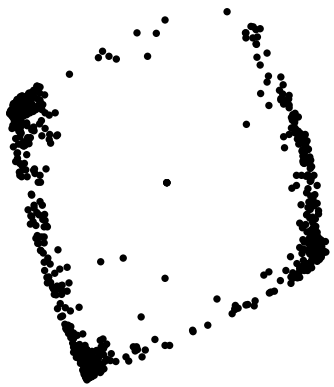
The pattern persists



From top to bottom: senate members for the years 2009, 2006, 2003, 2000, 1997, 1994, 1991.

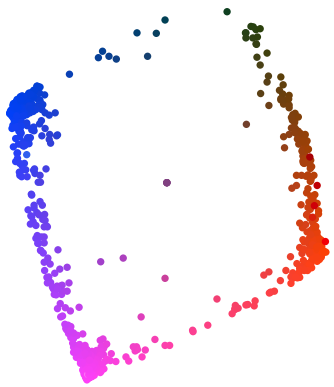
Parametrizing rollcalls

We can do the same analysis on the rollcalls instead of the members.
House of Representatives 2009...



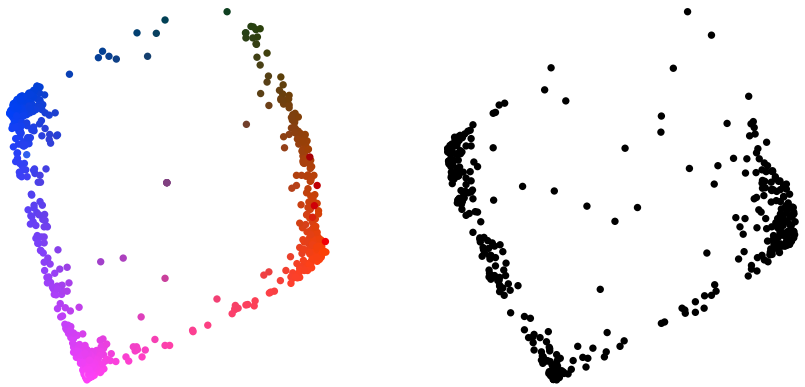
Parametrizing rollcalls

We can do the same analysis on the rollcalls instead of the members.
House of Representatives 2009...



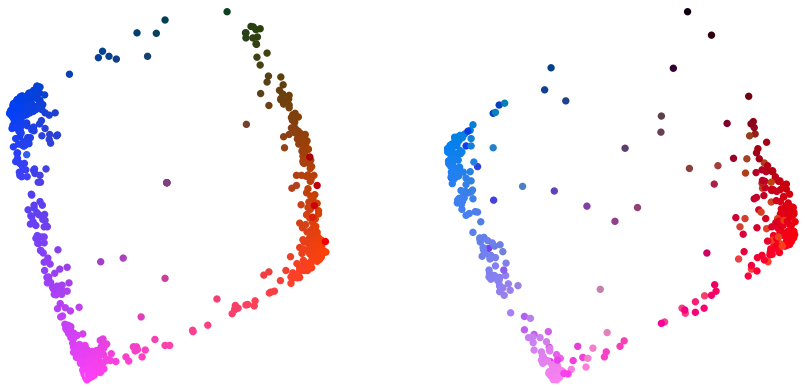
Parametrizing rollcalls

We can do the same analysis on the rollcalls instead of the members.
House of Representatives 2009... And the Senate...



Parametrizing rollcalls

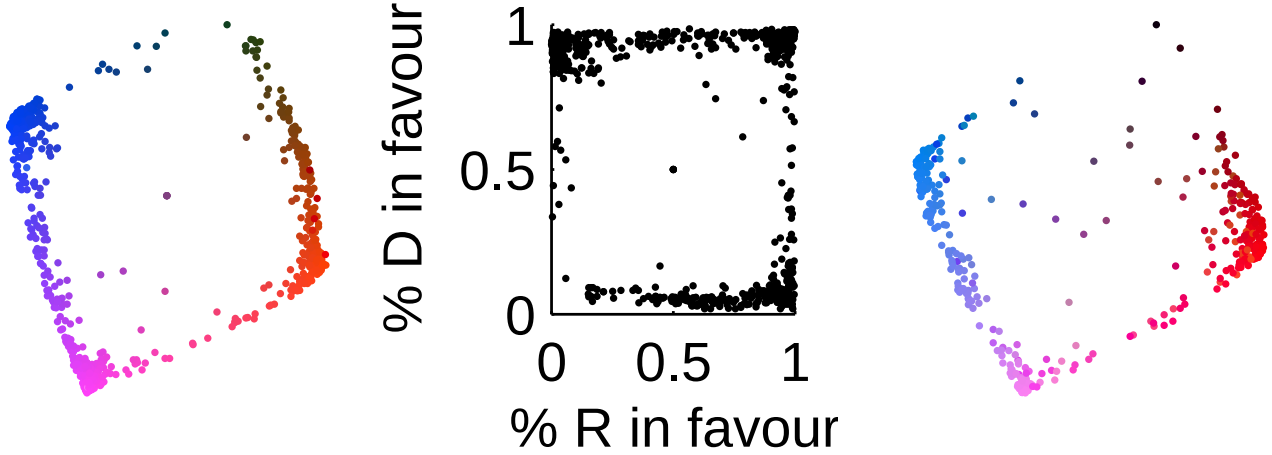
We can do the same analysis on the rollcalls instead of the members.
House of Representatives 2009... And the Senate...



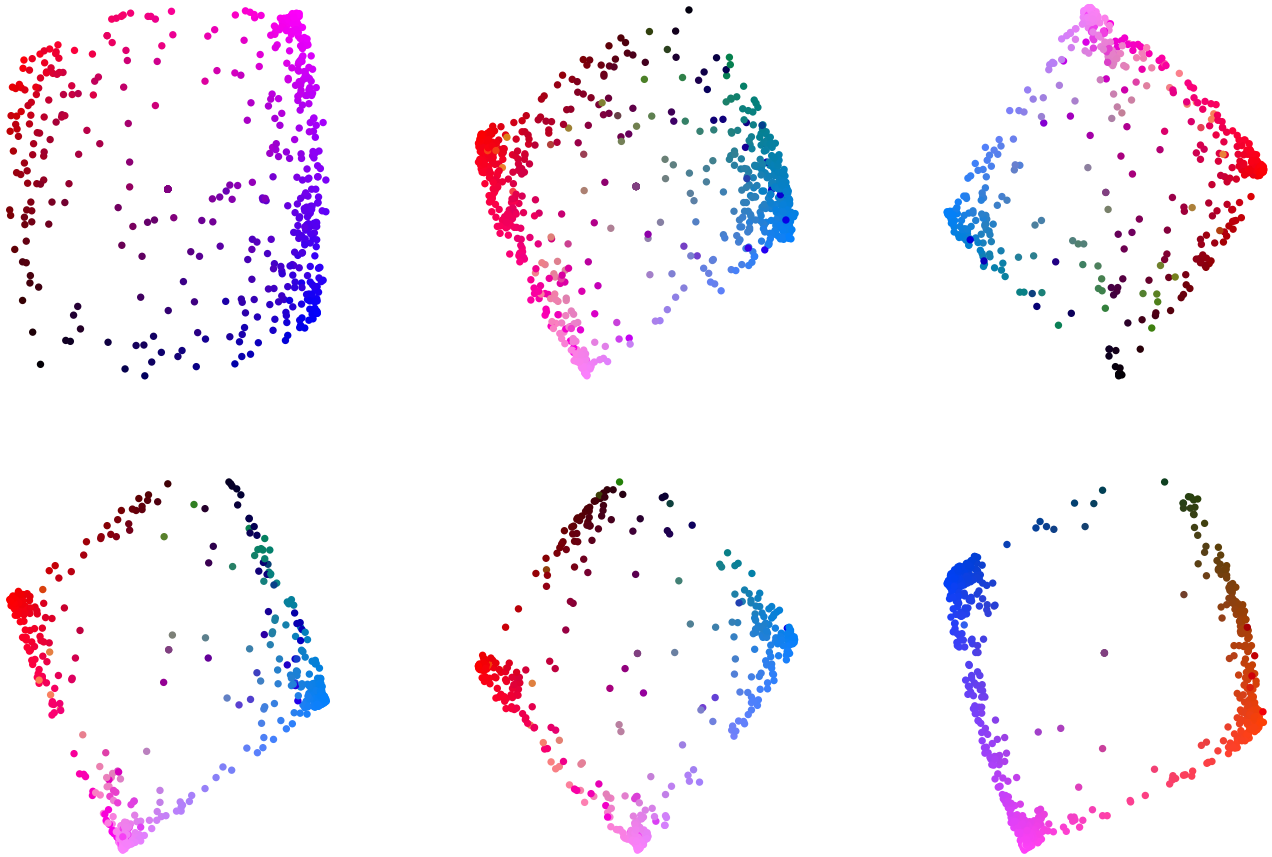
Parametrizing rollcalls

We can do the same analysis on the rollcalls instead of the members.

House of Representatives 2009... And the Senate...

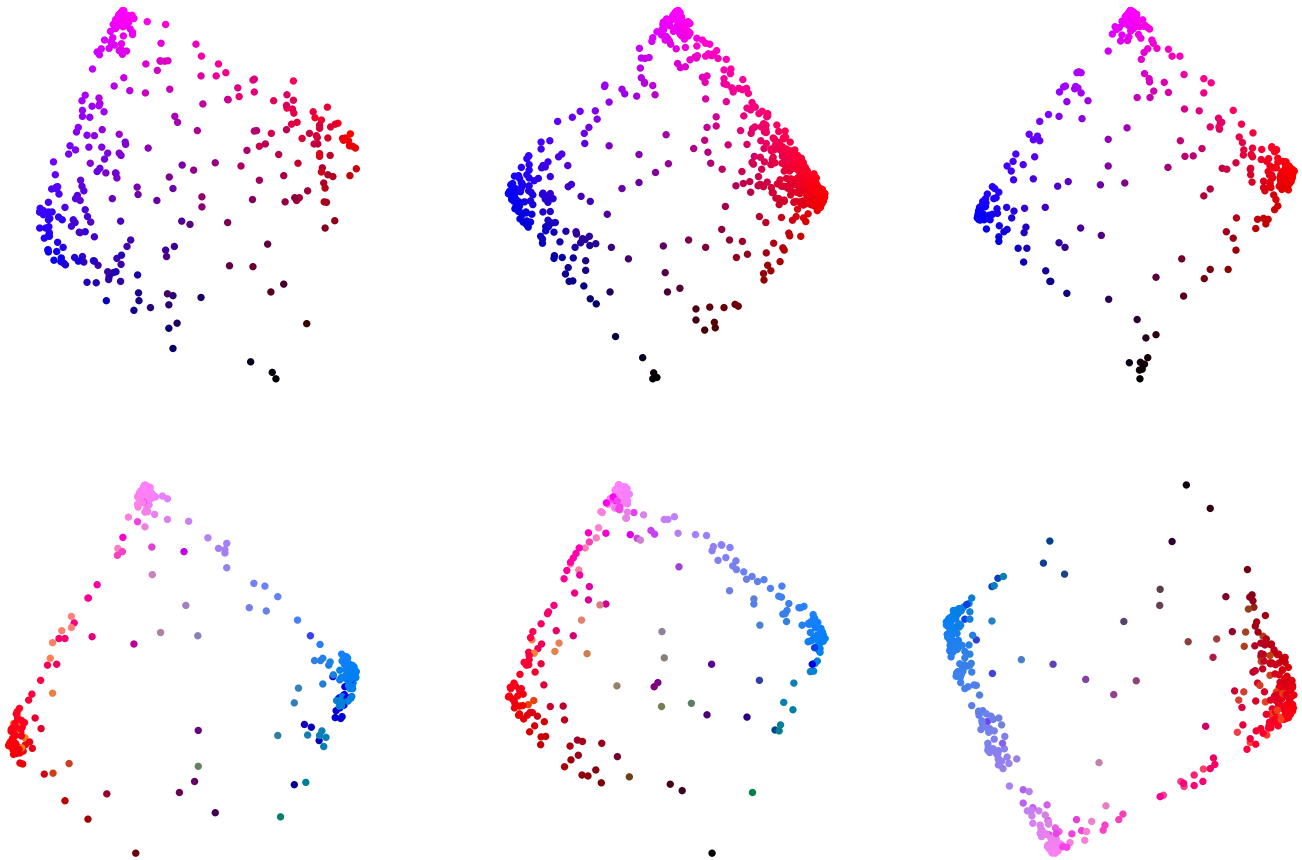


Again, the pattern persists



Left to right, top down: House of Representatives 1990, 1995, 1999, 2003, 2006, 2009.

Again, the pattern persists



Left to right, top down: Senate
1990, 1995, 1999, 2003, 2006, 2009.