Some notes on mean-median & partisan bias scores, and building **seats-votes curves** using lower house state legislative election results in the USA from 1971-2018. Also a place to organize some different non-geographical approaches to identifying partisan gerrymandering – eg – Warrington (2019); Warrington (2018); Gelman and King (1994); Katz, King, and Rosenblatt (2020).

# State legislative election results

library(tidyverse)

election\_results <- read.csv(url(git\_url)) %>% janitor::clean\_names() %>% mutate(d\_voteshare = round(d\_voteshare, 3))

A sample of the data set is detailed below. Results are presented year, state, and state house district; the

d\_voteshare column specifies the vote share received by the Democratic candidate.

election\_results %>%

filter(state == 'CA', year == 1980) %>% slice(1:5) %>%

select(-incumbent) %>% knitr::kable()

## state year district dem\_votes gop\_votes d\_voteshare party

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| CA | 1980 1 | 0 | 114547 | 0.000 R |
| CA | 1980 2 | 81884 | 36504 | 0.692 D |
| CA | 1980 3 | 55339 | 75547 | 0.423 R |
| CA | 1980 4 | 76013 | 35894 | 0.679 D |
| CA | 1980 5 | 52121 | 75998 | 0.407 R |

A super simple imputation method: Per approach described in Gelman and King (1994), winning parties of uncontested elections are re-assigned a vote share of 0.75, and losing parties 0.25.

election\_results1 <- election\_results %>%

mutate(d\_voteshare = ifelse(d\_voteshare == 1, 0.75, d\_voteshare), d\_voteshare = ifelse(d\_voteshare == 0, 0.25, d\_voteshare))

# Summarizing election results

Next, we summarize election results per legislature. Summary stats include:

the number of seats in legislature,

the number/proportion of seats won by Democrats,

the average vote share received by Democratic candidates; and the median Democratic vote share.

Important to emphasize here is that the v\_mean value specifies the average Democratic vote share across individual state house races in a given year, and not the aggregate statewide house results.

full\_summary <- election\_results1 %>% group\_by(state, year) %>%

mutate(dseat = ifelse(dem\_votes > gop\_votes, 1, 0), rseat = ifelse(dem\_votes < gop\_votes, 1, 0), d\_above = ifelse(d\_voteshare > mean(d\_voteshare),

1, 0)) %>%

summarize(district\_n = n(),

d\_seats = sum(dseat), r\_seats = sum(rseat), v\_mean = mean(d\_voteshare),

v\_median = median(d\_voteshare), d\_above = sum(d\_above)) %>%

mutate(seat\_bar = d\_seats/district\_n) %>% ungroup()

**Election results for Colorado during the 2010s** are presented below. So, Dems took the majority – and then some – in the Colorado State House during the previous decade.

full\_summary %>% filter(state == 'CO',

year > 2008) %>% select(-r\_seats) %>%

mutate(across(c(v\_mean, v\_median, seat\_bar), ~round(., 2))) %>% knitr::kable()

## state year district\_n d\_seats v\_mean v\_median d\_above seat\_bar

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| CO | 2010 | 65 | 32 | 0.48 | 0.50 | 35 | 0.49 |
| CO | 2012 | 65 | 37 | 0.51 | 0.54 | 37 | 0.57 |
| CO | 2014 | 65 | 34 | 0.48 | 0.51 | 37 | 0.52 |
| CO | 2016 | 65 | 37 | 0.50 | 0.54 | 37 | 0.57 |
| CO | 2018 | 65 | 41 | 0.55 | 0.59 | 36 | 0.63 |

**The plots below** illustrate the shifting partisan balance for a selection of state houses since 1972.

south <- c('CO', 'FL', 'AL', 'TX',

'AR', 'TN', 'OK', 'KY')

full\_summary %>%

filter(state %in% south) %>% ggplot() +

geom\_line(aes(x = year,

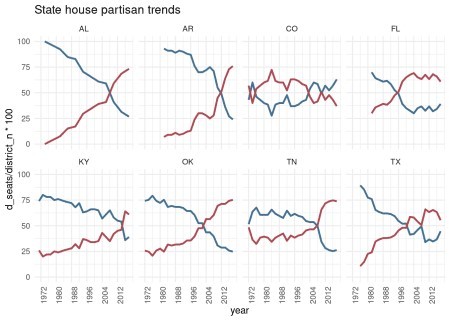
y = d\_seats/district\_n \* 100), color = '#437193', size = 1) +

geom\_line(aes(x = year,

y = r\_seats/district\_n \* 100), color = '#ae4952', size = 1) +

facet\_wrap(~state, ncol = 4) + theme\_minimal() +

theme(axis.text.x = element\_text(angle = 90, hjust = 1)) + scale\_x\_continuous(breaks=seq(1972, 2018, 8)) + ggtitle('State house partisan trends')



# Historical vote distributions

details1 <- election\_results1 %>% #filter(party %in% c('D', 'R')) %>%

left\_join(full\_summary) %>% group\_by(state, year) %>%

mutate(swing = 0.5 + v\_mean - d\_voteshare,

rank = rank(d\_voteshare, ties.method = 'first'), seat\_share = rank/n(),

seat\_share = 1 - seat\_share) %>%

## still not correct exactly --

mutate(swing = ifelse(seat\_share == seat\_bar, v\_mean, swing)) %>% arrange(seat\_share) %>%

ungroup()

Vote distributions for election results in Wisconsin since 1972 are illustrated below. Districts have been sorted in increasing order of **Democratic vote share**.

details1 %>%

filter(state == 'WI') %>%

ggplot() +

geom\_point(aes(x = factor(rank),

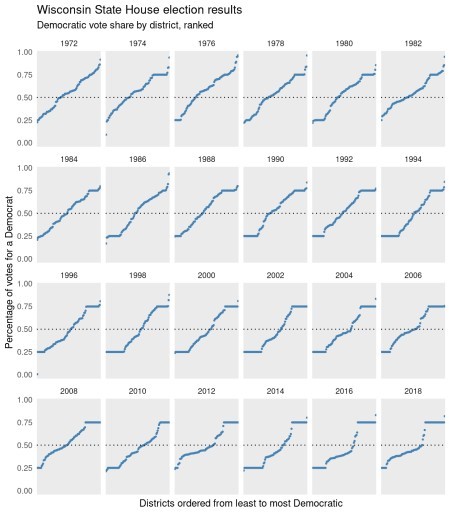
y = d\_voteshare), color = 'steelblue', size = .5) +

geom\_hline(yintercept = 0.5, lty = 3) + facet\_wrap(~year, ncol = 6) + theme\_minimal() + theme(axis.text.x=element\_blank()) +

xlab('Districts ordered from least to most Democratic') +

ylab('Percentage of votes for a Democrat') + labs(title = 'Wisconsin State House election results',

subtitle = 'Democratic vote share by district, ranked')



# Seats-votes curves

There are a host of metrics that aim to capture partisan asymmetries in vote distributions (see Warrington 2019 for a comparison). Here, we focus on mean-median scores and partisan bias scores, mainly because they are closely tied to the seats-votes curve.

The **mean-median score** is the difference between a party’s median *vote share* and its mean *vote share*

– divergence between these two values suggests a vote distribution that is skewed in favor of a particular party. In contrast, the **partisan bias score** is the difference between (1) a party’s actual *seat share* and

(2) that party’s hypothetical *seat share* if it garnered 50% of the statewide *vote share*. Both metrics are calculated below:

full\_summary1 <- full\_summary %>% mutate(mm = 0.5 + v\_mean - v\_median,

pb = (d\_above - 1) /district\_n) ##

As an example, we consider results from the **Wisconsin State House in 2018**. Again, results are presented from the perspective of Democrats.

xmm <- full\_summary1 %>% filter(state == 'WI', year == 2018)

## district\_n d\_seats v\_mean v\_median d\_above seat\_bar mm pb

**district\_n d\_seats v\_mean v\_median d\_above seat\_bar mm pb**

99 36 0.51 0.44 35 0.36 0.57 0.34

Per plot below, the green bar specifies the mean-median value; the red bar specifies the partisan bias score. So, if a seats-votes curve populates quadrant I, Democrats are over-represented in the legislature based on their statewide vote share; quadrant III, under-represented. The star specifies actual election results. **Extreme values in either quadrant are symptomatic of gerrymandering**.

see <- details1 %>%

filter(state == 'WI', year == 2018)

see %>% ggplot() +

geom\_hline(yintercept = .50) + geom\_vline(xintercept = .50) + geom\_step(aes(x = swing,

y = seat\_share,

color = factor(year)), size = 1) +

geom\_point(aes(x = v\_mean,

y = seat\_bar), pch="\u2605", size = 4) +

annotate('segment',

x = 0.5,

y = xmm$pb, xend = 0.5,

yend = 0.5,

color = '#913a40', size = 3, alpha = .5) + annotate('segment',

x = xmm$mm, y = 0.5,

xend = 0.5,

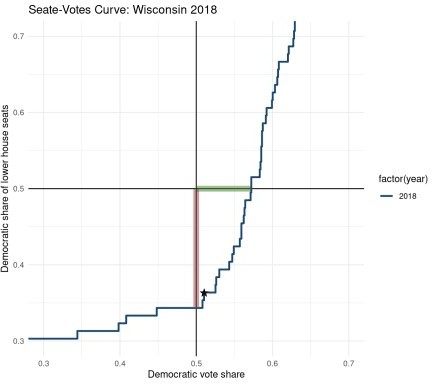
yend = 0.5,

color = '#3c811a', size = 3, alpha = .5) +

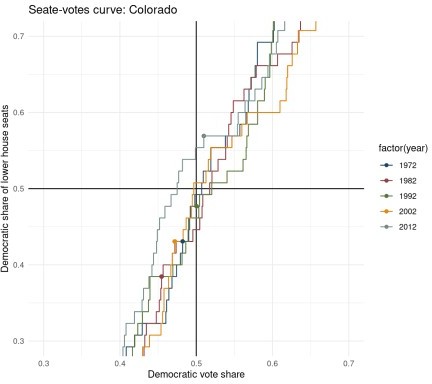
theme\_minimal() + theme(legend.position = 'right')+ ggthemes::scale\_color\_stata() + coord\_equal(xlim = c(0.3, 0.7),

ylim = c(0.3, 0.7)) +

ggtitle('Seate-Votes Curve: Wisconsin 2018') + ylab('Democratic share of lower house seats') + xlab('Democratic vote share')



**A historical example** from the state of Colorado –



# Resources

Gelman, Andrew, and Gary King. 1994. “A Unified Method of Evaluating Electoral Systems and Redistricting Plans.” *American Journal of Political Science*, 514–54.