GUIDE TO DATA AND REPLICATION FILES for

Catalinac, Amy, 2025, *Dominance Through Division: Group-Based Clientelism in Japan*, New York, NY: Cambridge University Press.[1]

Please cite this as the replication data for Catalinac, Amy, 2025, *Dominance Through Division: Group-Based Clientelism in Japan*, New York, New York: Cambridge University Press.

TABLE OF CONTENTS

1. THE BASICS p. 1
2. WORKING WITH DATA IN JAPANESE p. 2
3. GUIDE TO RAW DATA p. 2
4. GUIDE TO MASTER DATA p. 15
5. GUIDE TO REPLICATION SCRIPTS FOR EMPIRICAL CHAPTERS p. 26
6. APPENDIX I p. 28
7. APPENDIX II p. 33
8. WORKS CITED p. 34

---

1. THE BASICS

This repository contains two main data sets. The first is the raw data we collected for all Japanese municipalities that existed in any form between 1980 and 2014. As I explain below, this entailed merging JED-M voting data with demographic and fiscal data from Nikkei NEEDs. The second is the master data, which is a data set created from the raw data, but with additional years of demographic and fiscal variables added and without municipalities that straddle more than one electoral district (I call these "split municipalities"). The "master data"

[1] I owe a great debt to the following individuals, who helped with data collection and coding: Shiro Kuriwaki, Kuniaki Nemoto, Lucia Motolinia, Tatsuya Koyama, and my co-authors on the 2020 paper, Bruce Bueno de Mesquita and Alastair Smith. I want to especially thank Shiro, who wrote a codebook for an earlier version of the raw data, some of which I draw on below.

contains all the variables used to run the analyses in my book, and thus, is the primary data set used in the replication R scripts. Below, I explain pertinent features of both data sets.

---

2. WORKING WITH DATA IN JAPANESE

Both data sets are large and have Japanese characters in the municipality name and candidate name columns. The replication R scripts also sometimes have Japanese characters in them. Thus, care is required when opening the data in R and Excel.[2]

I use Windows in Japanese, which means that my version of R allows me to read in and enter data with Japanese characters when I select the correct encoding (Shift-JIS). However, I cannot directly open most .csv files with Japanese characters in them. If I want to view a .csv file as an Excel Workbook, I have to open a blank Workbook and use the Data tab to read in a csv file, selecting the correct encoding (Shift-JIS).

To read in a .csv file with Japanese characters encoded in shift-JIS directly into R, you could use:

```
library(readr)

dat <- read_csv("out_final_edit.csv", locale = locale(encoding = "SHIFT-JIS"))
```

Alternatively, you could try:

```
dat <- read.csv("out_final_edit.csv", fileEncoding = "SHIFT-JIS")
```

Importantly, the R script to implement the analyses in Chapter 6 *must be opened with the correct encoding* (Shift-JIS). This is written at the top of the R script. This is because Japanese characters become unreadable if they are set to the incorrect encoding, so the code will not work. Also, if you save the script with the unreadable characters, it will not be readable later.

---

3. RAW DATA (“out_final_utf8.csv”)

This is the raw data we collected for all Japanese municipalities that existed in any form between 1980 and 2014. Specifically, it includes all municipality-years for which we have voting data in the twelve HOR elections held between 1980-2014 (even if we do not have fiscal and demographic variables for some of them), and all municipality-years for which we have fiscal

[2] If you open an R script without specifying the encoding in the script, the Japanese characters will be garbled. If you save the script with the garbled characters, the code is then lost and will not work.

and demographic variables in each year from 1980-2012 (even if we do not have voting data for some of them).

This data was made by combining two data sets. One is the JED-M voting data for all municipalities in the twelve Lower House elections held between 1980 and 2014.[3] Because we merged the JED-M voting data with data supplied from Nikkei NEEDs, which records municipality-level information in a given *fiscal* year (not a calendar year), the raw data (and master data) records variables in *fiscal years*. The fiscal year in Japan runs from April 1st until March 31 of the subsequent calendar year. Thus, Lower House elections were held in the fiscal years 1980, 1983, 1986, 1989 (note that the calendar year was 1990), 1993, 1996, 2000, 2003, 2005, 2009, 2012 and 2014.

The unit of analysis in the JED-M data is technically not municipality-year, but *municipality-in-electoral district-year,* because some municipalities straddle more than one electoral district.[4] We maintained this structure in the raw data: municipalities that are perfectly contained within a single electoral district (the vast majority) appear only once per (election) year (in one row in this dataset), but municipalities that straddle *two* electoral districts, for example, appear twice for that (election) year (in two rows). One row has the voting data for one electoral district, while the other has the voting data for the other electoral district. This means that a municipality can appear in one row in one election, and then in two rows in another, if it became a *split municipality* between the two elections.

I explain the second data source, from which I retrieved fiscal and demographic variables, in further detail below, but let me point out here that in the raw data, split municipalities (which straddle more than one electoral district) are matched to fiscal and demographic variables for the *entirety of the municipality*. Thus, a split municipality straddling two electoral districts in the 1996 election is in two rows in 1996. Each row contains *different voting data* (pertaining to the different electoral districts it is in), but the *same demographic and fiscal data* (which is duplicated). This is because municipality-level variables are only reported at the level of the municipality, not at the level of the municipality-within-electoral district.

---

[3] Most of the variables below come from the JED-M files with suffix "D": e.g. (SH48D, SH49D). These files record the number of votes cast for all candidates competing in the electoral district in which the municipality is located, where electoral districts are MMDs prior to 1994 and SMDs after 1994. Thus, for each election, the number of suffix "D' files pertain to the number of electoral districts that existed in that year. The PR-related variables below come from the JED-M files with suffix "HD" (SH48HD, SH49HD, etc.) These files record the number of votes cast for all parties presenting a list in the PR bloc in which the municipality is located for all municipalities in the electoral district (MMD/SMD) in which the municipality is located. Thus, for each election, the number of suffix "HD" files pertain to the number of electoral districts that existed in that year.

[4] Electoral districts are multi-member (中選挙区) from 1980-1993 and single-member（小選挙区）from 1996 until 2014.

Continuing with my description of the JED-M data, municipalities are recorded under the names they had at the time of the respective elections, and there is nothing to connect municipalities with earlier or later versions of themselves.[5] Thus, if a municipality changes its name between the 1980 and 1983 Lower House elections, for example, it will appear in the 1983 file under its new name, without any indicator connecting it to the version of itself that existed in 1980. While municipality name changes are rare, a municipality's status as a village, town, or city appears as Japanese characters at the end of its name (村、町、市). Because our data spans more than 30 years, we have cases in the data of municipalities changing their status, which results in a name change.

To create the raw data, we took all the observations for which we had JED-M data (Lower House elections 1980-2014), and searched for an observation that matched these municipalities in the second (fiscal and demographic) dataset (described below). Due to features of the fiscal and demographic data, there were municipalities in the JED-M data for which we could not find a match, meaning that we were unable to retrieve fiscal and demographic variables for them. For completeness, the raw data keeps the small number of JED-M observations for which we could not find a match in the fiscal/demographic data.

Note that in the period of municipal mergers, municipalities can have *voting data* for the 2003 or 2005 Lower House elections, but are not matched to fiscal or demographic variables for that fiscal year. This is because these municipalities ceased to exist within the fiscal year. Thus, the absence of these latter variables is not a case of failing to find a match for them.

A description of all voting-related variables in the raw data, most of which came from JED-M, are as follows. Note that these variables are only populated for Lower House election years. To make it easier to merge in other pertinent characteristics of candidates and electoral districts, some of the below variables were taken from Reed and Smith's JHRED data (version published in 2015). This is all indicated below:

- **elec** (source: JED-M)**:** official number of the Lower House election held (our time period corresponds from the 36th to 47th).

- **HOR_electoral_district** (source: JED-M). Official government code for the Lower House electoral district the municipality is located in.

    - This is the number given to the electoral district by the government and does not change when the borders of an electoral district change. Thus, Electoral District

[5] This is because the JED-M data appears to have been collected from prefectural electoral commissions directly at the time of each election. Japan has 47 prefectures.

101 in the 1993 election (the last election before electoral reform) is not the same as Electoral District 101 in 1996 (the first election after electoral reform, which entailed changes to the boundaries of almost all electoral districts). Later in this document, I explain how the master data uses **kucoder** from Reed and Smith (2015) to conduct a panel. This variable adds a decimal suffix every time the borders of an electoral district change.

    - **HOR_electoral_district** contains either three numbers or four. If three, the first number indicates the prefecture (1 through 9) and the latter two numbers indicate the district within the prefecture (01 through 25). If four, the first two numbers indicate the prefecture (10 through 47) and the latter two numbers indicate the district within the prefecture (01 through 25).

- **mun_name1_jed** (source: JED-M). This and **mun_name2_jed** are the way municipalities are identified in the JED-M data. The first part of the name is usually a "gun" (province).

- **mun_name2_jed** (source: JED-M). This and **mun_name1_jed** are the way municipalities are identified in the JED-M data. The second part of the name is the municipality name. Some municipalities have the same mun_name2_jed, so using both mun_name1_jed and mun_name2_jed to identify municipalities is advisable.

- **split_mun_jed** (source: JED-M): 1 if the municipality is a split municipality, meaning that its area is not perfectly contained within a single electoral district, 0 otherwise.

    - Note that this is the primary way split municipalities are identified in the data. If we take a municipality that is split into two electoral districts in 1990, for example, this means that it appears on two rows of the dataset for the same election year. Each observation will have the *same* values on demographic and fiscal variables, but *distinct* **HOR_electoral_district** codes and voting data.

- **contains_split_mun** (source: JED-M): 1 if the municipality's electoral district contains a split municipality, 0 otherwise.

- **mun_voting_pop** (source: JED-M) (有権者数): number of people eligible to vote in the electoral district tier in the Lower House election in the municipality. Note that electoral districts were MMDs until 1994 and SMDs thereafter. The electoral district tier was the *only tier* in elections until 1996, after which it was one of *two tiers* (the second one was a "proportional representation" (PR) tier).

- **mun_voted** (source: JED-M) (有効投票者数): number of valid votes cast for the electoral district race in municipality.

    - Note that this is *valid votes*, not total votes. Thus, we exclude spoiled ballots. This was calculated by taking the sum of the vote tallies received by all candidates who were contesting the electoral district race in the municipality.

- **mun_turnout**: **mun voted/mun voting pop**

- **voting_pop_in_electoral_district (**source: Reed and Smith 2015; variable: **ku_electorate)**: number of eligible voters in the Lower House electoral district that houses the municipality. Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district.

- **voted_in_electoral_district (**source: Reed and Smith 2015; variable: **ku_totvote):** total number of valid votes cast in the electoral district housing the municipality. Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district.

- **turnout_in_electoral_district: voted_in_electoral_district/voting_pop_in_electoral_district.** Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district.

- **totmun_in_electoral_district** (source: JED-M): number of municipalities contained within the electoral district housing the municipality. Note that this number *includes* sections of a municipality that are located in the district. Thus, technically, it is a count of the number of vote-counting units (開票区) located within each electoral district. Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district.

- **ncands_electoral_district** (source: JED-M): number of candidates who ran in the electoral district housing the municipality. Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district.

- **totseat_in_electoral_district (**source: Reed and Smith 2015; variable: **ku_M)**: total number of seats available in the electoral district. Note that all electoral districts receive 1 for elections after 1994 as they were all single-member districts. Because this is an

electoral district-specific variable, it is identical for all municipalities within the electoral district.

- **cand_01 votes … cand_17_votes** (source: JED-M)**:** these variables record the number of votes each of the candidates running in the electoral district race in the Lower House election obtained in the municipality.

    - Critically, for all variables preceded by 'cand', the *order of candidates* (whether a given candidate appears as 'cand_01' or 'cand_05') is *determined by the number of votes the candidate obtained in the electoral district as a whole*. The candidate in the '01' position is the first place-getter; the candidate in the '02' position is the second place-getter; the candidate in the '03' position is the third place-getter and so on.

    - The total number of 'cand' variables runs until cand_17 because the largest number of candidates who ran in an electoral district in this period was 17. For each electoral district in a given election, the number of cand_ variables populated amounts to the number of candidates who ran in that district. A district seeing four candidates run will see data in the cand_01, cand_02, cand_03, and cand_04 variables, and not in the other cand_ variables.

- **cand_01_name … cand_17_name** (source: Reed and Smith 2015; variable **name_jp**)**:** this records the candidate's name in Japanese, for each candidate in positions 1 through the number of candidates who ran in the district.

- **cand_01_pty … cand_17_pty:** an English-language abbreviation of the candidate's political party, for each candidate in positions 1 through the number of candidates who ran in the district.

    - Originally, we used Reed and Smith 2015's **party_jp** variable to record the party names of the candidates running. We did not use the JED-M data. Then, because it was difficult to work with Japanese in the cells of the dataset, we converted the Japanese names to the abbreviations in the **name_en** variable in Reed and Smith 2015. In Appendix 1 of this codebook, we present the list of parties to which candidates in our data belonged and the English-language abbreviations used in the cand_01_pty … etc. variables.

- **cand_01_pid … cand_17_pid** (source: Reed and Smith 2015; variable: **pid**): a unique number identifying each candidate, for all candidates in positions 1 through the number of candidates who ran in the district. Note that pid variables were taken from Reed and

Smith 2015. Later versions of the same data published by Smith (in 2017) may use different pids. Our pid numbers come from the 2015 version of this data.

- **cand_01_terms … cand_17_ terms** (source: Reed and Smith 2015; variable: **totcwins**): this records the number of times the candidate has won an election, for all candidates in positions 1 through the number of candidates who ran in the district.

- **cand_01_votes_in_electoral_district** (source: JED-M)**:** the number of votes each candidate obtained in the electoral district as a whole, for all candidates in positions 1 through the number of candidates who ran in the district. For each candidate, we took the votes cast for her in all municipalities (including sections of municipalities) housed in the electoral district and summed them. Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district.

- **pr_bloc** (source: Reed and Smith 2015; variable: **prcode**): a numeric code for the PR bloc in which the municipality is located. Since Japan's electoral reform in 1994, municipalities have been in *both* electoral districts (single-member electoral districts) *and* PR (proportional representation) blocs. Voters cast two votes, one for a candidate in their electoral district and one for a party presenting a list in their PR bloc. Thus, PR-related variables are recorded only for Lower House elections held in and after 1996. PR blocs are drawn around prefectures (of which Japan has 47), and no municipalities span more than one prefecture, so all municipalities are perfectly contained within a single PR bloc.

- **votingpop_pr** (source: JED-M): (有権者数): number of people eligible to vote in the PR tier in the Lower House election in the municipality. Note that this tier was introduced with electoral reform in 1994, so this variable is only populated for elections after 1994.

- **mun_voted_PR** (source: JED-M): (有効投票者数): number of valid votes cast in the PR race in the municipality.

    - Note that this is *valid votes*, not total votes. Thus, we exclude spoiled ballots. This was calculated by taking the sum of the votes received by all parties presenting lists in the PR bloc in which the municipality is located. It is populated only for elections held after 1994.

- **mun_turnout_PR** (source: JED-M): **mun_voted_PR/ votingpop_pr**.

- **totmun_in_PR bloc** (source: JED-M): number of municipalities located in the PR bloc housing the municipalities. All municipalities are perfectly nested within a single PR bloc. It is populated only for elections held after 1994.

- **totseat_in_pr_bloc** (source: Reed and Smith 2015; variable **pr_m**): number of seats available in the PR bloc housing the municipality. It is populated only for elections held after 1994.

- **nparties_PR_bloc** (source: JED-M): number of parties that presented lists in the PR bloc housing the municipality and obtained at least one vote.

- **voted_for_[partyname]_in PR** (source: JED-M): number of votes obtained by each party presenting lists in the PR tier in the municipality (with name inserted where [partyname] is).

    - There are 26 different variables with PR votes for parties as there were 26 different parties that competed in these elections. Variables are only populated for elections in which a party presented a list. If a party presented a list in 1996, but no other election year, data will be populated for this party for 1996 and not in other election years.
    - Unlike the cand_ variables, the order in which parties appear has no relevance.

Note that the JED-M contains errors, some of which appear to be errors in the way the prefectural electoral commissions reported their voting data (and thus, not in the compilation process). We corrected these errors by hand as we made the data set.

Note also that municipalities in the Amami Islands were in a special electoral district called 奄美群島区 (**hor_electoral_district** == 4604) for the 1980, 1983, 1986 and 1990 elections, but in 1993, this district became part of Kagoshima 1st District. The raw data does not include voting data for these 14 municipalities for the 1980-1990 elections, and only includes them for the 1993 election, after they became part of Kagoshima 1st.

The second dataset used to compile the raw data was obtained from Nikkei NEEDs. This data is fiscal and demographic variables at the level of the municipality for the period 1980 until 2012 (fiscal years). Note that while the base data includes voting data from 1980 until 2014, the fiscal and demographic variables only run from 1980 until 2012 (later, I explain how I added the subsidy and demographic variables for 2013, 2014, and 2015 from data supplied by the Japanese government).

For every Japanese municipality in existence from 2000 onwards, the NEEDs data we used to compile the raw data records fiscal, demographic, and other variables for each municipality, spanning the period 1980-2012.

Critically, this inclusion rule has the following implications:

- Municipalities that ceased to exist prior to 2000 (a very small number) are not in this data. Thus, while we have elections data for these observations (because they were in the JED-M data set), we do not have fiscal and demographic variables for them.

- Municipalities created through a merger in the period after 2000 (in, say, 2003) have data going back to 1980. Thus, data prior to the year of a municipality's creation was imputed. Thus, for all municipalities that were created in the 2000s, we deleted data for the years prior to their existence. To help us, we used "boundary_changes_formatted.xls", which is a comprehensive list of all municipalities that ceased to exist since 2000 and all municipalities that were newly-created since 2000. Helpfully, this documents what happened to all the municipalities that ceased to exist (i.e. which municipality they were absorbed into).

In the NEEDs data we used to compile the raw data, data is recorded under both the (Japanese) name the municipality had in the post-2000 period and an official municipality code. Some of the files with variables received from Nikkei NEEDs contain only municipality names or only municipality codes, while others contain both.

This naming convention also has implications. If a municipality changed its name (or its designation as a town/village etc.) at some point prior to 2000, it will be recorded in the NEEDs data under the name it had in the post-2000 period. Thus, if a municipality was a city in the post-2000 period, then the NEEDs data records data under this name, which includes "city", for all prior years as well. This makes this observation difficult to match to voting data in prior years if it was a town or village (or had a different name) in those years. We used data kindly supplied by Kuniaki Nemoto to match municipalities that experienced a change to prior versions of themselves.

A description of the fiscal and demographic variables in the raw data, which come from Nikkei NEEDs, are as follows. Generally-speaking, data is presented annually for every year from 1980-2012, unless a municipality did not exist in a given year:

- **code** (source: NEEDs, supplemented with Kuniaki Nemoto's data): official municipality code (市区町村コード). Note that the final three digits are a code for the municipality within its prefecture, while the remaining leading digits (which range from 1 to 47) indicate which prefecture the municipality is in (there are 47 prefectures). Importantly,

this code is assigned to municipalities *without regard to how their borders have changed over time*.

- For example, old Hakodate City existed until December 1, 2004. On this date, new Hakodate City came into being by absorbing the old Hakodate City, as well as Toi Town, Esan Town, Todohokke Village, and Minamikayabe Town. While old Hakodate City and new Hakodate City have different borders, they have the same code (1202). The four towns and villages also have their own codes (1339, 1340, 1341, and 1342) until they disappeared in fiscal year 2004.

- Note that municipalities subsumed into other municipalities exist in the data as independent municipalities until the fiscal year prior to the creation of the new municipality. Intuitively, this makes sense because if a municipality ceases to exist during fiscal year 2004, it will not have completed fiscal year 2004, so will not have data for that year. Continuing with the above example, old Hakodate City, Toi Town, Esan Town, Todohokke Village, and Minamikayabe Town exist as independent municipalities in the data until 2003. New Hakodate City, created by merging these municipalities, exists in the data from fiscal year 2004 onwards. Because the creation of new Hakodate City occurred on December 1, 2004, in the middle of fiscal year 2004, the data for fiscal year 2004 was likely created by adding values for the independent municipalities prior to December 1, 2004 to values for the new entity from December 1 2004 until March 31 2005.

- Note that where possible, we used the **code** values obtained from the NEEDs data. When we could not match municipalities in the JED-M data to observations in the NEEDs data, we searched for a match to observations in data provided by Kuniaki Nemoto. If we found a match, we used the code from his data.

- Even when we did find a match to an observation in Nemoto's data, note that **code** is not populated for municipalities that experienced an election *followed by a merger within the same fiscal year*. Because these municipalities did not exist for the *full* fiscal year, we do not have fiscal and demographic variables for them. Not assigning them a **code** helped to avoid merging in imputed data from NEEDs. We kept the voting data attached to these municipalities in the raw data, but did not assign them a **code.**

- **muncode_num** (source: NEEDs, supplemented with Kuniaki Nemoto's data): **code** is useful if we want to merge in other government data. But if we want to study the same

municipality over time (in a panel regression), **code** is problematic because municipalities with the same code are not always the same entity. This is why we created **muncode_num**. This retains the original code, but affixes a decimal suffix to indicate whether a given municipality is a prior version of a similar municipality with the same name.

- Continuing with the above example, the old Hakodate City has a different **muncode_num** (1202.1) to the new Hakodate City (1202), whereas their **code** is identical. This indicates that the former, which exists in the data until 2003, was an earlier version of the latter, which exists in the data from 2004. Because Toi Town, Esan Town, Todohokke Village, and Minamikayabe Town, which exist in the data until 2003 (after which they were subsumed into new Hakodate City), are *not* earlier versions of a municipality with same name, their code and muncode_num is the same (1339, 1340, 1341, and 1342, respectively).

- Warning: programs like Excel have started to drop the decimal suffix when reading data in. Verifying that **muncode_num** contains decimal suffixes is important, particularly for any analyses that examine the post-2000 period.

- **mun_name_needs** (source: NEEDs): the municipality's name. For municipalities that were later subsumed into new municipalities, whether of the same or different name, values in **mun_name_needs** incorporate this information. This information is from the file "boundary_changes_formatted.xls". We entered this information to help us gauge which municipalities were prior versions of themselves and to conduct sanity checks as to which years municipalities existed. The information is incorporated into **mun_name_needs in** two ways:

  1. If an entity has the suffix "旧...date… まで", it is an older version of a municipality with the same name, which became that new municipality on the date noted. The newer municipality will have the same **mun_name_needs** as the old municipality but without the suffix. Continuing with the above example, old Hakodate City has a **mun_name_needs** (in Japanese) of "Hakodate City (kyuu 2004/12/01 made)" and the new Hakodate City has a **mun_name_needs** (in Japanese) of just "Hakodate City".

  2. If an entity has the suffix "(municipality name, date)" in its **mun_name_needs**, this means that it became part of a new municipality of that name on that date. Continuing with the above example, Toi Town, Esan Town, Todohokke Village, and Minamikayabe Town all have a **mun_name_needs** of their name, followed by the

suffix “(Hakodate City 2004/12/01)”, which indicates that they were subsumed into the new Hakodate City when it was created on December 1, 2004.

- **mun_stxlo** (source: NEEDs): amount of Local Allocation Tax (Ordinary) (LAT ordinary) received by the municipality in the fiscal year (地方交付税普通交付金). Units: JPY 1,000,000 (百万円). Note that municipalities that are fiscally strong or are wards of ordinance-designated cities do not receive LAT.

- **mun_stxls** (source: NEEDs): amount of Local Allocation Tax (Special) (LAT special) received by the municipality in the fiscal year (地方交付税特別交付金). Units: JPY 1,000,000 (百万円).

- **mun_stxl:** sum of **mun_stxlo and mun_stxls**

- **mun_ngaid** (source: NEEDs): amount of National Treasury Disbursements (NTD) received by the municipality in the fiscal year (国庫支出金). Units: JPY 1,000,000 (百万円).

- **mun_ipbf** (source: NEEDs)**:** Standardized fiscal revenue (基準財政収入額). This is a standardized measure of the amount of fiscal revenue a municipality generates from taxation. It is an indicator of a municipality's revenue-collecting capacity. Units: JPY 1,000,000 (百万円).

- **mun_dpbf** (source: NEEDs)**:** Standardized fiscal demand (基準財政需要額). This is a standardized measure of the amount of expenditure needed by a municipality to maintain public services. Units: JPY 1,000,000 (百万円).

- **mun_ceif** (source: NEEDs)**:** Index of fiscal strength (財政力指数). Higher values indicate greater fiscal soundness.

- **mun_population** (source: NEEDs)**:** Population in the municipality on March 31 of that year (住民基本台帳人口要覧). Note that in the 1980-2012 data, figures are limited to people with Japanese citizenship (later years appear to include those without citizenship).

- **mun_taxable_income** (source: NEEDs)**:** total taxable income in the municipality (課税対象所得額). Sum of per capita taxable income in the municipality. Units: JPY 1,000,000 (百万円).

- **mun_area_size** (source: NEEDs): geographic size of the municipality (面積). Units: kilometers squared. Note that the data source used by NEEDs records the area size of municipalities from 1998. Thus, for the years 1980-1997, we transplant the area size measure for the oldest year on record for a given **muncode_num**. This means that our pre-1998 measures are roughly accurate, but do not take into account changes in area size due to reclamation (whereas the post-1998 data does).

- **mun_population_density**: **mun_population/mun_area_size.** Units in population per kilometer squared.

- **year closest5** (source: NEEDs): this records the 5-year interval closest to the year of that observation. For example, the years 1983, 1984, 1985, 1986, and 1987 all receive a **year closest5** value of 1985. We do this because five of the NEEDs variables (mun_pop_14_and_under, mun_pop_65_and_over, mun_employment_primary, mun_employment_secondary, and mun_employment_tertiary) are recorded every 5 years. We use **year_closest5** to merge in this data.

- **mun_pop_14_and_under** (source: NEEDs): Adolescent population (14 years and under) in the municipality (１５歳未満人口). Data is collected at 5-year intervals (in 1980, 1985, 1990, etc.) on October 1st of those years. For the off-years, values are obtained from the survey year that is closest (year_closest5). For example, values for municipalities in 2002 are from 2000 and for municipalities in 1983 are from 1985.

- **mun_pop_65_and_over** (source: NEEDs): Elderly population (65 years old and older) (65歳以上人口). Population that is 65 years and over in the municipality. Data is collected at 5-year intervals (in 1980, 1985, 1990, etc.) on October 1st of those years. For the off-years, values are obtained from the survey year that is closest (year_closest5).

- **mun_employment_primary** (source: NEEDs): Population employed in primary sector (第一次産業). Population 15 years and over employed in the primary sector in the municipality (agriculture, fisheries, etc.) Data is collected at 5-year intervals (in 1980,

1985, 1990, etc.) on October 1st of those years. For the off-years, values are obtained from the survey year that is closest (year_closest5).

- **mun_employment_secondary** (source: NEEDs): Population employed in secondary sector (第二次産業). Population 15 years and over employed in the secondary sector in the municipality (construction, manufacturing, etc.) Data is collected at 5-year intervals (in 1980, 1985, 1990, etc.) on October 1st of those years. For the off-years, values are obtained from the survey year that is closest (year_closest5).

- **mun_employment_tertiary** (source: NEEDs): Population employed in tertiary sector (第三次産業). Population 15 years and over employed in the tertiary sector in the municipality (health, education, sales, etc.) Data is collected at 5-year intervals (in 1980, 1985, 1990, etc.) on October 1st of those years. For the off-years, values are obtained from the survey year that is closest (year_closest5).

The following variables in the raw data are incomplete: mun_DIDpopulation, mun_disaster_stricken, mun_is_seirei.

Note that the raw dataset used for Catalinac's book ("out_final_utf8.csv") is not exactly the same as the raw dataset used for the 2020 article as it is the product of correcting errors and incorporating new data.

---

4. MASTER DATA (Master_plus_Snow_Turn_Trans_Dis6.rds)[6]

To make the master data from the raw data, I proceeded in the following steps:

In Step 1, I read the raw data into a Stata script ("OrganizeJapanDataMay2018_lmc.do"). This is a revised version of the Stata script used to create the variables for Catalinac, Bueno de Mesquita, and Smith (2020). This script makes a host of voting-related variables and drops split municipalities. It saves the data as tempMay2018.dta; this was converted to Amy.RData and then MasterDataDec2019.csv (this is a .csv file encoded in SHIFT-JIS, so it should not be opened in Excel directly).

[6] The R script for Chapter 5 uses "Master_plus_Snow_Turn_Trans_Dis4.rds". This is the same file minus a single variable: (tmt_possible).

In Step 2, I took the data created in Step 1 (MasterDataDec2019.csv) and added fiscal and demographic variables for three additional years (2013, 2014, and 2015). To do this, I examined the sources of the data supplied by Nikkei NEEDs and collected the same variables from the Japanese government's E Statistics portal for the years 2013, 2014, and 2015. Before merging in this data, I verified that the variables for 2012 (supplied to us by NEEDs) were identical to the values for the variables in 2012, which I downloaded from the portal.

This repository includes the downloaded (and cleaned) files with the raw data contained in them, a file with further errors corrected ("corrected_variables_13-15.csv"), and the R script I used to merge these new variables in (Book_AddingData.R). I also merged in several variables used in Catalinac and Muraoka 2023. I called this version of the data "Master_plus_Snow.csv".

In Step 3, I took "Master_plus_Snow.csv" and made new variables used in my book, using the script Adding_variables_to_MasterPlusSnow2.R.

Below, I describe the variables in the master data. Remember that the following voting-related variables are populated for election years only. Note also that they were made *after dropping split municipalities.* This means that the master data does not contain the universe of electoral districts present in each of these elections because districts comprised *solely* of split municipalities have been dropped.

First, indicator variables were used to construct the electoral support variables, where "i" refers to "candidate i":

- **win_i**: did candidate i win or lose (note: this is also coded for non-election years because it uses **cand_01** and **totseat_in_electoral_district**)

- **senior_i:** was candidate i senior or not (senior is defined as having won five elections).

- **LDP_i**: was candidate i an LDP candidate or not

- **COM_i**: was candidate i either a JCP or JSP candidate or not

- **LDP_iplus**: was candidate i an LDP or LDPI candidate (an LDP-leaning conservative independent) or not

- **Vsharei**: what proportion of votes cast in the municipality went to candidate i (**cand_0'i'_votes/mun_voted**)

- **VshareVPi:** what proportion of the voting population in the municipality cast votes for candidate i (**cand_0i_votes/mun_voting_pop**)

- **PropSupporti:** proportion of candidate i's votes in the electoral district as a whole that came from this municipality (**cand_0`i'_votes/cand_0`i'_votes__in_electoral district**)

Next, here are the main variables used:

- **bestLDP_VshareVP:** for each municipality, this takes the VShareVPs it returned for each LDP winner who ran in the electoral district and records the maximum of these.

- **sumLDP_VshareVP:** for each municipality, this takes the VShareVPs it returned for each LDP winner who ran in the electoral district and records the sum of these.

- **rbestLDP_VshareVP:** a ranked version of bestLDP_VshareVP, in which municipalities are ranked on this variable alongside other municipalities in the same electoral district in the same year. Municipalities with the highest bestLDP_VshareVP in their electoral district receive 1, whereas municipalities with the lowest bestLDP_VshareVP receive 0.

    - Concretely, we first rank all municipalities in the same district-year according to bestLDP_VshareVP. Then, we take a municipality's absolute rank and subtract 1, and then divide that value by the total number of municipalities in the district-year minus 1.

    - Note that ranked variables were constructed after excluding split municipalities.

    - Note also that when a municipality is the only municipality in its electoral district, it cannot be ranked, so its **rbestLDP_VshareVP** is NA**.** When a municipality is in an electoral district without an LDP candidate running, all municipalities in that district receive **bestLDP_VshareVP**==0, which also means they cannot be ranked, so their **rbestLDP_VshareVP** values are also NA.

- **rsumLDP_VshareVP:** a ranked version of **sumLDP_VshareVP,** in which municipalities are ranked alongside other municipalities in the same electoral district. Municipalities with the highest sumLDP_VshareVP in their electoral district receive 1, whereas municipalities with the lowest sumLDP_VshareVP receive 0. See above for further discussion on ranked variables.

- **HighLDPVS**: this takes the LDP candidate who won the most votes in the electoral district housing the municipality and records that candidate's VshareVP in the municipality.

- **bestLDPp_VshareVP:** for each municipality, this takes the VShareVPs it returned for each LDP and LDPI winner who ran in the electoral district and records the maximum.

- **sumLDPp_VshareVP:** for each municipality, this takes the VShareVPs it returned for each LDP and LDPI winner in the electoral district as a whole and records the sum of these.

- **bestSeniorLDP_VshareVP:** for each municipality, this takes the VShareVPs it returned for each senior LDP winner in the district and records the maximum.

- **bestnonLDP_VshareVP:** for each municipality, this takes the VShareVPs it returned for all non-LDP winners in the district and records the maximum.

- **bestloseLDP_VshareVP**: for each municipality, this takes the VShareVPs it returned for all LDP losers in the district and records the maximum.

- **sumSeniorLDP_VshareVP:** for each municipality, this takes the VShareVPs it returned for each senior LDP winner in the district and records the sum of these.

- **sumnonLDP_VshareVP**: for each municipality, this takes the VShareVPs it returned for all non-LDP winners in the district and records the sum of these.

- **sumloseLDP_VshareVP:** for each municipality, this takes the VShareVPs it returned for all LDP losers in the district and records the sum of these.

- **sumDPJ_vsharevp: this** records the share of the voting population in the municipality who voted for a DPJ candidate. Populated for elections in which the DPJ ran.

- **votes_first: cand_01_votes/mun_voting_pop**

- **ngaid_pc: mun_ngaid/mun_population**

- **logngaid_pc**: log(**ngaid_pc**)

- **F1ngaid_pc**: the municipality m's **ngaid_pc** in the subsequent year (made with muncode_num)

- **F1logngaid_pc:** the municipality 's **logngaid_pc** in the subsequent year (made with muncode_num)

- **Lngaid_pc:** the municipality's **ngaid_pc** for the prior year (made with muncode_num)

- **Llogngaid_pc:** the municipality's **logngaid_pc** for the prior year (made with muncode_num)

- **income_pc: mun_taxable_income/mun_population**

- **logincome_pc:** log(**income_pc)**

- **mun_needy: mun_pop_14_and_under** + **mun_pop_65_and_over**

- **needy_pc: mun_needy/mun_population**

- **primary_pc: mun_employment_primary/mun_population**

- **lnpop:** log(**mun_population)**

- **logincome_pc**: log**(income_pc)**

- **mun_population_density**: **mun_population_density/1000.** Note that the Stata script rescales mun_population_density in the raw data to be thousands of people per kilometer squared. Thus, in the master data, mun_population_density==1.2 means 1.2 thousand people per km squared.

- **numLDPwin**: number of LDP candidates who won a seat in the electoral district housing the municipality.

- **numLDPcand**: number of LDP candidates who ran in the electoral district housing the municipality.

- **VotesLDP**: for each municipality, record the number of votes cast in the municipality for all LDP candidates who competed in the electoral district.

- **VotesLDP_cast: VotesLDP/mun_voted**

- **VotesWinLDP**: for each municipality, record the number of votes cast in the municipality for all LDP candidates who won a seat in the electoral district.

- **numLDPwinsen**: number of senior LDP winners in the district. Populated for election years only. Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district.

- **last**: this is populated for a single municipality in each electoral district. Subsetting data to last==1 can be used to construct a data set at the level of the electoral district.

- **VotePop**: sum(**mun_voting_pop**). This takes the voting population in all (non-split) municipalities in the electoral district and records the sum of these.

- **frac**: **mun_voting_pop/VotePop.** The proportion of the electoral district's voting population (constructed without split municipalities) who reside in the municipality.

- **PopIndex**: sum(**frac**^2). This takes the share of the electoral district's voting population constituted by each municipality in the district, squares these, and sums them. Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district.

    - Lower values: the voting population in an electoral district is more evenly spread out across the municipalities therein. Higher values: the population is more concentrated in a single municipality.
    - e.g. a district with relatively evenly-sized municipalities: (0.3^2) + (0.3^2) + (0.4^2) gets a PopIndex of 0.34. A district with unevenly-sized municipalities: (0.05^2) + (0.8^2) + (0.15^2) gets a PopIndex of 0.665.

- **HI:** (**PopIndex**-1/num)/(1-1/**num**). Populated for election years only. Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district.

- Electoral districts with fewer municipalities receive higher values on **PopIndex** than electoral districts with more municipalities, even when their voting populations are both evenly divided into constituent municipalities. **HI** takes a district's **PopIndex** and normalizes it across electoral districts, so that regardless of the number of municipalities in a district, districts in which votes are evenly divided across constituent municipalities get **HI**=0 and districts in which votes are concentrated in a single municipality get **HI**=1.

- **DISpop**: Population in the electoral district, calculated by taking the sum of **mun_population** for all (non-split) municipalities in the district. Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district.

- **fracpop: mun_population/DISpop.** The proportion of the electoral district's population (calculated without split municipalities) who live in the municipality.

- **DISmal**: **DISpop**/(100000* **totseat_in_electoral_district**). The number of people, in 100,000s, per seat in the municipality's electoral district. Populated for election years only. Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district.

- **DISceif: sum(mun_ceif*fracpop**). A district-level fiscal strength value, calculated by taking each (non-split) municipality's value for fiscal strength (**mun_ceif**) and multiplying it by its share of population in the electoral district. Populated for election years only. Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district.

- **DISneedy: sum(needy_pc*fracpop).** Share of a district's population who is needy, calculated by taking the share of needy residents in each (non-split) municipality (**needy_pc**), and multiplying this by its share of population in the electoral district. Populated for election years only. Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district.

- **DISprimary: sum(primary_pc*fracpop).** Share of a district's population employed in agriculture, calculated by taking the share of residents employed in agriculture in each (non-split) municipality (**primary_pc**), and multiplying that by its share of population in the electoral district. Populated for election years only. Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district.

- **DISarea:** sum(**mun_area_size**). A district-level area size variable, calculated by summing up the area size (in kilometers squared) of all (non-split) municipalities in a given district-year. Populated for election years only. Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district.

- **DISdensity**: **DISpop/(1000*DISarea).** The number of people, in 1,000s, per kilometer squared in the electoral district housing the municipality.

- **logDISpop**: log(**DISpop**)

- **DISincomepc:** sum**(income_pc*fracpop).** Per capita income in the district, calculated by taking the per capita income in each (non-split) municipality (**income_pc**), and multiplying that by its share of population in the district. Populated for election years only. Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district.

- **logDISincomepc:** log(**DISincomepc**)

- **DISngaid**: sum(**ngaid_pc*mun_population**). amount of NTD received by all (non-split) municipalities in the electoral district in that fiscal year. Calculated by taking the per capita NTD received by each (non-split) municipality in the electoral district, multiplying that value by the municipality's population, and then taking the sum of these values for all municipalities in the district-year. Populated for election years only. Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district.

- **DISngaidpc**: **DISngaid/DISpop.** A per capita measure of the amount of NTD received by the electoral district housing the municipality. Calculated by taking the NTD received by the electoral district and dividing it by the district's population. Populated for election years only. Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district.

- **FDISngaid**: sum(F1ngaid_pc*mun_population). Total NTD received by the municipality's electoral district in the subsequent year. Calculated by taking the per capita NTD received by each (non-split) municipality in the electoral district in the subsequent fiscal year, multiplying it by the population of the municipality, and then taking the sum of

these values. Populated for election years only. Because this is an electoral district-specific variable, it is identical for all municipalities within the district.

- **FDISngaidpc: FDISngaid /DisPop**. The per capita amount of NTD received by the municipality's electoral district in the subsequent year. Calculated by taking the total amount of NTD received in the following year, and dividing that by the district's population. Populated for election years only. Because this is an electoral district-specific variable, it is identical for all municipalities within the district.

- **logFDISngaidpc**: log (**FDISngaidpc**).

- **num**: number of (non-split) municipalities in each electoral district. Populated for election years only. Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district.

- **logNum**: log(**num**)

- **LDPseats**: the share of available seats in the district-year won by LDP candidates. Populated for election years only. Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district.

- **DisLDPvotes:** sum **(VotesLDP)** This takes the number of votes cast for LDP candidates in each (non-split) municipality in the electoral district and sums them. Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district).

- **DisLDPVS: DisLDPvotes/voting_pop_in_electoral_district**. This takes the number of votes cast for LDP candidates in the electoral district, divided by voting population in the district. Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district.

    - Note that **DisLDPvotes** is created after dropping split municipalities, whereas **voting_pop_in_electoral_district** is from Reed and Smith 2015's **ku_electorate**, which includes voters in split municipalities. For this reason, we constructed **DisLDPV2** (see below).

- **DisWinLDPvotes:** sum **(VotesWinLDP**). This takes the number of votes cast for winning LDP candidates in each (non-split) municipality in the electoral district and sums these.

Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district.

- **DisWinLDPVS: DisWinLDPvotes/voting_pop_in_electoral_district.** The number of votes cast for winning LDP candidates in the district, divided by the voting population in the district. Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district.

    - Note that **DisWinLDPvotes** is created after dropping split municipalities, whereas **voting_pop_in_electoral_district** is from Reed and Smith 2015's **ku_electorate**, which includes voters in split municipalities. For this reason, we constructed **DisWinLDPV2.**

- **DisLDPVS2**: **DisLDPvotes/VotePop.** The number of votes cast for LDP candidates in the district, divided by the voting population of the non-split municipalities in the district. Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district.

- **DisWinLDPVS2: DisWinLDPvotes/VotePop.** The number of votes cast for winning LDP candidates in the electoral district, divided by the voting population of the non-split municipalities in the district. Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district)

- **district_year:** concatenation of **year** and **hor_electoral_district**. Used for district_year fixed effects.

- **Lmun_turnout: mun_turnout** in the Lower House election immediately prior to this election.

- **votes_single_LDPw**: the maximum number of votes cast in the municipality for a single winning LDP candidate.

- **vs_single_LDPw**: **votes_single_LDPw/mun_voted**

- **votes_all_LDPw:** the sum of all votes cast in the municipality for winning LDP candidates.

- **vs_all_LDPw: votes_all_LDPw/mun_voted**

- **votes_all_LDPc:** the sum of all votes cast in the municipality for LDP candidates.

- **vs_all_LDPc: votes_all_LDPc/ mun_voted**

- **rvs_single_LDPw**: Takes the **vs_single_LDPw** values for all municipalities in the same electoral district and ranks them so that the municipality with the lowest value receives 0 and the municipality with the highest value receives 1.

    - Concretely, we first rank all municipalities in the same district-year according to **vs_single_LDPw**. Then, we take a municipality's absolute rank and subtract 1, and then divide that by the total number of municipalities in the district-year minus 1. Note that when a municipality is the only municipality in its district, it cannot be ranked, so receives NA. When a municipality is located in an electoral district without an LDP winner, it also receives NA.

- **votes_single_nonLDPIw:** the maximum number of votes cast in the municipality for a single winning candidate who was not LDP *nor* an LDP-aligned independent candidate.

- **vs_single_nonLDPIw: votes_single_nonLDPIw/mun_voted**

- **votes_single_LDPl:** the maximum number of votes cast in the municipality for an LDP candidate who lost the election.

- **vs_single_LDPl: votes_single_LDPl//mun_voted**

- **votes_single_DPJw:** the maximum number of votes cast in the municipality for a DPJ winner.

- **vs_single_DPJw: votes_single_DPJw/mun_voted**

- **marginvs**: The number of votes received by the candidate in the Mth position, minus the number of votes received by the candidate in the M+1 position, divided by the number of valid votes cast in the electoral district (**voted_in_electoral_district**). *M* is number of seats available in the district (district magnitude). It ranged from two to six in elections

held between 1980 and 1993, and is 1 thereafter. Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district.

- **kucoder** (source: Reed and Smith 2015). This is a code for a municipality's HOR electoral district that is similar to **hor_electoral_district** but has a decimal suffix attached at the end, indicating that an electoral district is a prior version of the same electoral district.

- **district_reform**: a code for the **hor_electoral_district** a municipality is located in, preceded by "0_" for elections prior to electoral reform and "1_" for elections after reform. It is an indicator variable for electoral district that is specific to pre- and post-1994. For example, 0_101 is the code Hokkaido First District receives in elections prior to 1994, while 1_101 is the code it receives after 1994.

- **district_reform2:** the same as **district_reform** except that instead of **hor_electoral_district, kucoder** is used. **Kucoder** is from Reed and Smith 2015. Like **hor_electoral_district, kucoder** does not take into account changes in electoral district boundaries before and after electoral reform, but it does take into account minor changes in electoral district boundaries within each period.

- **KomSSD**: 1 if the electoral district sees a Komeito candidate run and no LDP candidate, 0 otherwise. Coded only for election years in and after 1996. Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district.

- **resurrectSSD:** 1 if the electoral district sees an LDP candidate lose but enter the Lower House via the PR tier, 0 otherwise. Coded only for election years in and after 1996. Because this is an electoral district-specific variable, it is identical for all municipalities within the electoral district.

---

5. GUIDE TO REPLICATION SCRIPTS FOR EMPIRICAL CHAPTERS

The folder contains the separate R scripts used to implement all the analyses used in Chapters 5, 6, 7, and 8 of the book. Below, I present a guide to each R script, noting the files it reads in and the files it outputs.

Catalinac_2025_ch5replication.R

Files it uses:

- Raw data (consulted, not read in)
- Master_plus_Snow_Turn_Trans_Dis4.rds
- Reed-Smith-JHRED.csv (Reed and Smith version 2015)

Files it makes:

- no_tmts_districts2.csv
- Master_plus_Snow_Turn_Trans_Dis6.rds

Catalinac_2025_ch6replication.R[7]

Files it uses:

- Master_plus_Snow_Turn_Trans_Dis6.rds
- RSC_JHRED_12-23-2016.csv (Reed and Smith (version 2015) matched to file names from the 7,497 candidate manifestos used in Catalinac 2015)
- Catalinac_merged_with_topics.csv (topics data used in Catalinac 2015)
- no_tmts_districts2.csv
- 2012_voter_survey_data_edits.csv

Catalinac_2025_ch7replication.R

Files it uses:

- Master_plus_Snow_Turn_Trans_Dis6.rds
- pref_capitals.csv

Catalinac_2025_ch8replication.R

Files it uses:

- Master_plus_Snow_Turn_Trans_Dis6.rds

[7] Must choose SHIFT-JIS encoding to view this file in RStudio.

6. APPENDIX 1: PARTIES IN THE CANDIDATE VARIABLES

1	自民党	LDP
2	無所属 (社)	JSPI
3	社会党	JSP
4	民主党	DPJ
5	新進党	NFP
6	無所属（自民）	LDPI
7	民主社会党	DSP
8	さきがけ	Sakigake
9	公明党	Komeito
10	新生党	Renewal
11	無所属の会	Mushozokunokai
12	小沢の自由党	Ozawa Liberals
13	未来の党	TPJ
14	無所属（民主）	DPJI
15	日本新党	JNP

16	みんなの党	Your Party
17	新自由クラブ	NLC
18	無所属	Independent
19	保守党	Conservative
20	共産党	JCP
21	社会民主連合	SDL
22	国民新党	PNP
23	日本維新の会	JRP
24	社会民主党	SDP
25	民主改革連合	DRL
26	新党日本	NPJ
27	無所属（新進）	NFPI
28	無所属（日新）	JNPI
29	無所属（社）	JSPI
30	自由連合	Liberal Alliance
31	新党大地	Daichi
32	無所属（社民）	SDPInd

33	幸福実現党	Happiness Realization
34	改革クラブ	Minor Party
35	進歩党	Progressive Party
36	無所属（小自）	Ozawa Liberals Ind.
37	無所属 (公)	KomeitoInd
38	無所属（公）	KomeitoInd
39	無所属（さ）	SakigakeInd
40	無所属（民社）	DSPI
41	新社会党	New Socialist Party
42	無所属（共）	JCPI
43	無所属 (共)	JCPI
44	無所属 (民社)	DSPI
45	無所属（み）	Your Party Ind.
46	諸派	Minor Party
47	世界経済共同体党	Minor Party
48	国民党	Minor Party
49	無所属（ク）	NLCI

50	年金党	Minor Party
51	雑民党	Minor Party
52	大日本愛国党	Minor Party
53	政事公団太平会	Minor Party
54	地球維新党	Minor Party
55	日本労働党	Minor Party
56	日本世直し党	Minor Party
57	日本皇民党	Minor Party
58	憂国維新同志会	Minor Party
59	真理党	Minor Party
60	革自連	Minor Party
61	無所属（社民連）	SDLI
62	無所属 (ク)	NLCI
63	日本普遍政党	Minor Party
64	日本みどりの党	Minor Party
65	統一社会民主党	Minor Party
66	近畿政治ほほえみの会	Minor Party

67	尊皇義塾仁義社	Minor Party
68	日本自由民主党	Minor Party
69	日本人民党	Minor Party
70	クリーン新党	Minor Party
71	反ソ決死隊	Minor Party
72	緑の党	Minor Party
73	サラリーマン新党	Minor Party
74	世界新党	Minor Party
75	アジア建国党	Minor Party
76	日本道徳修身党	Minor Party
77	進歩自由連合	Minor Party
78	菊誠党	Minor Party
79	無所属 (社民連)	SDLI
80	緑社会党	Minor Party
81	立憲養正会	Minor Party
82	郷土美化・影山次郎後援会	Minor Party
83	大日本旭日社	Minor Party

84	社会主義労働者党	Minor Party

7. APPENDIX 2: PARTIES IN THE PARTY VARIABLES

民主 Democratic Party of Japan DPJ

自民 Liberal Democratic Party LDP

新社 New Socialist Party NSP

自連 Liberal Alliance LIBA

社民 Social Democratic Party SDP

共産 Japan Communist Party JCP

新進 New Frontier Party NFP

さきがけ New Harbinger Party (can be called "Sakigake", which is how Japanese is read) SAKI

民改 Democratic Reform League (民主改革連合) DRL

公明 Clean Government Party (usually called “Komeito”) KOM

自由 Liberal Party (Ozawa Liberals) LIBS

保守 Conservatives CONS

無会 Independents League (無所属の会) (this is read as Mushozoku no Kai) MUSH

社会 Socialist Party (社会党) SOC

日本 New Party Japan (新党日本) NPJ

国民 Peoples New Party PNP

大地 New Party Daichi (新党大地) NPD

幸福　Happiness Realization Party HRP

みな　Your Party YP

改ク　Reform Club RC

本質　New Party Essence (新党本質) NPE

維新　Japan Restoration Party JRP

未来　Tomorrow Party of Japan TPJ

改革　New Party Reform (新党改革) NPR

次世代 Party for Future Generations PFG

生活　People's Lives First PLF

支持なし Support No Party (literally, 支持政党なし is the name of a party that ran) NOPTY

---

8. WORKS CITED

Catalinac, Amy. 2016. "From pork to policy: The rise of programmatic campaigning in Japanese elections", *Journal of Politics*, 78.1: 1-18.

Catalinac, Amy, and Taishi Muraoka, 2023, "How incumbent politicians respond to the enactment of a programmatic policy: evidence from snow subsidies", *Japanese Journal of Political Science*, 24.1.

Catalinac, Bueno de Mesquita, and Smith, 2020, "A tournament theory of pork barrel politics: The case of Japan", *Comparative Political Studies* 53.10-11; 1619-1655.

Reed, Steven R., and Daniel M. Smith. The Reed-Smith Japanese House of Representatives Elections Data Set. Version: September 9, 2015.