

Multi Member District Analysis

Ian McDonald 14-Feb-2018

This notebook includes R code under development for my analysis of multi member districts in U.S. state legislatures. The description covers the R code and does not include background on the project's research questions or hypotheses.

The script adapts a dataset published by the Harvard Dataverse from Shor and McCarty (2011, 2015), and generates two preliminary lattice graphs. The dataset provides estimates of state legislator ideology scores, converted to a common scale. The data uses legislator roll calls analogous to DW-NOMINATE scores for U.S. Congress members, popularized by the work of Poole and Rosenthal. Shor and McCarty normalize these scores into a single scale using results from Project Votesmart National Political Awareness Test (NPAT), and then consolidate scores across multiple sessions to report a single career score for each state legislator (including both upper and lower houses).

I use data reported from 2015 covering 1993-2014.

The code applies Tidyverse modules, including dplyr, ggplot2, and purrr.

```
suppressMessages(library("tidyverse"))
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':
##
##      date
```

```
library(stringr)
```

We import the raw data file into the npat_june_2015 table. The raw table contains over 90 columns; npat_master contains only legislator attributes without columns labeled by year.

```
source_data <- "shor mccarty 1993-2014 state individual legislator scores public June 2015.tab"

npat_june_2015 <- source_data %>%
  read_delim(delim="\t", escape_double=FALSE, col_types=cols()) %>%
  mutate(member_id = paste0(st, sprintf("%04d", st_id)))

if (nrow(npat_june_2015) != length(unique(npat_june_2015$member_id))) {
  message("Error: There is a duplicate member_id somewhere in npat_june_2015")
}

npat_master <- npat_june_2015 %>%
  select(name, party, st, member_id, np_score)
```

Import data describing state legislatures, and create a vector with the states identified as multi-member (ten of 50 states).

```
#https://ballotpedia.org/State_legislative_chambers_that_use_multi-member_districts

leg_counts <- read_csv("district_numbers.csv", col_types=cols())

state_legislatures <- leg_counts %>%
  filter(double == TRUE) %>%
```

```
select(stcd)

state legislatures <- map_chr(state legislatures[[1]], as.character)
```

Identify the range of years in source data file to analyze. The raw data is organized into four ranges of columns named with explicit year labels. The column names are read and years. The gather function in the dplyr packages performs a similar operation but this custom routine works efficiently.

Note the use of map_dbl, one of the variations of map functions from the purrr package. This function, and others like it, replace the apply functions from base R.

```
fields <- names(npat_june_2015[6:93])
years <- c(1993:2014)
string_sections <- c("senate", "house", "sdistrict", "hdistrict")

make_field_name <- function(year, field_names, string_section) {

  s <- paste0(string_section, as.character(year))
  x <- str_detect(s, field_names)
  which(x)
}

make_f1 <- map_dbl(years, make_field_name, field_names = fields, string_section = "sdistrict")
make_f2 <- map_dbl(years, make_field_name, field_names = fields, string_section = "hdistrict")
make_f3 <- map_dbl(years, make_field_name, field_names = fields, string_section = "senate")
make_f4 <- map_dbl(years, make_field_name, field_names = fields, string_section = "house")

field_names <- bind_cols(year = years,
  senate_year = fields[make_f1],
  house_year = fields[make_f2],
  senate_districts = fields[make_f3],
  house_districts = fields[make_f4])

generate_chamber_table <- function(year, chamber="lower") {

  if (chamber == "lower") {
    district_field <- paste0("hdistrict", year)
    district_flag <- paste0("house", year)
  } else if (chamber == "upper") {
    district_field <- paste0("sdistrict", year)
    district_flag <- paste0("senate", year)
  } else {
    stop("what is the chamber?")
  }

  chamber <- npat_june_2015 %>%
    drop_na(!!district_flag) %>%
    rename(district = !!district_field) %>%
    mutate(district = paste0(st, "_", district)) %>%
    select(member_id, party, st, district, np_score) %>%
    mutate(year = year)
}

lower <- map_dfr(years, generate_chamber_table, chamber = "lower")
```

```
upper <- map_dfr(years, generate_chamber_table, chamber = "upper")
```

In this analysis, I want to consider the spread of NPAT scores in districts that elect at least one member from each party (Democrat or Republican). I suspect that the spread has grown since 1994. The analysis is complicated by the differences in electoral procedures and setup by state.

The script generates two graphs latticed by state:

1. The difference in the average NPAT score spread of districts with both parties represented.
2. The proportion of districts with both parties, vs. districts where all representatives come from the same party. In states with multi-member districts, the districts tend to elect representatives from one party or the other, but 16% send representatives from both (a surprisingly high percentage).

```
overall_mean_scores <- lower %>%
  group_by(st, year) %>%
  summarise(mean = mean(np_score)) %>%
  spread(year, mean)

# relabel Washington State and Idaho districts from WA_XXX-0[1 or 2] to WA_XXX
lower <- lower %>%
  mutate(district = if_else(st == "WA" | st == "ID", substr(district, 1, 6), district))

# include only states appearing in the state legislatures vector.
lower <- lower %>%
  filter(st %in% state legislatures)

upper <- upper %>%
  filter(st %in% state legislatures)

double_dists <- lower %>%
  group_by(district, year) %>%
  summarise(freq = n()) %>%
  filter(freq > 1)

lower <- inner_join(lower, double_dists, messages = FALSE)

## Joining, by = c("district", "year")

#split districts are those with at least one D and one R in the same term
split_districts <- lower %>%
  group_by(district, year) %>%
  count(party) %>%
  spread(party, n) %>%
  mutate(D = if_else(is.na(D), 0, as.double(D))) %>%
  mutate(I = if_else(is.na(I), 0, as.double(I))) %>%
  mutate(R = if_else(is.na(R), 0, as.double(R))) %>%
  mutate(split = if_else(D >= 1 & R >= 1, TRUE, FALSE)) %>%
  mutate(split_label = if_else(split==TRUE, "2 Party Dists", "1 Party Dists")) %>%
  mutate(state = substr(district,1,2))
```

The two graphs are generated by this code:

```
#compute means by year, state, and split status
state_group <- group_by(lower, district, year)

state_means_and_range <- state_group %>%
```

```

    summarise(mean=mean(np_score), max=max(np_score), min=min(np_score))

state_means_and_range <- inner_join(split_districts, state_means_and_range, by=c("district","year"))

state_means_and_range <- inner_join(leg_counts, state_means_and_range, by=c("stcd"="state"))

state_means_and_range<- state_means_and_range %>%
  mutate(range = max - min) %>%
  select(-one_of(c("upper", "upper_term", "lower_term")))

anal <- state_means_and_range %>%
  group_by(stcd, year) %>%
  mutate(denom = n()) %>%

  group_by(split, split_label, add=TRUE) %>%
  mutate(range = mean(range), count = n()) %>%
  mutate(spct = count / denom) %>%
  mutate(ss = if_else(split==TRUE,
    paste0(stcd,"-", "2 Party Dists"),
    paste0(stcd,"-", "1 Party Dists")))

```

The states shown have multi-member districts. The solid line represents districts where each representative comes from the same party. The dashed lines are districts with one (or more) representative from each party.

```

title_1 <- "Multi Member District State Legislatures"
subtitle_1 <- "Differences between Difference between Max and Min Average NPAT Score"
caption_1 <- "Solid Line = Members from Same Party, Dashed Line = Members from 2+ parties"

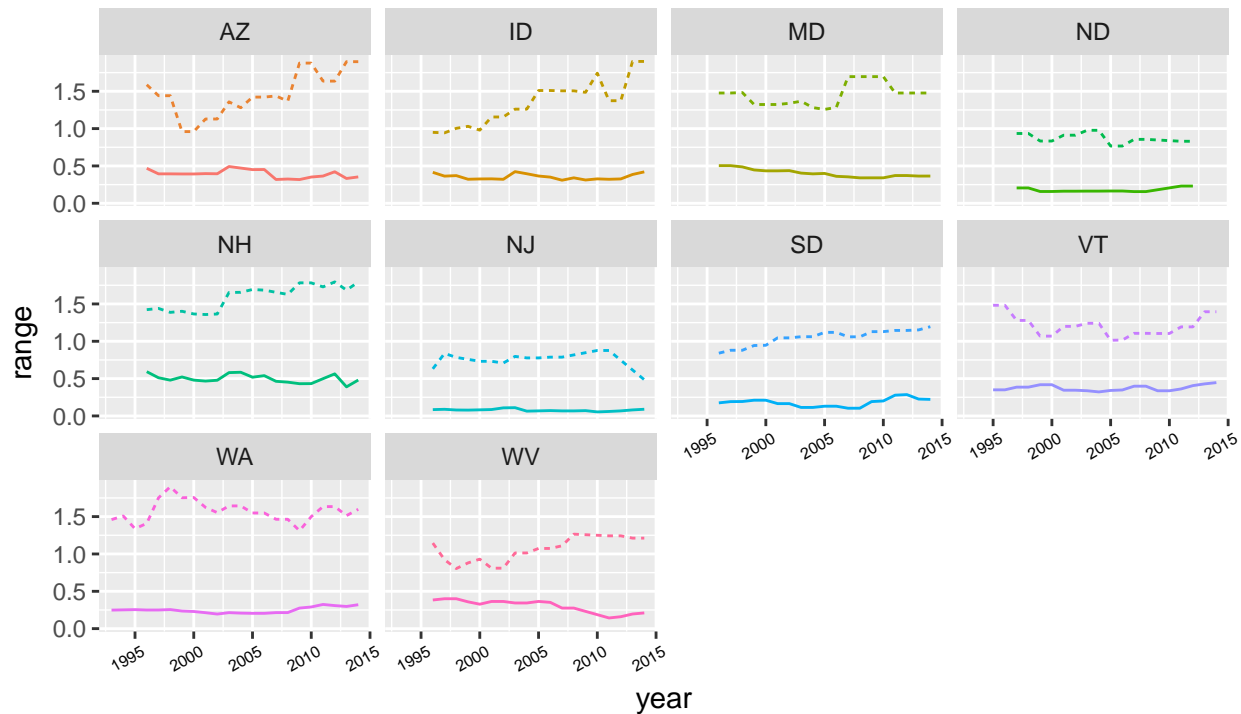
mmd_plot_1 <- ggplot(data=anal, aes(x=year, y=range, color=ss)) +
  #ylab("Difference between Max and Min Average NPAT Score") +
  labs(title = title_1, subtitle = subtitle_1, caption = caption_1) +
  geom_line(aes(linetype=split_label), show.legend=FALSE) +
  facet_wrap(~stcd)+
  theme(axis.text.x=element_text(color = "black", size=6, angle=30, vjust=.8, hjust=0.8))

mmd_plot_1

```

Multi Member District State Legislatures

Differences between Difference between Max and Min Average NPAT Score



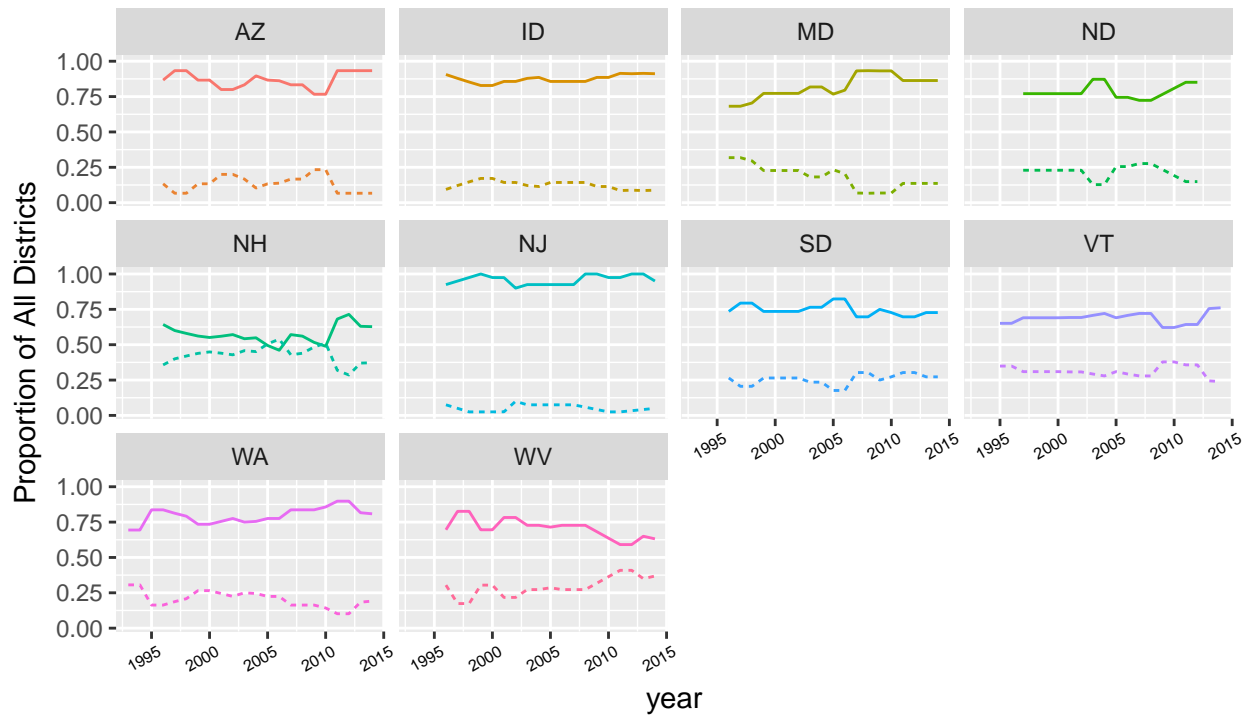
```
save.image(file = "State Means and Ranges.png")
```

```
title_2 <- "Multi Member District State Legislatures"
subtitle_2 <- "Proportion of Districts with One Party v. Two Party"
caption_2 <- "Solid Line = Members from Same Party, Dashed Line = Members from 2+ parties"
mmd_plot_2 <- ggplot(data=anal, aes(x=year, y=spct, color=ss)) +
  ylab("Proportion of All Districts") +
  labs(title = title_2, subtitle = subtitle_2, caption = caption_2) +
  geom_line(aes(linetype=split_label), show.legend=FALSE) +
  facet_wrap(~stcd) +
  theme(axis.text.x=element_text(color = "black", size=6, angle=30, vjust=.8, hjust=0.8))
```

```
mmd_plot_2
```

Multi Member District State Legislatures

Proportion of Districts with One Party v. Two Party



Solid Line = Members from Same Party, Dashed Line = Members from 2+ parties

```
save.image(file = "Proportion of Districts with One Party v. Two Party")
```