Analysing Australian Election Results

Carlos Yanez Santibanez

2023-06-08

Contents

Chapter 1

Acknowledgements

Chapter 2

Introduction

Chapter 3

Strucure

_ Problem statement - Obtaining the data - EDA - Modelling - Conclusions

APPENDICES - HOW we waht the data obtained - Detailed EDa - TEchnical note on buulding packages - vignettes

Chapter 4

Data

4.1 Data Sources

The first in the process was to source demographic and electoral data, which has been provided from two sources:

- **The Australian Electoral Commission** (AEC) [Commission, 2023a] . The AEC contains detailed online records for every federal election held in the 21st century, through their Tally Room website [Commission, 2023b].
- The **Australian Bureau of Statistics** (ABS) [of Statistics, 2023a]. The ABS provides a wide number of national statistics and is responsible to conduct a national census of population and housing every 5 years. Comprehensive census data is provided in multiple formats, including csv files through Census Data Packs [of Statistics, 2023b], which are available for censuses from 2006 onwards.

Both organisations are the authoritative source for electoral and statistical data in Australia, and the data is provided openly. Although there are no quality issues, the way that data is provided presents other challenges, namely:

- In both cases, data are provided in large volumes and exhaustive granularity. If not done effectively, data extraction and aggregation can be time-consuming and resource intensive.
- Census data points are provided using the ABS own geographical standard - and only a small selection of census data is provided at the electoral division level. Conversion between ABS geographical structures and electoral divisions is not straightforward as there is no 1:1 correspondence. Both geographical systems change from election to election and census to census.
- Despite the best efforts of both organisations in keeping consistency, names of electorates, parties, and census attributes change over time - to compare

similar statistics manual mapping is necessary.

To address these issues and ensure repeatability, three R packages have been written to undertake this task:

- **{auspol}** [**Yáñez Santibáñez, 2023b**], which extracts and presents electoral results.
- **{auscensus}** [**Yáñez Santibáñez, 2023a**], which allows to interact with Census Data Packs to extract different statistics across geographical units, and across censuses.
- **{aussiemaps}** [**Yáñez Santibáñez, 2023c**], which assists with aggregating census data into electoral divisions, by matching and apportioning different geographical structures.

The appendix contains a vignette for each package, explaining their respective *modus operandi*. At a higher level, the extraction pipeline for this project is represented by figure 4.1.

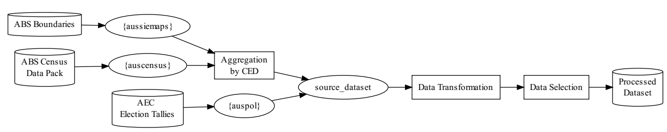


Figure 4.1: Flow of data from sources to dataset

In summary, the process followed consisted of the below steps:

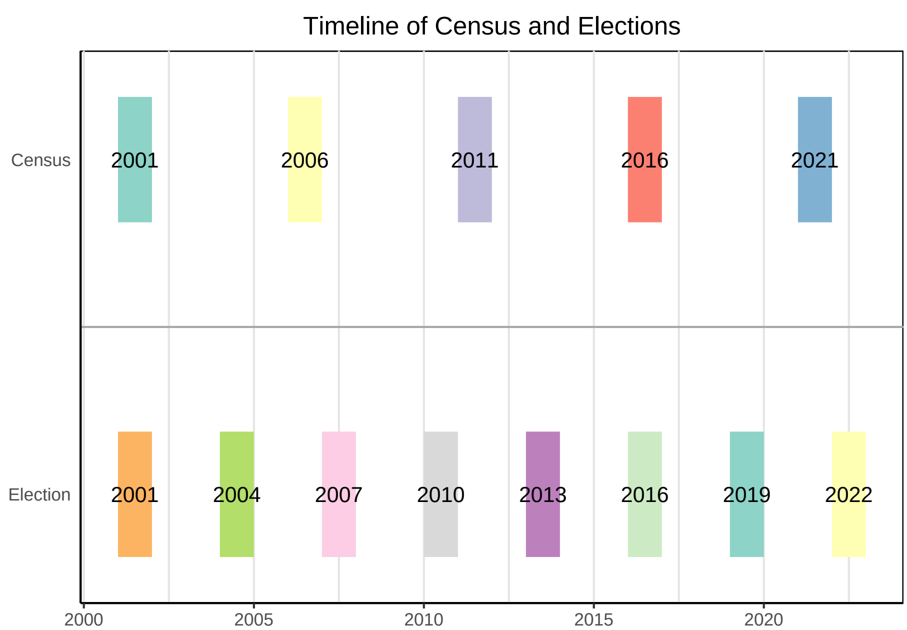
1. Census data was extracted from the respective Census Data Pack using **{auscensus}**. Using the package workflow, key attributes were identified in each census, extracted from the respective files and given common names. Data were extracted for statistical areas and apportioned into Commonwealth Electoral Divisions by overlapping area, with the help of functions written into **{aussiemaps}**

2. Primary vote results for each division were extracted using the **auspol** package.
3. All the data was stored in a local database, from where was extracted and put together in a single dataset.
4. From there, the "raw" data was further processed and stored in a single "consolidated" dataset.

4.2 Data Selection

4.2.1 Census and Election Years

The first to address when extracting the data is to establish a correspondence between census and election data. Since election the census cycle (5 years) does not match the electoral cycle (determined by the incumbent government, with a 3-year term for the House of Representatives), there is a potential problem of the census data not being completely representative of the population on a given election day. Figure **??** presents the best matches between both events held in the 21st century.



Considering the census data available and selecting the elections closer to each census, four pairs of events were selected for data extraction. there are presented in table 4.1

Table 4.1: Selected Census-Election pairings

Census	Election
2006	2007
2010	2011
2016	2016
2021	2022

Please note that this selection will remove half of the elections within the period, which may have an effect on model accuracy. However, since the objective is not to obtain an accurate prediction this has been accepted as a trade-off to avoid having to interpolate demographic attributes between censuses - which is also subject to inaccuracies given the rapid demographic changes experienced in Australia's main cities.

4.2.2 Electoral Data

In the case of the electoral data. not much processing was required. The source data already contains records of primary voting for each electorate and only percentages have been calculated. In addition, the number of total votes per party at the national and state level have been calculated. A sample of the extracted data is presented in table 4.2.

Table 4.2: Sample extraction - Canberra 2022

Year	Division	Abbreviation	Party	Votes	Percentage
2022	Canberra	ALP	ALP	34,574	45.2%
2022	Canberra	GRN	GRN	19,240	25.2%
2022	Canberra	COAL	Liberal (Coalition)	16,264	21.3%
2022	Canberra	Other	Other Parties	6,417	8.4%

4.2.3 Census Data

As mentioned in section 4.1, a major challenge with respect of census data is the large volume of data points collected. For instance, the data pack for the 2022 Census contains 62 different tables, ranging from 8 [1] to 1,590 [2] attributes.

[1] *02 -Selected Medians and Averages*
[2] *09 - Country of Birth of Person by Age by Sex*

To select which variables to extract, literature and journalistic sources were consulted ([Biddle and McAllister, 2022], [Parliament], [Jakubowicz and Ho, a]) to inform an initial set of covariates. In total (XYZ) variables were selected, which correspond to below to the following groups:

1. **Income** : Distribution of population in pre-set income brackets.
2. **Education Level**: Distribution of educational achievement (from incomplete secondary to vocational education and academic degrees).
3. **Age**: Distribution of the population in generational cohorts. Taking into account the selected elections, the four groups of interest are Baby Boomers (1946 to 1964), Generation X (1965 to 1980), Generation Y (1981 to 1996) and Generation Z (1997 to 2021).
4. **Relationship status**: Variables describing civil status (e.g. living alone, married, in a de facto relationship).
5. **Household type:** Descriptors of type of housing , (e.g. standalone house, semi-detached, flats).
6. **Household tenure:** Descriptors of house ownership, rental or other arrangement (e.g. public housing).
7. **Citizenship**: Percentage of the population that hold Australian citizenship. Although non-citizens are not entitled to vote, this variable can be taken as a proxy for relative integration of migrant communities into civic life.
8. **Religion:** Percentage of the population declaring to profess a religion. For this analysis, largest and high growth religious groups were selected (No religion/secular, Roman Catholic, Anglican-Presbyterian-Uniting, Christian Orthodox, Other Christianity, Islam, Hinduism, Buddhism).
9. **Language**: Languages spoken in the community. Similar to religion, a selection of relevant language have been included to reflect the historic and current migrant communities.

Apart from those, each electorate has been classified as **metropolitan** if it lies within the boundaries of Australian capital cities or **non-metropolitan** if not. Altogether, these variables try to reflect wealth and education (cited by [Biddle and McAllister, 2022] as key factors in deciding political persuasion), as well as stage in life and belonging to a particular migrant community (sometimes cited as an influential factor, for instance in [Jakubowicz and Ho, b]).

A sample of the resulting dataset is present in table 4.3.

Table 4.3: Dataset sample

election_year	DivisionNm	ALP	COAL	GRN	Other	Australian_Citizens	Age_Baby_Boomers
2022	Spence	11.70	-10.76	-0.61	-6.58	86.74	17.77
2007	La Trobe	-2.77	3.96	1.24	-2.90	88.64	26.85
2022	Gellibrand	10.18	-8.68	4.59	-6.94	76.50	13.71
2010	Hughes	-0.42	6.28	-5.49	-1.43	89.39	16.16
2010	Eden-Monaro	5.74	-1.52	-2.24	-1.31	89.88	22.55
2016	Chisholm	1.87	2.84	2.13	-5.38	70.26	15.47
2007	Longman	4.57	1.73	-3.44	-2.62	87.59	24.41
2010	Maranoa	-17.22	20.39	-6.20	1.47	90.26	20.56

4.3 Data Exploration

In total, the resulting dataset is made up of 4 response variables and 55 potential predictors, plus identificatory attributes like division mane and election year. As expected, an initial inspection shows that some of the covariates are loosely correlated with primary vote. Also expected, many of the covariates exhibit medium to high correlation levels amongst themselves, e.g. negative correlation between high and low level income groups, and certain age brackets with houselhold type and tenure.

As examples, figure 4.2 show a somewhat weak correlation between Coalition primary vote and percentage of baby boomer population. Figure 4.3 presents the correlation values for religion and language attributes that aside from expected pairings (e.g. Hinduism and South Asian languages or Italian speakers and percentage of declared catholics), there is a almost exclusive positive correlation between membership to Anglican, Presbyterian and Uniting churches and percentage of monolingual English speakers. The percentage of monolingual English speakers is also negative correlated to all other language groups.

Besides from this, it is worth noticing that :

- There is no apparent change in the relationship between a given covariate

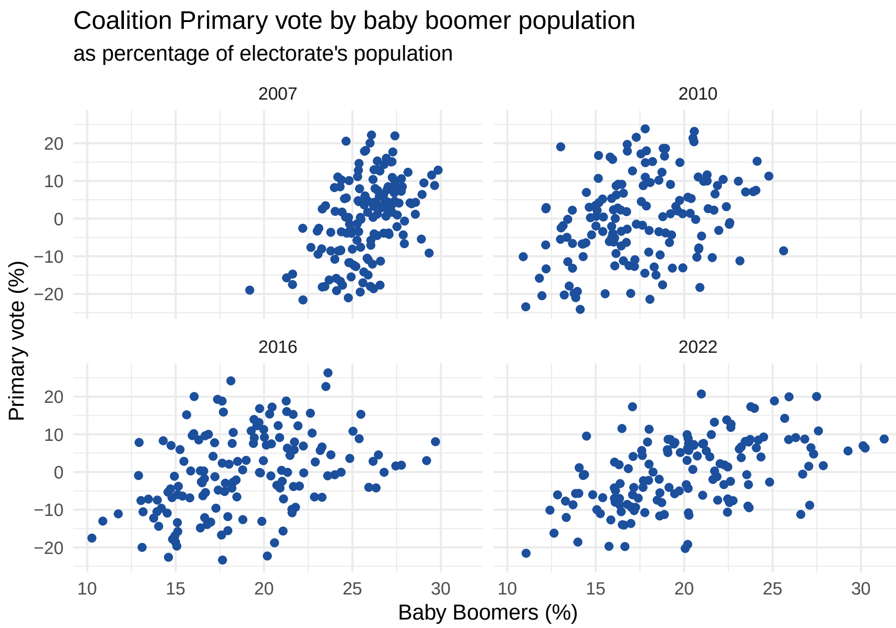


Figure 4.2: Correlation between Coalition vote and Baby boomer population

and the responses when broken down by state or capital city.

- There are also no obviously distinguishable differences when splitting results by each election.

4.3.1 Dimensionality reduction using Multiple Factor Analysis

Given the large number of variables and considering their correlation, it is worth exploring if a change of space could help to better identify variation, and whether the number of covariates can be reduced in a meaningful way. For this **multiple factor analysis** (MFA) [Escofier and Pagès, 2008] was used, given that:

- MFA allows to use variables that belong to groups.
- Allows to combine quantitative and qualitative variables.

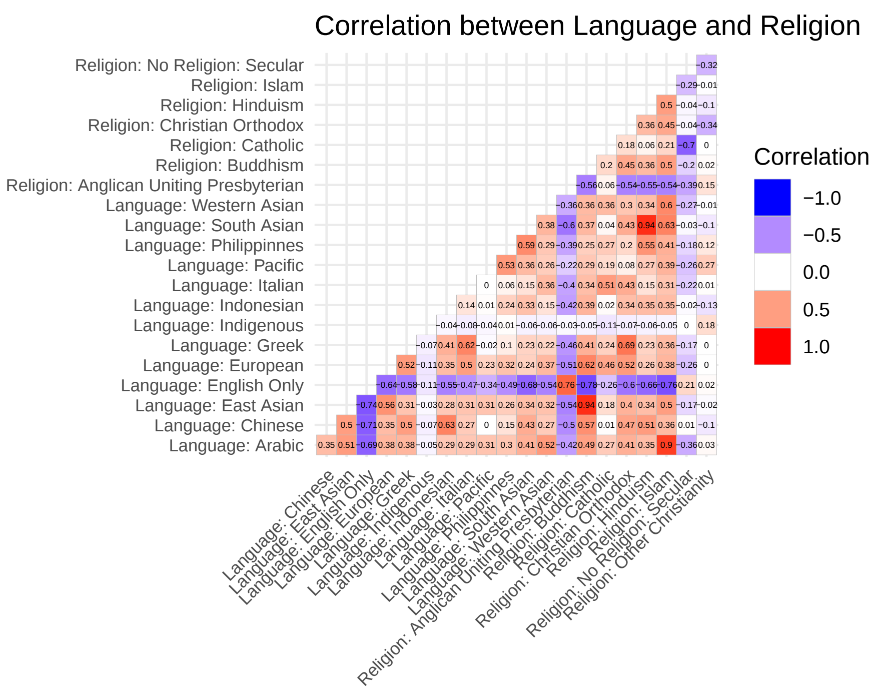


Figure 4.3: Correlation for selected covariates

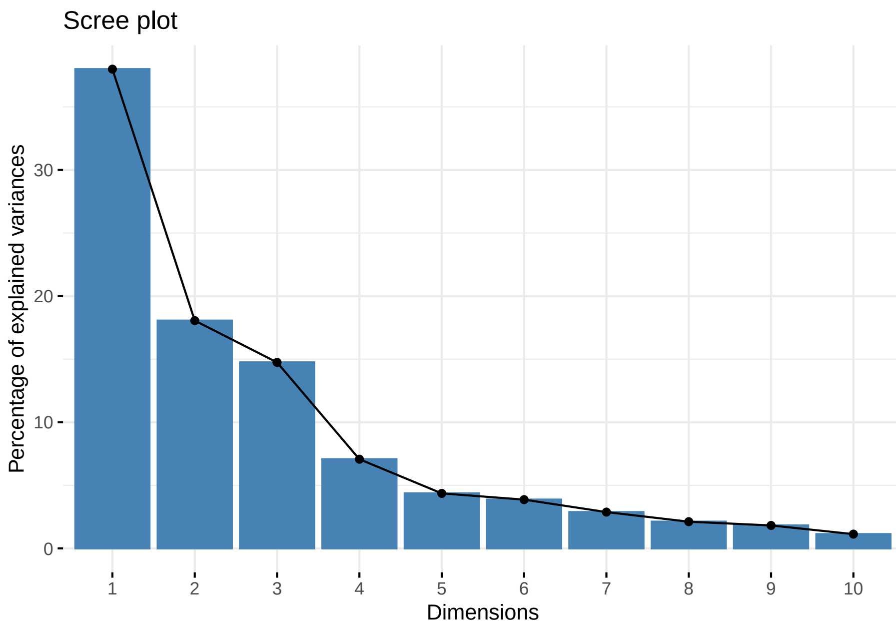


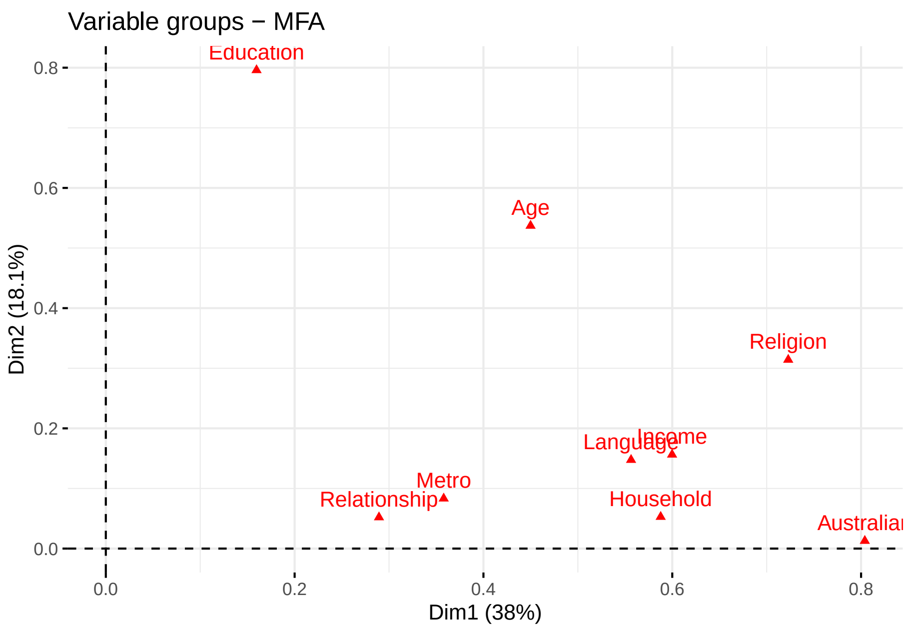
Figure 4.4: Scree plot for MFA

Table 4.4: Eigenvalues and cumulative variance

Dimension	eigenvalue	percentage of variance	cumulative percentage of variance
Dim 1	4.528	37.989	37.989
Dim 2	2.152	18.056	56.045
Dim 3	1.757	14.742	70.788
Dim 4	0.842	7.067	77.855
Dim 5	0.519	4.358	82.213
Dim 6	0.460	3.864	86.076
Dim 7	0.343	2.876	88.952
Dim 8	0.252	2.116	91.069
Dim 9	0.217	1.817	92.886

Table 4.4: Eigenvalues and cumulative variance

Dimension	eigenvalue	percentage of variance	cumulative percentage of variance
Dim 10	0.135	1.129	94.015



Chapter 5

(APPENDIX) {auscensus} Vignette

Extracted from https://carlosyanez.github.io/auscensus/articles/complex_case.html on Sunday 22 January 2023

This vignette shows a more complex use case of auscensus. Let's assume we want to extract the percentage of Australian Citizens for all Commonwealth Electoral Divisions, as measured in last 4 Censuses (2006-2021).

An initial exploration shows that this data can be found in table 01 (across all four censuses) - which provided an statistical summary. However, is not published aggregated by electorate across all censuses.

```
## # A tibble: 4 x 3
##   Year  table_number CED
##   <chr> <chr>        <lgl>
## 1 2011  01           TRUE
## 2 2016  01           TRUE
## 3 2021  01           TRUE
## 4 2006  01           NA
```

Therefore, we will retrieve the data from the lowest statistical unit. However, SA1 were not available in 2006 - where the smallest area was a "CD".

```
## # A tibble: 4 x 4
##   Year  table_number CD    SA1
##   <chr> <chr>        <lgl> <lgl>
## 1 2006  01           TRUE  NA
## 2 2011  01           NA    TRUE
## 3 2016  01           NA    TRUE
## 4 2021  01           NA    TRUE
```

The next step is to figure the attributes for the numbers of Australian citizen and total population, which are presented below:

```
## # A tibble: 6 x 3
##   Table Attribute                   Year
##   <chr> <chr>                       <chr>
## 1 01    Total Persons_males         2006
## 2 01    Total Persons_females       2006
## 3 01    Total Persons_persons       2006
## 4 01    Age Groups: 0-4 Years_males   2006
## 5 01    Age Groups: 0-4 Years_females 2006
## 6 01    Age Groups: 0-4 Years_persons 2006
```

```
## # A tibble: 4 x 5
##   Attribute                   `2006` `2011` `2016` `2021`
##   <chr>                       <lgl>  <lgl>  <lgl>  <lgl>
## 1 Australian Citizen_persons  TRUE   NA     NA     NA
## 2 Australian_citizen_persons  NA     TRUE   TRUE   TRUE
## 3 Total Persons_persons       TRUE   NA     NA     NA
## 4 Total_persons_persons       NA     TRUE   TRUE   TRUE
```

Using *attribute_tibble_to_list*, this data frame can be converted into the required format.

Now, we can cycle through the four censuses and extract the data. Please note that CDs and SA1s are not equivalent, but they are stored together for convenience:

To aggregate the data, **aussiemaps::geo_aggregate()** can help using area to apportion on non-overalpping cases. Then, this package's *calculate_percentage()* will take the totals from the list and calculate percentages.

```
## # A tibble: 601 x 6
##    Unit       Year Attribute             Value  Total Percentage
##    <chr>     <dbl> <chr>                 <dbl>  <dbl>      <dbl>
##  1 Adelaide   2006 Australian Citizens 115908 140151       82.7
##  2 Aston      2006 Australian Citizens 115010 129463       88.8
##  3 Ballarat   2006 Australian Citizens 118247 129221       91.5
##  4 Banks      2006 Australian Citizens 104181 118371       88.0
##  5 Barker     2006 Australian Citizens 133033 145410       91.5
##  6 Barton     2006 Australian Citizens 105287 131234       80.2
##  7 Bass       2006 Australian Citizens  85631  94270       90.8
##  8 Batman     2006 Australian Citizens 105561 128107       82.4
##  9 Bendigo    2006 Australian Citizens 121930 131885       92.5
## 10 Bennelong  2006 Australian Citizens 104473 129340       80.8
## # i 591 more rows
```

Chapter 6

(APPENDIX) {auspol} Vignette

Extracted from https://carlosyanez.github.io/auspol/articles/house_primary_vote.html on Sunday 22 January 2023

auspol includes two functions to interact with the preference distribution data:

- get_house_primary_vote()
- house_primary_vote_summary()
- house_primary_comparison_plot()
- house_primary_historic_plot()

6.1 What is this?

If you are unfamiliar with the Australian electoral system and preferential voting, please look at this [explainer(https://www.aec.gov.au/learn/preferential-voting.html) before proceeding.

6.2 Getting the data

get_house_primary_vote() is the basic function to retrieve primary vote data published by the AEC. Without any arguments, it will deliver all the results for all elections, but it comes with parameters to facilitate filtering. For instance, to get the results for Brisbane for 2022:

```
## # A tibble: 344 x 17
##     Year StateAb DivisionID DivisionNm PollingPlaceID PollingPlace CandidateID Surname  GivenN
##    <dbl> <chr>        <int> <chr>               <int> <chr>              <int> <chr>    <chr>
##  1  2022 QLD            156 Brisbane            83397 Alderley           37204 KENNEDY  Tiana
```

```
## 2  2022 QLD            156 Brisbane             83397 Alderley             35972 KNUDS
## 3  2022 QLD            156 Brisbane             83397 Alderley             37338 BATES
## 4  2022 QLD            156 Brisbane             83397 Alderley             37230 JARR
## 5  2022 QLD            156 Brisbane             83397 Alderley             37482 EVANS
## 6  2022 QLD            156 Brisbane             83397 Alderley             38213 HOLD
## 7  2022 QLD            156 Brisbane             83397 Alderley             37311 BULL
## 8  2022 QLD            156 Brisbane             83397 Alderley               999 Infor
## 9  2022 QLD            156 Brisbane              6017 Ascot                37204 KENN
## 10 2022 QLD            156 Brisbane              6017 Ascot                35972 KNUDS
## # i 334 more rows
## # i 4 more variables: PartyNm <chr>, OrdinaryVotes <int>, Swing <dbl>, SittingMembe
```

Both parameters can include more than one value, e.g.

```
## # A tibble: 712 x 17
##     Year StateAb DivisionID DivisionNm PollingPlaceID PollingPlace CandidateID Surna
##    <dbl> <chr>        <int> <chr>               <int> <chr>              <int> <chr
##  1  2022 QLD            156 Brisbane            83397 Alderley           37204 KENN
##  2  2022 QLD            156 Brisbane            83397 Alderley           35972 KNUDS
##  3  2022 QLD            156 Brisbane            83397 Alderley           37338 BATES
##  4  2022 QLD            156 Brisbane            83397 Alderley           37230 JARR
##  5  2022 QLD            156 Brisbane            83397 Alderley           37482 EVANS
##  6  2022 QLD            156 Brisbane            83397 Alderley           38213 HOLD
##  7  2022 QLD            156 Brisbane            83397 Alderley           37311 BULL
##  8  2022 QLD            156 Brisbane            83397 Alderley             999 Infor
##  9  2022 QLD            156 Brisbane             6017 Ascot              37204 KENN
## 10  2022 QLD            156 Brisbane             6017 Ascot              35972 KNUDS
## # i 702 more rows
## # i 4 more variables: PartyNm <chr>, OrdinaryVotes <int>, Swing <dbl>, SittingMembe
```

```
## # A tibble: 1,783 x 17
##     Year StateAb DivisionID DivisionNm PollingPlaceID PollingPlace CandidateID Surna
##    <dbl> <chr>        <int> <chr>               <int> <chr>              <int> <chr
##  1  2022 QLD            156 Brisbane            83397 Alderley           37204 KENN
##  2  2022 QLD            156 Brisbane            83397 Alderley           35972 KNUDS
##  3  2022 QLD            156 Brisbane            83397 Alderley           37338 BATES
##  4  2022 QLD            156 Brisbane            83397 Alderley           37230 JARR
##  5  2022 QLD            156 Brisbane            83397 Alderley           37482 EVANS
##  6  2022 QLD            156 Brisbane            83397 Alderley           38213 HOLD
##  7  2022 QLD            156 Brisbane            83397 Alderley           37311 BULL
##  8  2022 QLD            156 Brisbane            83397 Alderley             999 Infor
##  9  2022 QLD            156 Brisbane             6017 Ascot              37204 KENN
## 10  2022 QLD            156 Brisbane             6017 Ascot              35972 KNUDS
## # i 1,773 more rows
## # i 4 more variables: PartyNm <chr>, OrdinaryVotes <int>, Swing <dbl>, SittingMembe
```

By default, the results are presented by polling place, with the possibility to aggregate them.

```
## # A tibble: 37 x 14
##     Year StateAb DivisionID DivisionNm CandidateID Surname   GivenNm  BallotPosition Elected H
##    <dbl> <chr>        <int> <chr>            <int> <chr>     <chr>             <int> <lgl>   <
##  1  2019 QLD            156 Brisbane           999 Informal  Informal            999 FALSE   F
##  2  2019 QLD            156 Brisbane         32751 PERRY     Anne                  1 FALSE   F
##  3  2019 QLD            156 Brisbane         32946 NEWBURY   Paul                  6 FALSE   F
##  4  2019 QLD            156 Brisbane         32960 WHITTAKER Aaron                 3 FALSE   F
##  5  2019 QLD            156 Brisbane         33144 BARTLETT  Andrew                4 FALSE   F
##  6  2019 QLD            156 Brisbane         33206 EVANS     Trevor                2 TRUE    T
##  7  2019 QLD            156 Brisbane         33224 EMANUEL   Kamala                7 FALSE   F
##  8  2019 QLD            156 Brisbane         33326 JEANNERET Rod                   5 FALSE   F
##  9  2019 WA             245 Perth              999 Informal  Informal            999 FALSE   F
## 10  2019 WA             245 Perth            32155 PERKS     Caroline              6 FALSE   F
## # i 27 more rows
## # i 1 more variable: OrdinaryVotes <int>
```

```
## # A tibble: 12 x 17
##     Year StateAb DivisionID DivisionNm PollingPlaceID PollingPlace CandidateID Surname     Giv
##    <dbl> <chr>        <int> <chr>               <int> <chr>              <int> <chr>       <ch
##  1  2022 WA             245 Perth                8203 Yokine North       37417 BAILEY      Cam
##  2  2022 WA             245 Perth                8203 Yokine North       36515 POWELL      Dea
##  3  2022 WA             245 Perth                8203 Yokine North       37748 CONNOR      Sea
##  4  2022 WA             245 Perth                8203 Yokine North       37803 VOS         Dav
##  5  2022 WA             245 Perth                8203 Yokine North       37233 SZMEKURA-M~ Sar
##  6  2022 WA             245 Perth                8203 Yokine North       37327 GORMAN      Pat
##  7  2022 WA             245 Perth                8203 Yokine North       37273 NICKOLS     Eva
##  8  2022 WA             245 Perth                8203 Yokine North       36507 EBERHART    Son
##  9  2022 WA             245 Perth                8203 Yokine North       36628 PERKS       Car
## 10  2022 WA             245 Perth                8203 Yokine North       36601 DWYER       Dav
## 11  2022 WA             245 Perth                8203 Yokine North       37290 GYURU       Aid
## 12  2022 WA             245 Perth                8203 Yokine North         999 Informal    Inf
## # i 4 more variables: PartyNm <chr>, OrdinaryVotes <int>, Swing <dbl>, SittingMemberFl <lgl>
```

It is also possible to restrict the results to selected polling places

Additionally, it is possible to select one or more states instead of a group of divisions, e.g.:

```
## # A tibble: 86 x 14
##     Year StateAb DivisionID DivisionNm CandidateID Surname  GivenNm    BallotPosition Elected
##    <dbl> <chr>        <int> <chr>            <int> <chr>    <chr>               <int> <lgl>
##  1  2019 TAS            192 Bass               999 Informal Informal              999 FALSE
##  2  2019 TAS            192 Bass             32124 ARCHER   Bridget                 4 TRUE
##  3  2019 TAS            192 Bass             32327 HART     Ross                    2 FALSE
##  4  2019 TAS            192 Bass             32379 WOODBURY Susan                   3 FALSE
##  5  2019 TAS            192 Bass             32399 COOPER   Carl                    7 FALSE
##  6  2019 TAS            192 Bass             32540 HALL     Tom                     1 FALSE
##  7  2019 TAS            192 Bass             32545 ROARK    Allan John              6 FALSE
```

```
## 8  2019 TAS           192 Bass           33590 LAMBERT  Todd
## 9  2019 TAS           193 Braddon          999 Informal Informal            9
## 10  2019 TAS           193 Braddon        32094 BRAKEY   Craig
## # i 76 more rows
## # i 1 more variable: OrdinaryVotes <int>
```

It is also possible to filter results by one or more parties:

```
## # A tibble: 8 x 14
##    Year StateAb DivisionID DivisionNm CandidateID Surname    GivenNm   BallotPositi
##   <dbl> <chr>        <int> <chr>            <int> <chr>      <chr>             <in
## 1  2019 NT             306 Lingiari         32740 SNOWDON    Warren
## 2  2019 NT             306 Lingiari         33045 PRICE      Jacinta
## 3  2019 NT             307 Solomon          32743 GOSLING    Luke John
## 4  2019 NT             307 Solomon          33053 GANLEY     Kathy
## 5  2022 NT             306 Lingiari         36968 RYAN       Damien
## 6  2022 NT             306 Lingiari         37286 SCRYMGOUR  Marion
## 7  2022 NT             307 Solomon          36937 MACFARLANE Tina
## 8  2022 NT             307 Solomon          37280 GOSLING    Luke
## # i 1 more variable: OrdinaryVotes <int>
```

house_primary_vote_summary() builds on the basic function and summarises data .

```
## # A tibble: 8 x 11
##    Year StateAb DivisionNm PartyAb  PartyNm                              OrdinaryVo
##   <dbl> <chr>   <chr>      <chr>    <chr>                                        <i
## 1  2022 QLD     Brisbane   AJP      Animal Justice Party                          1
## 2  2022 QLD     Brisbane   ALP      Australian Labor Party                       20
## 3  2022 QLD     Brisbane   GRN      Queensland Greens                            20
## 4  2022 QLD     Brisbane   Informal Informal                                      1
## 5  2022 QLD     Brisbane   LDP      Liberal Democrats                             1
## 6  2022 QLD     Brisbane   LNP      Liberal National Party of Queensland         26
## 7  2022 QLD     Brisbane   ON       Pauline Hanson's One Nation                   1
## 8  2022 QLD     Brisbane   UAPP     United Australia Party                        1
## # i abbreviated name: 1: Percentage_with_Informal
```

Using the previous filters, it is possible to get ad-hoc summaries, for instance - all the ALP votes in Queensland in 2022, or the historic Liberal vote in Franklin.

```
## # A tibble: 30 x 11
##     Year StateAb DivisionNm  PartyAb PartyNm                OrdinaryVotes GivenNm S
##    <dbl> <chr>   <chr>       <chr>   <chr>                          <int> <chr>   <
##  1  2022 QLD     Blair       ALP     Australian Labor Party         27323 Shayne  N
##  2  2022 QLD     Bonner      ALP     Australian Labor Party         20930 Tabatha Y
##  3  2022 QLD     Bowman      ALP     Australian Labor Party         23196 Donisha D
##  4  2022 QLD     Brisbane    ALP     Australian Labor Party         20346 Madonna J
##  5  2022 QLD     Capricornia ALP     Australian Labor Party         20543 Russell R
##  6  2022 QLD     Dawson      ALP     Australian Labor Party         18921 Shane   H
```

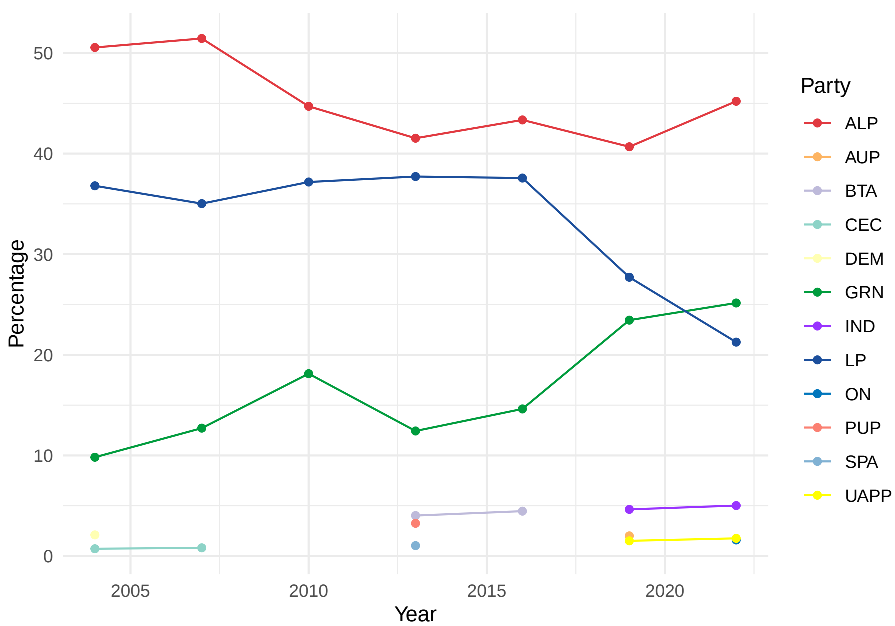
```
## 7  2022 QLD     Dickson     ALP     Australian Labor Party         22988 Ali      FRANCE
## 8  2022 QLD     Fadden      ALP     Australian Labor Party         18140 Letitia DEL FABBRO
## 9  2022 QLD     Fairfax     ALP     Australian Labor Party         18001 Sue      FERGUSON
## 10 2022 QLD     Fisher      ALP     Australian Labor Party         19804 Judene   ANDREWS
## # i 20 more rows
```

```
## # A tibble: 7 x 11
##    Year StateAb DivisionNm PartyAb PartyNm OrdinaryVotes GivenNm      Surname Percentage_with_
##   <dbl> <chr>   <chr>      <chr>   <chr>           <int> <chr>        <chr>
## 1  2004 TAS     Franklin   LP      Liberal         21337 Henry        FINNIS
## 2  2007 TAS     Franklin   LP      Liberal         22616 Vanessa      GOODWIN
## 3  2010 TAS     Franklin   LP      Liberal         18386 Jane         HOWLETT
## 4  2013 TAS     Franklin   LP      Liberal         21867 Bernadette   BLACK
## 5  2016 TAS     Franklin   LP      Liberal         20754 Amanda-Sue   MARKHAM
## 6  2019 TAS     Franklin   LP      Liberal         18591 Dean         YOUNG
## 7  2022 TAS     Franklin   LP      Liberal         14374 Kristy Maree JOHNSON
```

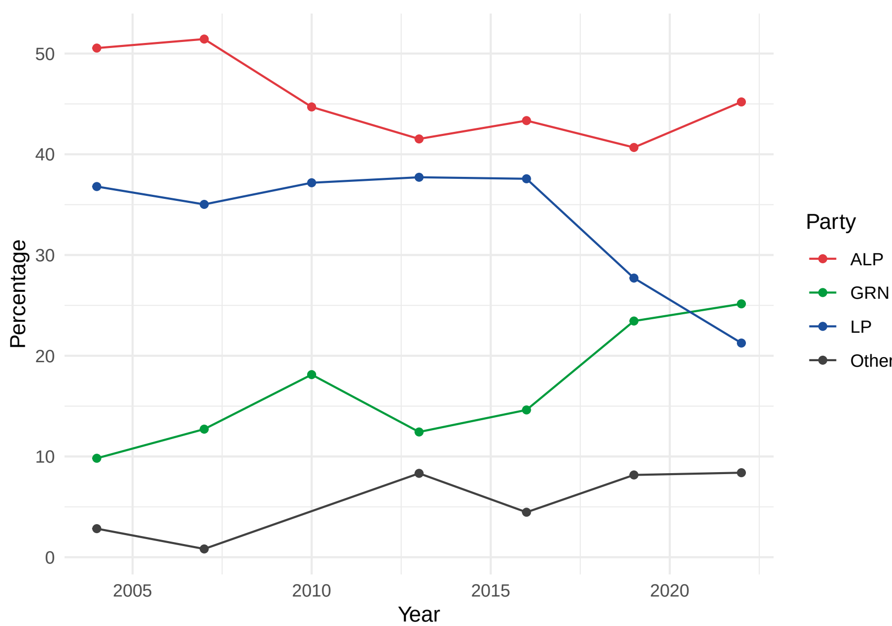
6.3 Plotting

6.3.1 Historic Trends

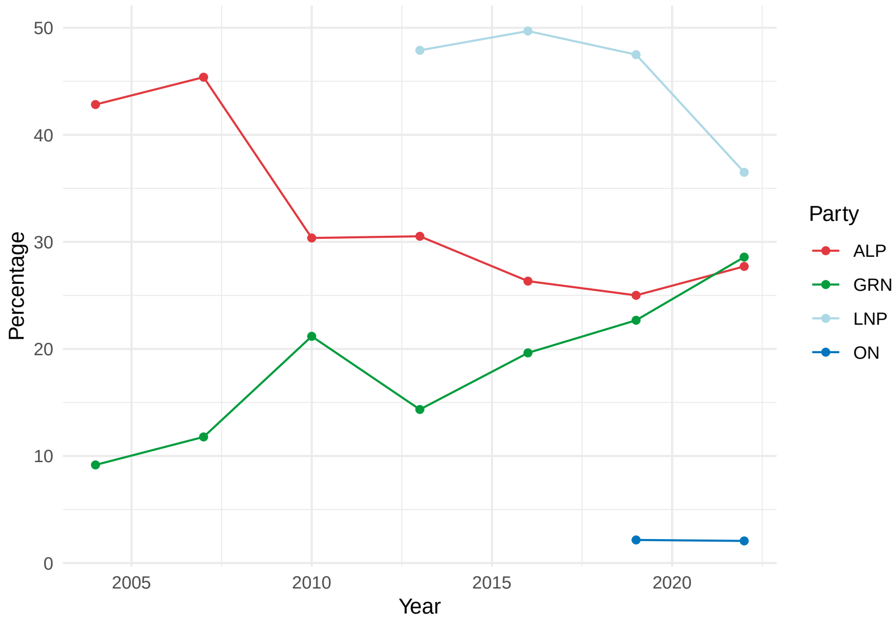
The first plotting convenience function in this package allows comparing the evolution of primary voting across time. This function relies on house_primary_summary and uses many of its options. Its first use is to represent party trends in one electorate:



As they can be many minor parties, it is sometimes useful just to focus on a number of parties. This function allows filtering by a number of parties or by filtering by the most voted in a certain year. In both cases, it is possible to consolidate others' votes.

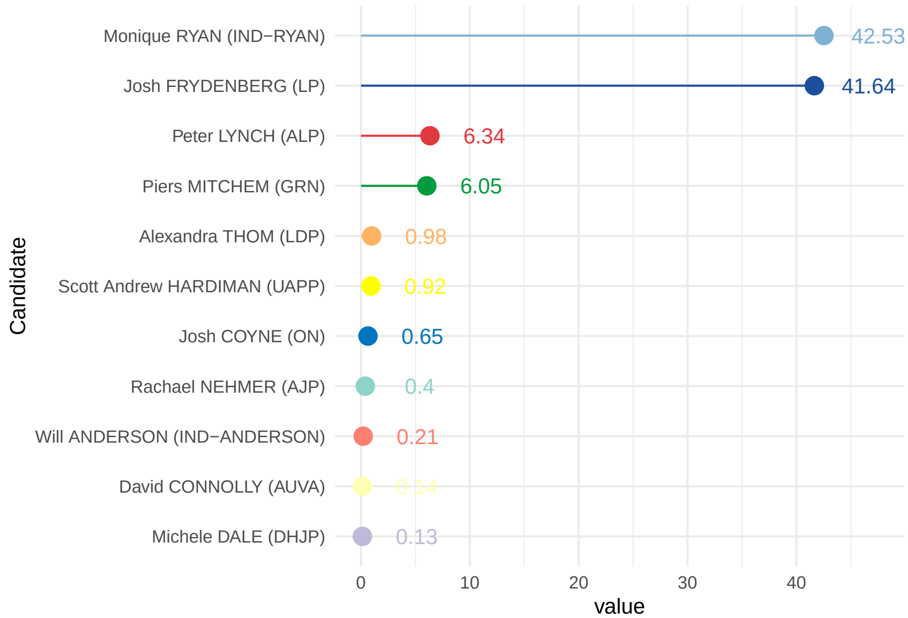


Finally, it is possible to aggregate party acronyms - sometimes the same party has changed named or registered differently

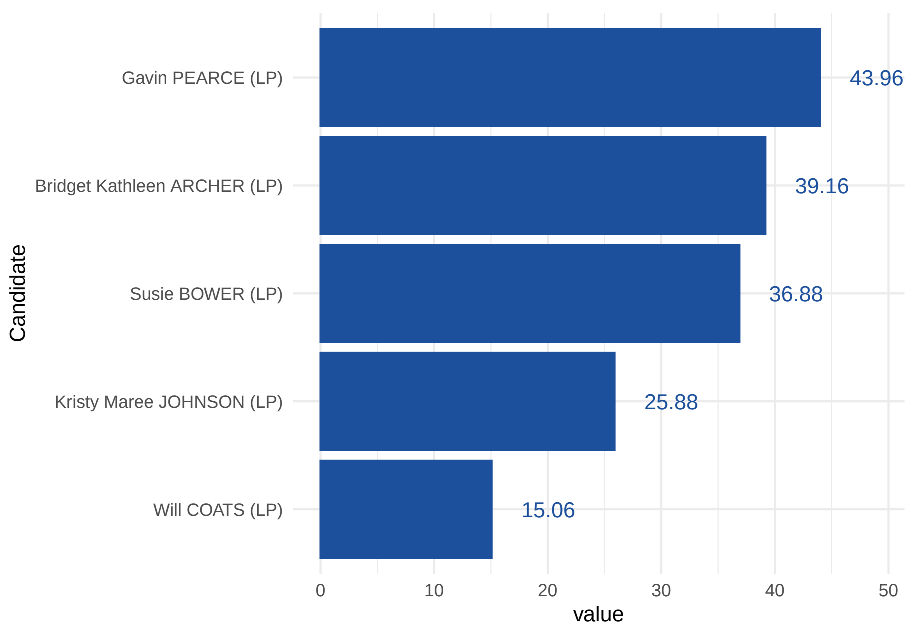


6.4 Results for one election

This package also contains a convenience function to look at the primary vote results for one division. Lile the previous function, this also inherits many of the attributes of *get_house_primary_vote*.



The plots can also be displayed using bars, as shown below



Chapter 7

(APPENDIX) {aussiemaps} Vignette

Extracted from https://carlosyanez.github.io/aussiemaps/articles/aussiemaps.html on Sunday 22 January 2023

7.1 {aussiemaps} - Yet another maps package

This package has been built to facilitate the use of the geographic boundary files published by the Australian Bureau of Statistics (ABS). The ABS has published several boundary files - i.e. the Australian Statistical Geography Standard (ASGS) from 2006 onwards and the Australian Standard Geographical Classification (ASGC) before that - covering both:

- Statistical Geographic Structures created and maintained by the ABS - and used to collect data.
- Non-ABS structure, e.g Postal Areas, Electoral Divisions, LGA boundaries.

This package has four versions of the above, aligned with Census years 2006, 2011,2016 and 2021. This makes it easy to mix use with Census data packs or the {auscensus} package.

This package provides access to a processed version of those boundaries - as sf objects, allowing it to cater for the following scenarios:

- Get the boundaries of an electoral division across time.
- Get all the S1 or S1 areas within a Council area.
- Get all postcodes in a state or territory.

This repository also contains the R script used to process the files. Although not tested, the functions could also accommodate BYO structures for other years.

7.2 Getting started.

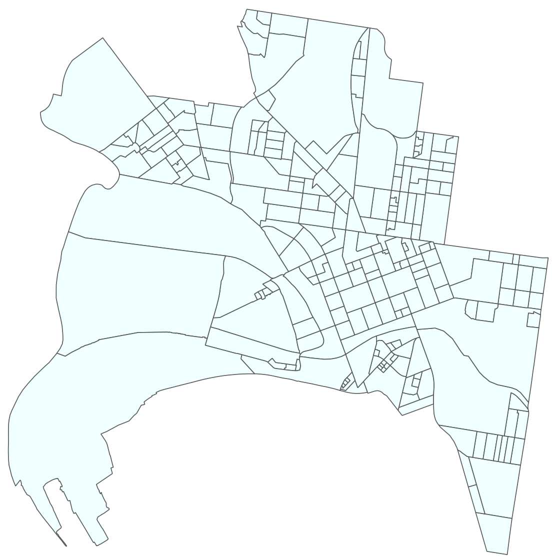
The core function of this package is get_map(), which retrieves the sf files. get_map provides several filters to narrow down the data retrieved and avoid getting everything unless is needed. The key parameters for this function are:

- How the data will be filtered (e.g. return only objects in a particular state, council or metro area)
- Which year/version of the data will be retrieved?
- Which aggregation will be used (e.g. which will be the resulting objects)

Filters and column names follow the same name convention used in the original ABS files. The function list_attributes(), will present them in tibble format:

```
## # A tibble: 10 x 5
##    attributes `2006`         `2011`         `2016`         `2021`
##    <chr>      <chr>          <chr>          <chr>          <chr>
##  1 CD_CODE    CD_CODE_2006   <NA>           <NA>           <NA>
##  2 CED_CODE   CED_CODE_2006  CED_CODE_2011  CED_CODE_2016  CED_CODE_2021
##  3 CED_NAME   CED_NAME_2006  CED_NAME_2011  CED_NAME_2016  CED_NAME_2021
##  4 IARE_CODE  IARE_CODE_2006 IARE_CODE_2011 IARE_CODE_2016 IARE_CODE_2021
##  5 IARE_NAME  IARE_NAME_2006 IARE_NAME_2011 IARE_NAME_2016 IARE_NAME_2021
##  6 ILOC_CODE  ILOC_CODE_2006 ILOC_CODE_2011 ILOC_CODE_2016 ILOC_CODE_2021
##  7 ILOC_NAME  ILOC_NAME_2006 ILOC_NAME_2011 ILOC_NAME_2016 ILOC_NAME_2021
##  8 IREG_CODE  IREG_CODE_2006 IREG_CODE_2011 IREG_CODE_2016 IREG_CODE_2021
##  9 IREG_NAME  IREG_NAME_2006 IREG_NAME_2011 IREG_NAME_2016 IREG_NAME_2021
## 10 LGA_CODE   LGA_CODE_2006  LGA_CODE_2011  LGA_CODE_2016  LGA_CODE_2021
```

Let's say we want to retrieve all SA1 in the City of Melbourne for 2016 - this can be done via:



7.3 Filtering via regular expressions

The filter arguments are intended to be regular expressions, for instance:

```
## Simple feature collection with 8 features and 3 fields
## Geometry type: POLYGON
## Dimension:     XY
## Bounding box:  xmin: 115.6286 ymin: -41.3658 xmax: 152.0004 ymax: -20.34465
## Geodetic CRS:  GDA94
##               SSC_NAME_2016                           UCL_NAME_2016     STE_NAME_2016
## 1                  Prestons                                  Sydney   New South Wales POLYGON ((1
## 2  Preston (Toowoomba - Qld)  Remainder of State/Territory (Qld)        Queensland POLYGON ((1
## 3 Preston (Whitsunday - Qld)  Remainder of State/Territory (Qld)        Queensland POLYGON ((1
## 4            Preston (Tas.) Remainder of State/Territory (Tas.)          Tasmania POLYGON ((1
## 5             South Preston Remainder of State/Territory (Tas.)          Tasmania POLYGON ((1
## 6             Preston Beach   Remainder of State/Territory (WA) Western Australia POLYGON ((1
## 7        Preston Settlement   Remainder of State/Territory (WA) Western Australia POLYGON ((1
## 8            Preston (Vic.)                               Melbourne          Victoria POLYGON ((1
```

Whereas

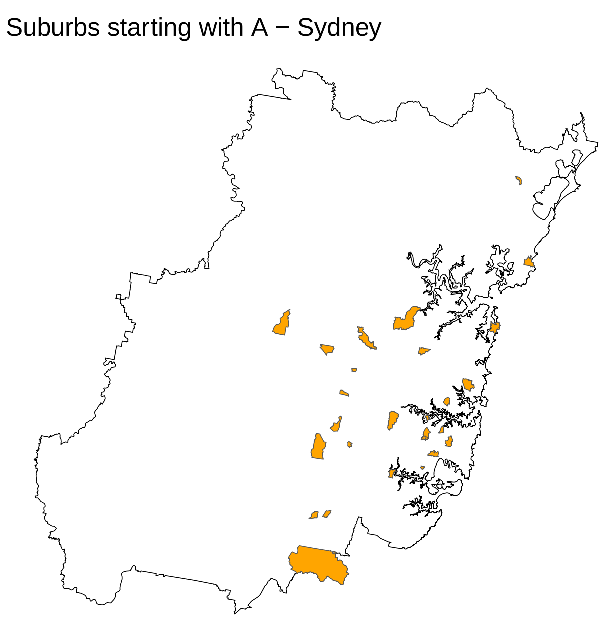
```
## Simple feature collection with 3 features and 3 fields
## Geometry type: POLYGON
## Dimension:     XY
## Bounding box:  xmin: 146.0066 ymin: -41.33851 xmax: 150.8979 ymax: -33.9263
```

```
## Geodetic CRS:  GDA94
##      SSC_NAME_2016                           UCL_NAME_2016   STE_NAME_2016
## 1         Prestons                                    Sydney New South Wales POLYGON ((150
## 2 Preservation Bay Remainder of State/Territory (Tas.)        Tasmania POLYGON ((146
## 3   Preston (Tas.) Remainder of State/Territory (Tas.)        Tasmania POLYGON ((146
```

7.4 Even more complex filtering

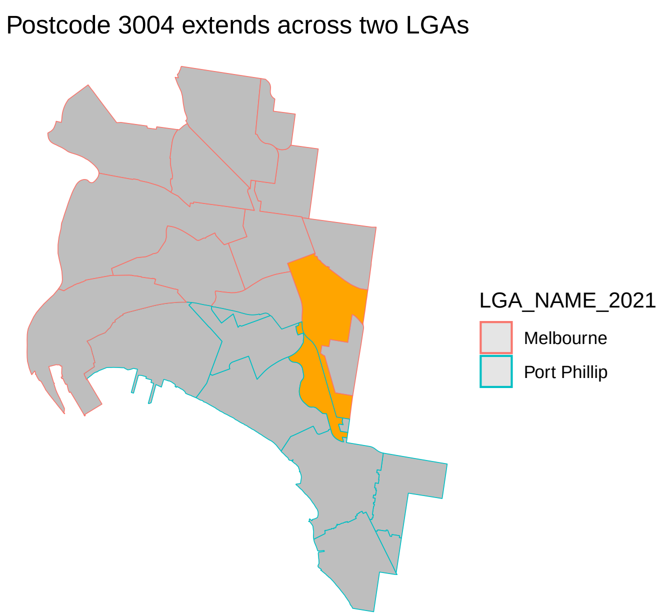
If more complex subsetting is needed, it is possible to pass a table with the elements to be selected. In order to do that, list_structure() comes to help. This function uses the same year and filters parameters than get_map() (actually this function calls the former if no table is provided). Once you have the dataset, you can use any ad-hoc filter to get the needed structures. For example

```
## Reading layer `cache_2021_6766fccc' from data source `C:\Users\carlo\OneDrive\Docume
## Simple feature collection with 1 feature and 36 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: 149.9719 ymin: -34.33116 xmax: 151.6306 ymax: -32.99606
## Geodetic CRS:  GDA2020
```



7.5 Aggregation

It is worth noticing that the *aggregation* parameter accepts more than one variable. Those parameters are passed to dplyr::group_by() before aggregation - thus more variables will impact how sf objects are aggregated. For instance, if we look at the postal areas (ABS approximation of a postcode) in the cities of Melbourne and Port Phillip:

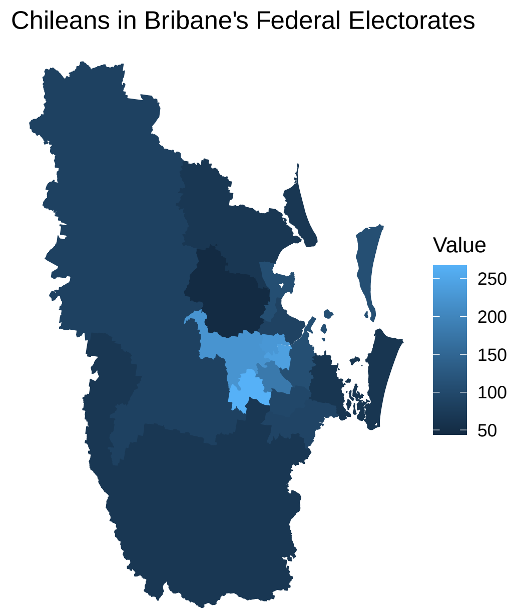


Using external data

This package provides sf data, thus the result can be easily merged with any other data frame. Since data has been taken from the ABS and the output contains both names and **codes** of geographic structures, data can be joined using an un-ambiguous key. Furthermore, with {auscensus}, this package can be used as data filters to retrieve said data in the first place. For example:

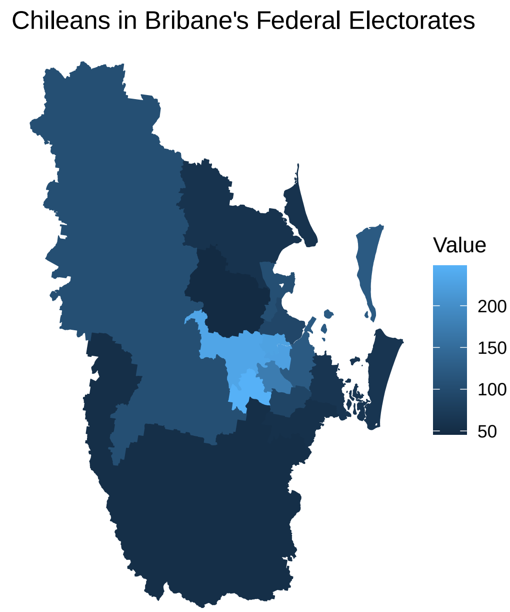
```
## Reading layer `cache_2021_4ec18365' from data source `C:\Users\carlo\OneDrive\Documents\.aussi
## Simple feature collection with 15 features and 36 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: 152.0734 ymin: -28.36387 xmax: 153.5467 ymax: -26.45233
## Geodetic CRS:  GDA2020
```

```
##  [1]  85 109  64 228  44  90 241  87  66 180 267 107  96 223  66
```



7.6 Data Aggregation

As a bonus function, *geo_aggregate()* aggregates data, transforming between geographic structures. For instance, let's imagine that for the previous case, it is only possible to get data by SA2. *geo_aggregate()* can aggregate the data to obtain an approximation for each electorate. When an SA1 is not fully contained by an electorate, the function will use the overlapping area as the weighting factor.


Chileans in Bribane's Federal Electorates
Value
200
150
100
50

Bibliography

Nicholas Biddle and Ian McAllister. Explaining the 2022 australian federal election result. Technical report, 06 2022. URL https://apo.org.au/node/318286.

Australian Electoral Commission. Website, 2023a. URL https://www.aec.gov.au/.

Australian Electoral Commission. Tally room archive, 2023b. URL https://results.aec.gov.au/.

Brigitte Escofier and Jérome Pagès. *Analyses factorielles simples et multiples. Objectifs méthodes et interprétation*. Sciences Sup. Dunod, 2008. URL https://hal.science/hal-00382085.

Andrew Jakubowicz and Christina Ho. Was there an 'ethnic vote' in the 2019 election and did it make a difference?, a. URL http://theconversation.com/was-there-an-ethnic-vote-in-the-2019-election-and-did-it-make-a-difference-117911.

Andrew Jakubowicz and Christina Ho. Was there an 'ethnic vote' in the 2019 election and did it make a difference?, b. URL http://theconversation.com/was-there-an-ethnic-vote-in-the-2019-election-and-did-it-make-a-difference-117911.

Australian Bureau of Statistics. Website, 2023a. URL https://abs.gov.au/.

Australian Bureau of Statistics. Census data packa, 2023b. URL https://abs.gov.au/census/find-census-data/datapacks/.

Commonwealth Parliament. Voting patterns by generation. URL https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/FlagPost/2022/April/Voting_patterns_by_generation. Archive Location: Australia Last Modified: 2022-04-29 Publisher: corporateName=Commonwealth Parliament; address=Parliament House, Canberra, ACT, 2600; contact=+61 2 6277 7111.

Carlos Yáñez Santibáñez. *auscensus: Access Australian Census Data (2006-*

2021), 2023a. URL https://carlosyanez.github.io/auscensus/. R package version 0.0.1.0008.

Carlos Yáñez Santibáñez. *auspol: Australian Federal Election Results (2004-2022)*, 2023b. URL https://carlosyanez.github.io/auspol/. R package version 0.0.1.0000.

Carlos Yáñez Santibáñez. *aussiemaps: Maps of Australia*, 2023c. URL https://carlosyanez.github.io/aussiemaps/. R package version 0.2.0.0013.