

Two linked datasets of all Australian federal politicians and elections (1901-2020) and associated R package*

Rohan Alexander[†]

January 31, 2021

Abstract

TBD

1 Introduction

We tidy and link two datasets about Australian federal-level politics from 1901 to 2020: biographical information and election results. Tidying and linking these datasets allows researchers to answer a variety of questions that each individual dataset cannot. Our datasets were created in a reproducible way that can be updated and corrected as necessary and form a resource of interest to a variety of academic disciplines.

Our biographical dataset - `AustralianPoliticians` - consists of a collection of datasets related to Australian federal politicians from 1901 through to 2020. The datasets are:

- `'all.csv'`: The main dataset.
- `'by_division_mps.csv'`: Adds information about the division ('seat') of the politician.
- `'by_party.csv'`: Adds information about the party of the politician.
- `'by_state_senators.csv'`: Adds information about the state that a senator was representing.
- `'cabinet.csv'`: A dataset of who was in the cabinet.
- `'ministry.csv'`^{***}: A dataset of who was in the ministry

*I thank Hidayat Ismail for excellent research assistance. I thank Jill Sheppard, Monica Alexander, Rory Haupt, Ben Readshaw, Kelly Lyons, Sharla Gelfand, Simon Munzert, and Edward Howlett, for their general assistance, helpful comments, and suggestions. Thank you to Patrick Leslie who generously donated data. Our code and datasets are available at: <https://github.com/RohanAlexander/AusPol>. Comments on the 31 January 2021 version of this paper are welcome at: rohan.alexander@utoronto.ca. For the latest version please click [here](#).

[†]University of Toronto, rohan.alexander@utoronto.ca

- ‘uniqueID_to_apID.csv’: A correspondence between the uniqueIDs used in these datasets and the IDs used by the Australian Parliament House.

Our elections dataset - AustralianElections - consists of a collection of datasets related to all Federal Australian elections and by-elections from 1901 through to 2020. The datasets are:

- TBD.

Figure 1 identifies the relationship between these datasets.

```
knitr::include_graphics(here::here("inputs/schema.png"))
```

Understanding who is, and has been, in parliament is a critical initial step for almost any type of quantitative analysis. Nonetheless, there is no publicly available comprehensive dataset of those who have been members of Australia’s federal, state, and colonial parliaments. In this paper we introduce a new dataset - AustralianPoliticians - which contains daily data on the members of parliament for the Australian State Parliament, the six state parliaments, the two territory parliaments, and the preceeding colonial parliaments. We discuss the creation of this dataset, and then show its validity through various existing datasets and internal consistency tests. Finally, we discuss some extensions to this dataset. This dataset will be of use to those interested in quantitative analysis of Australian politics.

A critical first step for modern quantitative political analysis is a publicly available, comprehensive, dataset. In this paper we introduce one such dataset - AustralianPoliticians - which documents everyone who has been a member of any Australian federal, state, and/or colonial parliament. By bringing together various sources in this way, researchers can use this dataset to answer critical questions related to representation. Researchers can also build extensions to this dataset that focus on their particular interest.

AustralianPoliticians is a record of everyone who has been a member of the Australian State Parliament (starting from 1901), the six state parliaments as well as their colonial predecessors, and the two territories. There are X unique people in the dataset, and the core dataset - all - provides basic biographical information about them. Related datasets are linked to this core dataset by a uniqueID, and deal with parliament and house, party, division, ministries, and finally translate the uniqueID into various other IDs used by other sources.

The datasets are up-to-date as of 1 July 2020, and will be maintained and updated on a quarterly basis, by the author for the next five years. The latest version of the datasets are publicly available via GitHub, with an annual version lodged with the Australian Data Archive.

These datasets represent a critical contribution to Australian political science. It is expected that they will be widely built on and expanded and we look forward to working with other researchers in this process.

The datasets are up-to-date as of 31 January 2021.



all.rda	by_division_mps.rda	by_
uniqueID 	uniqueID 	uni
surname	division	par
allOtherNames	stateOfDivision	par
firstName	enteredAtByElection	pa
commonName	mpFrom	p
displayName	mpTo	partyCh
earlierOrLaterNames	mpEndReason	partyS
title	mpChangedSeat	partyS
gender	mpComments	partyV
birthDate		
birthYear		
birthPlace		
deathDate		
member		
senator		
wasPrimeMinister		
wikidataID		
wikipedia		
adb		

Figure 1: Relationship between different datasets.

2 Dataset details

2.1 all.csv

This is the main dataset and contains one row per politician, with columns: uniqueID, surname, allOtherNames, firstName, commonName, displayName, earlierOrLaterNames, title, gender, birthDate, birthYear, birthPlace, deathDate, member, senator, wikidataID, wikipedia, adb, and comments.

```
head(AustralianPoliticians::all)
```

##	uniqueID	surname	allOtherNames	firstName	commonName	displayName	earlierOrLaterNames	title	gender	birthDate	birthYear	birthPlace	deathDate	member	senator	wasPrimeMinister	wikidataID	wikipedia	adb	comments
## 1	Abbott1859	Abbott	Richard Hartley Smith	Richard	<NA>															
## 2	Abbott1869	Abbott	Percy Phipps	Percy	<NA>															
## 3	Abbott1877	Abbott	Macartney	Macartney	Mac															
## 4	Abbott1886	Abbott	Charles Lydiard Aubrey	Charles	Aubrey															
## 5	Abbott1891	Abbott	Joseph Palmer	Joseph	<NA>															
## 6	Abbott1957	Abbott	Anthony John	Anthony	Tony															
## 1	Abbott, Richard		<NA>	<NA>	male	<NA>														
## 2	Abbott, Percy		<NA>	<NA>	male	1869-05-14														
## 3	Abbott, Mac		<NA>	<NA>	male	1877-07-03														
## 4	Abbott, Aubrey		<NA>	<NA>	male	1886-01-04														
## 5	Abbott, Joseph		<NA>	<NA>	male	1891-10-18														
## 6	Abbott, Tony		<NA>	<NA>	male	1957-11-04														
## 1	Bendigo	1940-02-28	0	1		NA														
## 2	Hobart	1940-09-09	1	1		NA														
## 3	Murrurundi	1960-12-30	0	1		NA														
## 4	St Leonards	1975-04-30	1	0		NA														
## 5	North Sydney	1965-05-07	1	0		NA														
## 6	London	<NA>	1	0		1														
## 1																				
## 2																				
## 3																				
## 4																				
## 5																				
## 6																				
## 1																				
## 2																				
## 3																				
## 4																				
## 5																				
## 6																				

uniqueID is usually the surname of the politician and the year that they were born, e.g. Abbott1859. In certain cases this is not enough to uniquely identify them and then we add the first name, e.g. AndersonCharles1897 and AndersonGordon1897. In cases where there is punctuation in the surname, e.g. Ashley-Brown or O'Brien, this has been removed but capitalisation has been retained, so those would become AshleyBrown or OBrien, respectively.

commonName is used to highlight the name that the politician tended to be known as e.g. Ted instead of Edward This is used in displayName which is a politicians surname and their common name (if they had one) or first name e.g. Abbott, Richard. In cases where this would not be unique, e.g. Francis Baker, an additional name has been added.

earlierOrLaterNames is mostly used to keep track of women changing their names at marriage. Similarly, title is mostly used to keep track of 'Dr', but both have been used inconsistently and should be only used sparingly.

Some politicians don't have a complete birth date, and instead only have a year of birth. In these cases their entry for birthDate will be empty, but they will have a birthYear. All death dates are complete, but in the case of one politician – John William Croft – this has been inputted, as the circumstances and timing (even year) of his death are unknown. birthPlace is mostly taken from WikiData, with a few updates. There are some issues that need to be addressed here.

member and senator are binary indicator variables used to signify whether the politician was in the lower or upper house. Most politicians are only in one or the other, but some were in both. One politician in the dataset was neither a senator nor an MP - Heather Elaine Hill. She remains in the dataset because she was elected to the senate, and the need for this dataset to exactly match the AustralianElections one), however her eligibility was challenged and her election was invalidated, so she was never a senator.

adb is a link to the Australian Dictionary of Biography.

2.2 by_division_mps.csv

This dataset adds information about the division ('seat') of the politician. One row per division-politician, with columns: uniqueID; mpsDivision; mpsState; mpsEnteredAtByElection; mpsFrom; mpsTo; mpsEndReason; mpsChangedSeat; and mpsComments.

```
head(AustralianPoliticians::by_division_mps)
```

##	uniqueID	division	stateOfDivision	enteredAtByElection	mpFrom
## 1	Abbott1869	New England	NSW	No	1913-05-31
## 2	Abbott1886	Gwydir	NSW	No	1925-11-14
## 3	Abbott1886	Gwydir	NSW	No	1931-12-19
## 4	Abbott1891	New England	NSW	No	1940-09-21
## 5	Abbott1957	Warringah	NSW	Yes	1994-03-26
## 6	Abel1939	Evans	NSW	No	1975-12-13
##		mpTo	mpEndReason	mpChangedSeat	mpComments

## 1	1919-11-03	Retired	NA	<NA>
## 2	1929-10-12	Defeated	NA	<NA>
## 3	1937-03-28	Resigned	NA	<NA>
## 4	1949-10-31	Retired	NA	<NA>
## 5	2019-05-18	Defeated	NA	<NA>
## 6	1977-11-10	Retired	NA	<NA>

Certain divisions change name. Sometimes this is minor, for instance Kingsford-Smith to Kingsford Smith, and sometimes it is total. In all cases this is being treated as change in division – the politician is treated as finishing with one division and moving to another – but `changedSeat` can be used to identify these cases and adjust for them if necessary.

`byElection` is a binary indicator variable as to whether the politician was entering the seat following a by-election.

`changedSeat` is a binary indicator variable as to whether the politician left a division because they were changing the division, as opposed to losing an election or retiring.

2.3 by_party.csv

This dataset adds information about the party of the politician. One row per party-politician, with columns: `uniqueID`; `partyAbbrev`; `partyName`; `partyFrom`; `partyTo`; `partyChangedName`; `partySimplifiedName`; `partySpecificDateInputted`; and `partyComments`.

```
head(AustralianPoliticians::by_party)
```

##	uniqueID	partyAbbrev	partyName	partyFrom	partyTo
## 1	Abbott1859	CP	Australian Country Party	<NA>	<NA>
## 2	Abbott1869	CP	Australian Country Party	1919-11-03	<NA>
## 3	Abbott1869	LIB	Liberal Party of Australia	<NA>	1917-02-17
## 4	Abbott1869	NAT	Nationalist Party	1917-02-17	1919-11-03
## 5	Abbott1877	CP	Australian Country Party	<NA>	<NA>
## 6	Abbott1886	CP	Australian Country Party	<NA>	<NA>
##	partyChangedName	partySimplifiedName	partySpecificDateInputted	partyComments	
## 1	NA	Nationals	NA	<NA>	
## 2	NA	Nationals	NA	<NA>	
## 3	NA	Liberals	NA	<NA>	
## 4	NA	Nationalist Party	NA	<NA>	
## 5	NA	Nationals	NA	<NA>	
## 6	NA	Nationals	NA	<NA>	

Party can be a little confusing in cases where a politician changed party. In general, in this dataset, the to/from dates are set-up so that when a politician is in parliament they will have the correct party. However the dataset should not be used to say anything about when they are out of parliament. For instance, some politicians lost their seat, changed party, and then regained a seat in parliament. The dataset does not know when they

changed party while they were out of parliament, and it assumes that they changed party either at the same time that they lost their seat or at the same time as they re-gained a seat. Similarly, there are plenty of cases where a politician has ceased being a member after they leave parliament, for instance, Malcolm Fraser left the Liberals. Again, that is not reflected in the dataset.

Certain parties, such as the Nationals, changed their name at various points in time. This is included as a party change for people at that time in `partyAbbreviationParlHandbook` and `partyNameParlHandbook`. However, `partySimplified` abstracts away from that.

Party name changes:

- The Country Party changed to the National Country Party on 3 May 1975 according to <http://nla.gov.au/nla.news-article110636121>. It then changed from the National Country Party to the National Party of Australia on 17 October 1982 according to <http://nla.gov.au/nla.news-article116476081>. And finally, it changed from the National Party of Australia to The Nationals on 11 October 2003 according to the party website.
- The Nick Xenophon Team changed to Centre Alliance on 10 April 2018, according to ABC news reports.

2.4 by_state_senators.csv

This dataset adds information about the state that a senator was representing. The variables are: `uniqueID`; `senatorsState`; `senatorsFrom`; `senatorsTo`; `senatorsEndReason`; `senatorsSec15Sel`; and `senatorsComments`.

```
head(AustralianPoliticians::by_state_senators)
```

##	uniqueID	senatorsState	senatorFrom	senatorTo	senatorEndReason	sec15Sel
## 1	Abbott1859	VIC	1928-12-18	1929-06-30	Retired	1
## 2	Abbott1869	NSW	1925-11-14	1929-06-30	Defeated	0
## 3	Abbott1877	NSW	1935-07-01	1941-06-30	Defeated	0
## 4	Abetz1958	TAS	1994-02-22	<NA>	<NA>	1
## 5	Adams1943	WA	2005-07-01	2012-03-31	Died	0
## 6	Adamson1857	QLD	1920-07-01	1922-05-02	Died	0
##	senatorComments					
## 1	<NA>					
## 2	<NA>					
## 3	<NA>					
## 4	<NA>					
## 5	<NA>					
## 6	<NA>					

This dataset is fairly similar to `by_division_mps`, expect that it also has `senatorsSec15Sel`. This is a binary indicator variable and indicates whether the senator has been appointed rather than elected.

2.5 uniqueID_to_aphID.csv

This dataset adds a correspondence between the unique identifiers used in these datasets and the identifier used by the Australian Parliament House on its website. the main issue with the APH identifier is that it is not clear who it is referring to without looking it up. Additionally, in certain cases it changes from time to time, and it is easy to accidentally change the format by opening it in Excel.

```
head(AustralianPoliticians::uniqueID_to_aphID)
```

```
##      uniqueID aphID idsComments
## 1 Abbott1957   EZ5      <NA>
## 2 Abbott1886   JLJ      <NA>
## 3 Abbott1891   JLL      <NA>
## 4 Abbott1877   JNT      <NA>
## 5 Abbott1869   JLM      <NA>
## 6 Abbott1859   JYU      <NA>
```

2.6 appointments.csv

This dataset adds the beginnings of a dataset about ministerial appointments. Don't use it yet.

```
# head(AustralianPoliticians::positions)
```

2.7 Sources

In the first instance, the Parliamentary Handbook was the main source of information. This was augmented with information from Wikipedia, the Australian Dictionary of Biography, and the Senate Biographies wherever possible. Limited information was obtained from other sources, such as state parliaments and newspapers (via Trove), and these have generally been specified in the comments. birthPlace is mostly from WikiData.

The uniqueID_to_aphID dataset was primarily drawn from a dataset put together by Patrick Leslie, and it was checked against a modern dataset from Open Australia, and Tim Sherratt's Historic Hansard records for the Reps and Senate.

3 Summary statistics

3.1 Australia

There have been X people elected to the Australian Federal Parliament from when it was established in 1901 through to 1 July 2020. This includes X people who did not take their seat either because of death or a nullified election.

3.2 Validity

3.2.1 Internal consistency

There are a variety of internal consistency measures that the dataset passes. These include: not having more members on any day than there are seats in parliament; X; and Y.

3.2.2 Existing datasets

4 R Package

Although it is possible to directly read in the CSV datasets, for those that use R (R Core Team, 2020) a wrapper R package has been developed. This package can be installed from GitHub and then the datasets read in.

```
install.packages("devtools")
devtools::install_github("RohanAlexander/AustralianPoliticians")
library(tidyverse)

all <- AustralianPoliticians::all %>% as_tibble()
by_division_mps <- AustralianPoliticians::by_division_mps %>% as_tibble()
by_party <- AustralianPoliticians::by_party %>% as_tibble()
by_state_senators <- AustralianPoliticians::by_state_senators %>% as_tibble()
uniqueID_to_aphID <- AustralianPoliticians::uniqueID_to_aphID %>% as_tibble()
```

The datasets could then be merged using `left_join`:

```
all_individuals_with_their_division <- all %>%
  left_join(by_division_mps, by = c("uniqueID"))
```

Monica Alexander has written a brief blog post where she uses the package to look at life expectancy of Australian politicians:

https://www.monicaalexander.com/posts/2019-08-09-australian_politicians/

5 Vignettes

In this section we walk through several vignettes to illustrate how to use the dataset. These vignettes include: the average age of parliaments and gender proportions over time.

5.1 Average age of parliaments

5.2 Gender proportions

6 Discussion

6.1 Weaknesses and next steps

The dataset represents a foundation that can be built on. In particular,

6.2 Roadmap

- Add information about education.
- Add information about relationships, for instance father-son, etc.
- Add voting based on JFG's dataset.

References

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.