**The Reed-Smith Japanese House of Representatives Elections Data Set (JHRED)**

Steven R. Reed

Faculty of Policy Studies

Chuo University

Daniel M. Smith

Department of Government

Harvard University

**CITATION**

Reed, Steven R., and Daniel M. Smith. *The Reed-Smith Japanese House of Representatives Elections Data Set*. Version: April 19, 2018. URL: https://sites.google.com/site/danielmarkhamsmith/data.

**DESCRIPTION**

This panel data set includes every single candidate who ran in any general election or by-election for the Japanese House of Representatives from 1947 to 2017. The data set includes a total of 28,740 observations, i.e., candidate-elections, for 10,434 unique individuals, across 26 general elections.

From 1947-1993, candidates competed for votes in multi-member districts (MMD) under the single non-transferable vote (SNTV) electoral system. Since electoral reform in 1994, candidates in the new mixed-member majoritarian (MMM) system compete in one of 300 (reduced to 295 in 2014) single-member districts (SMD), on a party list in one of 11 regional proportional representation (PR) blocs, or dual-listed in both tiers. Electoral variables in the data set that refer to SNTV/MMD or SMD election districts are denoted with “ku” (“district” in Japanese) whereas PR district variables are denoted with “pr.”

The data set is based on three separate data sets originally collected by Steven R. Reed, and then updated, expanded, and cleaned by Daniel M. Smith. The basic candidate information and electoral results were compiled from *Asahi*, *Yomiuri*, *Mainichi*, and other newspaper records, as well as official election statistics from the Japanese Ministry of Internal Affairs and Communications (*Sōmushō*). Some candidate background information was compiled from yearly almanacs such as *Seikan Yōran*, *Seiji Handbook*, *Kokkai Binran*, candidate website profiles, *Za Senkyo*, and *Wikipedia*. Other sources vary, as noted, for specific variables or groups of variables. We thank Yusaku Horiuchi, Kuniaki Nemoto, Amy Catalinac, and Petter Lindgren for their help in cleaning these data; Yusaku Horiuchi, Ko Maeda, Masaru Kohno, Benjamin Nyblade, Gary Cox, and Mike Thies for providing some of the data for specific variables or years; and Shiro Kuriwaki, Anthony Volk, and Eric Xiao for research assistance.

If you find a mistake in the data, please let us know: danielmarkhamsmith@gmail.com.

*Note on merging data:*Because the Japanese characters used in names can vary across data sources, merging our data with other data sources based on the names of candidates can sometimes be difficult, and may require some manual merging of observations that do not match up. The personal identification codes in this updated version of the Reed-Smith data set differ from past versions due to corrected mistakes and the addition of new candidates. If you wish to merge this data set with older versions, we recommend merging on **KUCODE**, **YR** (keeping by-elections in mind), and **VOTE**, and then crosschecking **NAME\_JP**. In the entire data set, there is only one instance of two candidates in the same district at the same election receiving the exact same number of votes (Gunma 1st District in 1952).

**CANDIDATE IDENTIFICATION VARIABLES**

**PID** = a unique, five-digit personal identification code for each individual candidate, assigned alphabetically based on **NAME\_JP**.

**NAME\_JP** (string) = the name of the candidate in Japanese (kanji, unless written in kana in election records). When two or more candidates have the exact same name, they are distinguished with numbers after the name. Note: some name characters may differ slightly from other sources due to variation in how characters can be written. For candidates who used different names, variations in kanji/kana, or aliases over time, we have picked a single name and kept it constant across observations.

**ELECTORAL VARIABLES**

**LEGIS** = the number of the legislature for which the election was held, beginning with the 23rd House of Representatives elected in 1947. By-elections are grouped with the legislature at the time of the election.

**YEAR** = the year of the election.

**YR** = the year of election coded as consecutive numbers, 1 (1947) through 25 (2014). By-elections are given half-year values, e.g., 1.5, 2.5, etc. All by-elections that were held between two general elections are given the same **YR** value, with the exception of Miyagi 6th District, which held two by-elections between the 1996 and 2000 general elections (1997 and 2000), so **YR** is coded 19.5 and 19.6, respectively.

**KEN** (string) = the name of the prefecture (*ken*) containing the district (*ku*), in English. There are 47 prefectures. The numbers below correspond to the base for each district code (**KUCODE**).

NA = the candidate did not run in a district race (e.g., a pure PR candidate).

1: Hokkaidō

2: Aomori

3: Iwate

4: Miyagi

5: Akita

6: Yamagata

7: Fukushima

8: Ibaraki

9: Tochigi

10: Gunma

11: Saitama

12: Chiba

13: Tōkyō

14: Kanagawa

15: Niigata

16: Toyama

17: Ishikawa

18: Fukui

19: Yamanashi

20: Nagano

21: Gifu

22: Shizuoka

23: Aichi

24: Mie

25: Shiga

26: Kyōto

27: Ōsaka

28: Hyōgo

29: Nara

30: Wakayama

31: Tottori

32: Shimane

33: Okayama

34: Hiroshima

35: Yamaguchi

36: Tokushima

37: Kagawa

38: Ehime

39: Kōchi

40: Fukuoka

41: Saga

42: Nagasaki

43: Kumamoto

44: Ōita

45: Miyazaki

46: Kagoshima

47. Okinawa

**KUNR** = the number of the district within the prefecture. 0: a pure PR candidate.

**KUNAME** (string) = the name of the district, formed as **KEN** + **KUNR**, e.g., “Hokkaido 1.” This is the official name of the district, and as with **KUCODE**, does not change when redistricting. For pure PR candidates, this variable is the name of the PR bloc, formed as **REGION** + bloc, e.g., “Hokkaido bloc.”

**KUCODE** = the number of the district within the prefecture. Each code is created using the base code for each district (see the list below **KEN**) + **KUNR**. 0: a pure PR candidate.

**KUCODER** = same as **KUCODE**, however, when a district is split into two new districts, one of the new districts retains the original district number. For example, in 1967 Tokyo 1st was split into a new Tokyo 1st and Tokyo 8th districts. We thus code **KUCODER** for Tokyo 1st as 1301 from 1947 through 1963 but as 1301.1 from 1967 through 1993. Other divided districts are handled analogously. One district was formed from a merger of two districts. In 1993, Kagoshima 1st was formed from the old Kagoshima 1st district and the old Amami single-member district. The code for Kagoshima 1st thus becomes 4601.1 in 1993. In a few cases, such as Hokkaido 13th, the name of the district changed (to Hokkaido 7th) but the boundaries did not. In these cases, **KUCODER** remains unchanged even as **KUCODE** changes. **KUCODER** resets following electoral reform, when many district boundaries changed, so **KUCODER** from 1996 should not be considered equal to pre-1996 **KUCODER**, even if the code does not change. Redistricting occurred again in 2003 and 2014, and **KUCODER** again includes .1 or .2 in some cases after the base code for affected districts. 0: a pure PR candidate.

**REGION** (string) = the name of the region containing the district, in English; also corresponds to the eleven post-1994 PR blocs.

**PRCODE** = the number of the PR regional bloc (see **REGION** above). 0: a pure district (MMD or SMD) candidate.

10000: *Hokkaido*

20000: *Tohoku*: Aomori; Iwate; Miyagi; Akita; Yamagata; Fukushima

30000: *Kita Kanto*: Ibaraki; Tochigi; Gunma; Saitama

40000: *Tokyo*

50000: *Minami Kanto*: Chiba; Kanagawa; Yamanashi

60000: *Hokuriku Shinetsu*: Niigata; Toyama; Ishikawa; Fukui; Nagano

70000: *Tokai*: Gifu; Shizuoka; Aichi; Mie

80000: *Kinki*: Shiga; Kyoto; Osaka; Hyogo; Nara; Wakayama

90000: *Chugoku*: Tottori; Shimane; Okayama; Hiroshima; Yamaguchi

100000: *Shikoku*: Tokushima; Kagawa; Ehime; Kochi

110000: *Kyushu*: Fukuoka; Saga; Nagasaki; Kumamoto; Oita; Miyazaki; Kagoshima; Okinawa

**RESULT** = the candidate’s election result.

0: Lost the election.

1: Was elected in MMD or SMD in a general election.

2: Ran in SMD but lost; was elected via PR list (a “zombie”).

3: Was elected as a pure PR list candidate.

4: Was elected midterm due to the resignation or death of sitting MP, either because the candidate was the runner-up within 3 months of the election under SNTV, or was in the next position on the party list under MMM (so-called *kuriage tōsen*).

5: Was elected in a by-election.

**INC** = the candidate’s incumbency status.

Note: this variable codes only parliamentary incumbency, and does not take into account whether the candidate was an incumbent in the same district in which s/he ran.

0: Not an incumbent; did not win the by-election; was not promoted through the PR list or within 3 months of the election due to a resignation.

1: Won in MMD or SMD in the last general election; did not resign in the midterm.

2: Ran in SMD but lost; won a PR seat in the last election as a so-called “zombie.”

3: Won a PR seat in the last election as a pure PR candidate.

4: Promoted in the midterm through the PR list, or in cases of exit within 3 months under SNTV electoral system, due to a death or resignation.

5: Won a by-election; including zombie incumbents who resigned to win in the SMD.

6: Won the last election but resigned in the midterm (does not include those running in a by-election who resigned a seat in PR to run in the SMD).

7: Incumbent from a different district who changed districts (e.g., Oota Kazumi, Yoshino Masayoshi, etc.). This does not apply to incumbents whose geographical base didn’t change, but district number did (i.e., due to redistricting). In such cases (e.g., Hosono Goshi in 2003), and in 1996, there can be two district incumbents (**INC** = 1).

**PARTY\_JP** = the candidate’s party affiliation at the time of the election, in Japanese.

**PARTY\_EN** = the candidate’s party affiliation at the time of the election, in English.

**PARTY\_ID** = the candidate’s party affiliation code. See the **PARTY APPENDIX** below for a complete list.

**KU\_VOTE** = the raw number of valid votes won by the candidate in the SNTV/SMD district.

**KU\_TOTVOTE** = the total number of valid votes cast in the SNTV/SMD district in that election. The value of this variable in elections pre-1958 may not be the same as that obtained by adding the votes of all the candidates in the district because of the presence of fringe candidates who are not included in the data set.

Note: A candidate’s vote share can be calculated with **KU\_VOTE / KU\_TOTVOTE**

**KU\_ELECTORATE** = the total number of eligible voters in the SNTV/SMD district for that election.

Note: Dividing **KU\_TOTVOTE** by **KU\_ELECTORATE** will yield slightly different figures from official turnout because official figures include invalid votes.

**KU\_RANK** = the rank order of the candidates in the SNTV/SMD district.

**KU\_M** = district magnitude; the number of Diet members elected in the SNTV/SMD district.

**KU\_ NCAND** = the raw number of candidates in the SNTV/SMD district.

**PR\_PARTYVOTE** = the raw number of valid votes won by the candidate’s party in the PR district.

**PR\_TOTVOTE** = the total number of valid votes cast for all parties in the PR district in that election.

Note: the vote share of the party in the PR district can be calculated with **PR\_PARTYVOTE / PR\_TOTVOTE**

**PR\_ELECTORATE** = the total number of eligible voters in the PR district for that election.

Note: Dividing **PR\_TOTVOTE** by **PR\_ELECTORATE** will yield slightly different figures from official turnout because official figures include invalid votes.

**PR\_M** = district magnitude; the number of Diet members elected from the PR district.

**PR\_PARTYSEATS** = number of seats won by the party in the PR district.

**PR\_RANK** = the candidate’s initial position on the party list in the party PR district.

**PR\_NCAND** = the number of candidates on the party list in the party PR district.

**PR\_RERANK** = the candidate’s re-ranked position on the party list in the party PR district (after SMD candidates removed from the list). Coded 0 if dual-listed candidate removed from the list due to winning an SMD. Coded 999 if dual-listed candidate failed to qualify because of a low SMD performance (less than 1/10 of SMD vote). NOT YET CODED FOR 2017

**BYELECTION** = 0 if a general election, 1 if a by-election or a special election.

**BYCAUSE** (string) = the reason for the by-election.

**MTENTRY** = a dummy variable coded 1 if the candidate entered the legislature midterm (i.e., through a by-election).

**MTEXIT** = a dummy variable coded 1 if the candidate exited the legislature midterm (i.e., due to death or resignation).

**PREWRUNS** = the number of times the candidate ran in the prewar period.

**PREWWINS** = the number of times the candidate won in the prewar period.

**RUN1946** = coded 1 if the candidate ran in 1946. Note: may overlook some very fringe candidates.

**WIN1946** = coded 1 if the candidate was elected in 1946.

**TOTCRUNS** = the total number of times since 1947 (election time *t* inclusive) that the candidate has ever run in an election, including by-elections, and regardless of district or jiban.

Note: For the total number of prewar and postwar runs, add **TOTCRUNS** + **PREWRUNS + RUN1946**.

**TOTCWINS** = the total number of times since 1947 (election time *t* inclusive) that the candidate has ever won in an election, including by-elections, and regardless of district or jiban.

Note: For the total number of prewar, 1946, and postwar wins, you can add **TOTCWINS** + **PREWWINS + WIN1946**. This variable is NOT exactly the same as the total number of terms served because a few candidates (such as Machimura Nobutaka in Hokkaido 5th District in 2010) resigned their seats midterm and then were re-elected in a by-election. This must be calculated using **PID**, **RESULT**, and **LEGIS** after dropping losing candidates. Previous wins (election time *t* – 1) must also be calculated.

**DEMOGRAPHIC VARIABLES**

These data are based on municipality-level census data, but all variables are aggregated to the district-level. All inter-census year data are estimated values. Data for by-elections are estimated as the mean of the values of the preceding and following general election, or simply the values of the most recent election if it was held the same year.

Data up to 2004 come from: “Horiuchi, Yusaku and Masaru Kohno. 2004. Japanese Election District-Level Census Data (JEDC Data). Version: 23 April 2004. A machine-readable data file.” **POPDID** for 2014 kindly provided by Ko Maeda. Data are not available for all years.

**POPTOTAL** = the total population of the district.

**POPMALE** = the number of males in the district.

**POPFEMALE** = the number of females in the district.

**POPUNDER15** = the number of people under the age of 15 in the district.

**POPOVER65** = the number of people over the age of 65 in the district.

**POPDID** = the number of people living in census-defined “densely inhabited districts.”

**POPDENSITY** = the proportion of people living in census-defined “densely inhabited districts.” Equal to **POPDID**/**POPTOTAL**.

**WORK1ST** = the number of people in primary sector employment (agriculture, forestry, fishery, and mining industries).

**WORK2ND** = the number of people in secondary sector employment (manufacturing, construction, and public utility industries).

**WORK3RD** = the number of people in tertiary sector employment (all other industries, such as retail and services).

**CANDIDATE BACKGROUND VARIABLES**

Unless otherwise noted, each of these variables is a dummy variable coded 1 for candidates who have the relevant background characteristic. We used many sources in coding the data, including yearly almanacs and Wikipedia biographies, but many entries are based solely on newspaper summaries that are, in turn, based on candidate self-reports. The data should thus be considered somewhat noisy. The user should also note that an entry of 1 indicates that the candidate had the particular background characteristic but an entry of 0 does not necessarily indicate that the candidate did not have that characteristic.

**DEPURGEE** = the candidate had been purged by the US Occupation and was able to run again starting in a later election (most candidates returned in 1952).

**FEMALE** = the candidate is a female.

**BYEAR** = the year of birth of the candidate. Birth year is missing for one by-election candidate (Kubo Haruzou) in 1948.

**AGE** = the age of candidate at the time of the election. Some ages are estimated by subtracting the election year from the year of birth, and thus may be (at most) one year off. Age is missing for one by-election candidate (Kubo Haruzou) in 1948.

**CAMPAIGN FINANCE VARIABLES**

**EXP** = the amount (in yen) of the candidate’s declared campaign expenditures.

Note: Available only for nominal district candidates (no PR) in general elections from 1967 on. Data for 1967 through 1990 were provided by Gary Cox and Michael Thies.

**LIMIT** = the official limit (in yen) for campaign expenditures in that district for that election. Note: Available only for nominal districts (no PR) in general elections from 1967 on. Data for 1967 through 1990 were provided by Gary Cox and Michael Thies.

**TOTEXP** = the total amount of campaign expenditures declared by all candidates in that district for that year.

Note: Available only for nominal districts (no PR) in general elections from 1967 on. Data for 1967 through 1990 were provided by Gary Cox and Michael Thies.

**PARTY APPENDIX**

Codes for **PARTYID**:

1: Liberal Party (日本自由党/自由党/吉田派の自由党), Democratic Liberal Party (民主自由党), or Liberal Democratic Party (LDP, 自民党)

1.5: Liberal Democratic Party-affiliated independents (LDPI, 無所属（自民）). Liberal Party-affiliated independent (無所属（自由）)

2: Japan Socialist Party (日本社会党), the Left Socialist Party (左派社会党), the Social Democratic Party (社会民主党)

2.5: JSP-affiliated independents

3: Kōmeitō (公明党)

3.5: Kōmeitō-affiliated independents

4: Democratic Socialist Party (DSP, 民主社会党), or the Right Socialist Party (右派社会党)

4.5: DSP-affiliated independents and Okinawa Social Mass Party (沖縄社会大衆党)

5: Japan Communist Party (共産党)

5.5: JCP-affiliated independents and Okinawa Citizen’s Party (沖縄人民党)

6: The second largest conservative party before the formation of the LDP, i.e., Democrats (民主党1947/日本民主党), Reform Party (改進党)

7: National Cooperative Party (国民協同党)

8: Reform Socialist Party (社会革新党) and Cooperative Party (協同党)

9: Hatoyama Liberals (鳩山派の自由党)

10: Independent (無所属)

11: New Liberal Club (新自由クラブ)

11.5: NLC-affiliated independent

12: Sakigake (さきがけ)

12.5: Sakigake-affiliated independent

13: Japan New Party (日本新党)

13.5 JNP-affiliated independent

14: Renewal (新生党)

15: New Frontier Party, NFP (新進党)

15.5: NFP-affiliated independent

16: Democratic Party of Japan, DPJ (民主党), Democratic Party, DP (民進党)

16.5: DPJ or DP-affiliated independent

17: Ichirō Ozawa’s Liberal Party (自由党)

17.5: Liberal Party-affiliated independents

18: Liberal Alliance (自由連合).

19: Social Democratic League (社民連)

19.5: SDL-affiliated independent

20: Japan Farmers Party (日本農民党)

21: New Farmers Party (農民新党)

22: Labor-Farmer Party (労働者農民党)

23: Reconstruction Party (日本再建連盟).

24: Progressive Party (進歩党).

25: Conservative Party (保守党)

26: New Socialist Party (新社会党)

27: New Liberal Party (新自由党)

28: Mushozoku no Kai (無所属の会)

29: Reform Club (改革クラブ) ; New Renaissance Party, NRP (新党改革)

30: Democratic Reform League, DRL (民主改革連合)

31: People’s New Party, PNP (国民新党)

32: New Party Japan, NPJ (新党日本)

33: New Party Daichi (新党大地)

34: Your Party (みんなの党)

34.5 Your Party-affiliated independent

35: Japan Restoration Party, JRP (日本維新の会)・Japan Innovation Party, JIP (維新の党)・Initiatives for Osaka, IFO(おおさか維新の会)・Ishin no Kai （日本維新の会）

36: Tomorrow Party of Japan, TPJ (未来の党)・People’s Lives First Party生活の党・(new) Liberal Party of Ozawa (自由党)

37: Party of Hope (希望の党)

38: Happiness Realization Party (幸福実現党)

39: Party for Future Generations (次世代の党)・Party for Japanese Kokoro, PJK(日本のこころを大切にする党、日本のこころ)

40: Reform Club (改革の会) Hatoyama Kunio and Nishioka Takeo’s party in 1994; inter-electoral only

41: Constitutional Democratic Party. CDP (立憲民主党)

99: Minor (諸派or other minor party as noted in **PARTYJP**)