MONASH University

ETC5512: Wild Caught Data

Combining Australian census and election data

Lecturer: *Kate Saunders*

Department of Econometrics and Business Statistics

✉ ETC5512.Clayton-x@monash.edu

📅 Week 6

Today you will:

- look at the ABS geographical boundaries for the 2021 census
- integrate data from different sources (census and election) to make exploratory inferences

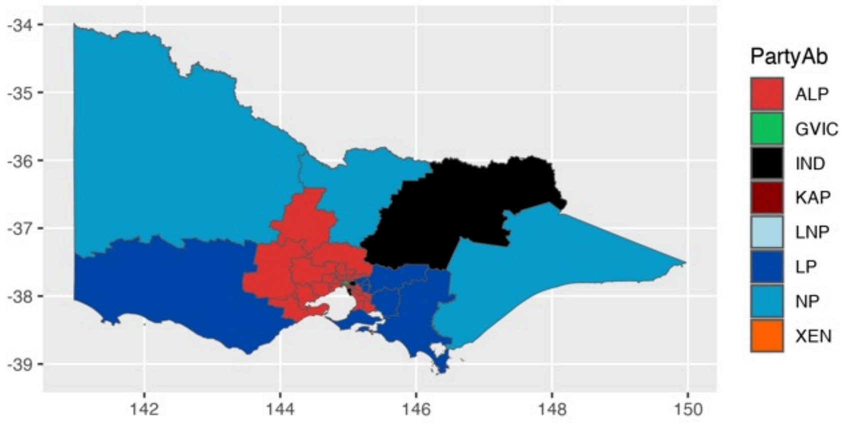
Coding Perspective:

- Further expand our understanding of how to read and use spatial data in R
- Better understand how spatial data is organised
- Learn how to intersect two spatial objections
- Practice re-projecting maps

Recall 2021 Federal Election Data

```r
library(tidyverse)
library(sf)
aec_map <- read_sf(here::here("data/vic-july-2021-esri/E_VIC21_region.shp"))
votes <- read_csv("https://results.aec.gov.au/27966/Website/Downloads/HouseDopByDivisionDownload-27966.csv", skip = 1)

electoral_winners = votes |>
  mutate(DivisionNm = toupper(DivisionNm)) |>
  filter(Elected == "Y") |>
  select(PartyAb, DivisionID, DivisionNm, Elected) |>
  distinct()

aec_map = aec_map |>
  mutate(Elect_div = toupper(Elect_div))

winners_with_map = electoral_winners |>
  left_join(aec_map, by = c("DivisionNm" = "Elect_div"))

aus_colours <- c(
  "ALP" = "#DE3533", "LNP" = "#ADD8E6", "KAP" = "#8B0000", "GVIC" = "#10C25B", "XEN" = "#ff6300",
  "LP" = "#0047AB", "NP" = "#0a9cca", "IND" = "#000000"
)

ggplot(winners_with_map) +
  geom_sf(aes(fill = PartyAb, geometry = geometry)) +
  scale_fill_manual(values = aus_colours)

#UPDATED CODE - previous variable assignments from past weeks
winners_fix = winners_with_map
auscolors = aus_colours
```

Recall 2021 Federal Election Data

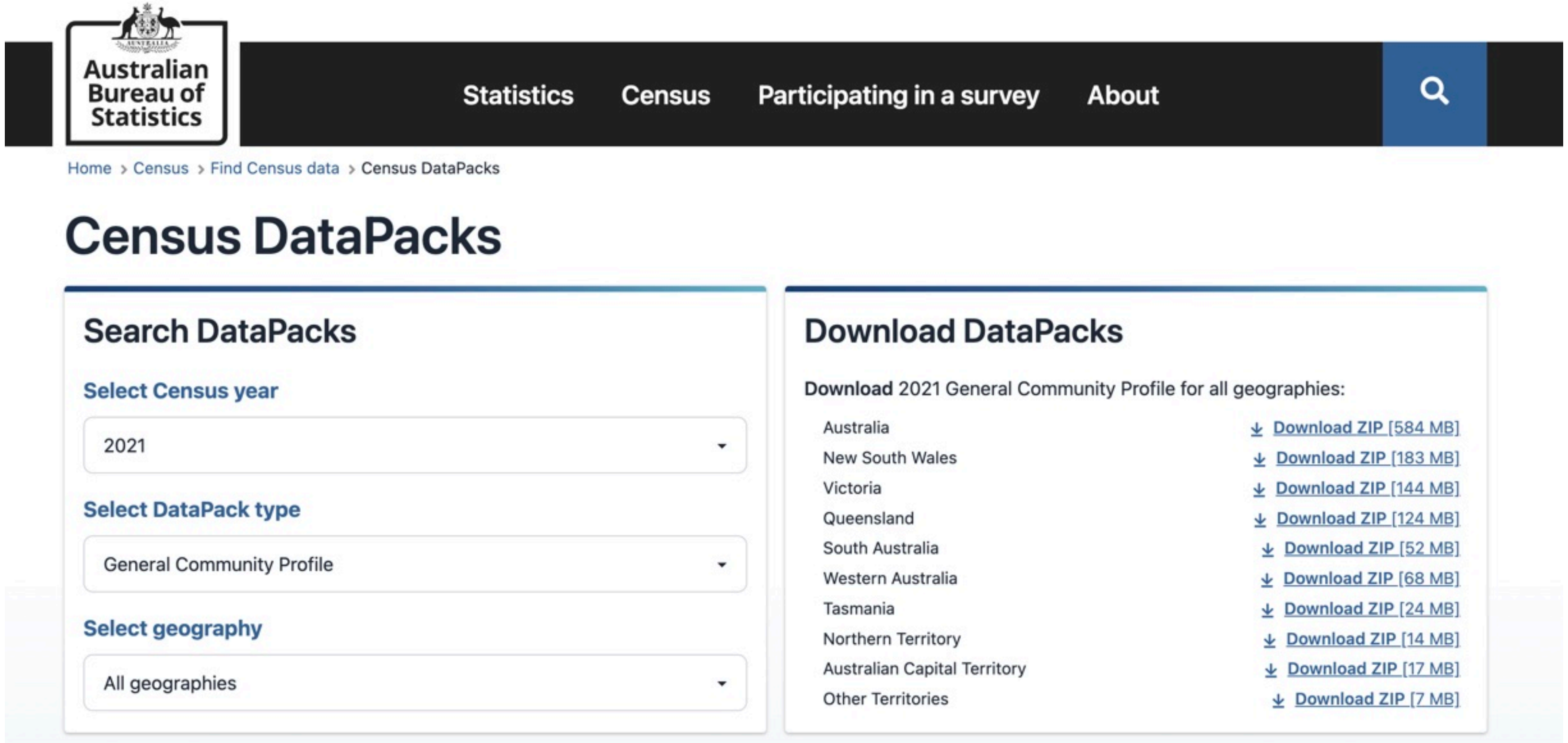


There are two sources of data:

1. Electoral boundary
2. The votes for candidates in each electorate

Recall 2021 ABS Census Data

- DataPacks https://datapacks.censusdata.abs.gov.au/datapacks/
- GeoPackages https://datapacks.censusdata.abs.gov.au/geopackages/



Australian Bureau of Statistics
Statistics
Census
Participating in a survey
About
Home > Census > Find Census data > Census DataPacks

Census DataPacks

Search DataPacks

Select Census year: 2021

Select DataPack type: General Community Profile

Select geography: All geographies

Download DataPacks

Download 2021 General Community Profile for all geographies:

Australia	Download ZIP [584 MB]
New South Wales	Download ZIP [183 MB]
Victoria	Download ZIP [144 MB]
Queensland	Download ZIP [124 MB]
South Australia	Download ZIP [52 MB]
Western Australia	Download ZIP [68 MB]
Tasmania	Download ZIP [24 MB]
Northern Territory	Download ZIP [14 MB]
Australian Capital Territory	Download ZIP [17 MB]
Other Territories	Download ZIP [7 MB]

ABS Census 2021 GeoPackages

GeoPackage

> *A **GeoPackage** (GPKG) is an open, non-proprietary, platform-independent and standards-based data format for geographic information system implemented as a SQLite database container. Defined by the **Open Geospatial Consortium** (OGC) with the backing of the US military and published in 2014, GeoPackage has seen widespread support from various government, commercial, and open source organizations.*
>
> — *Wikipedia*

Recall: OGC also defines the WKT

ABS GeoPackage (2021)

https://datapacks.censusdata.abs.gov.au/geopackages/

1. 2021
2. Victoria
3. Topic: Employment and Income or Table: G17
4. GDA2020

- **Or use the `strayr` package!** We'll use the one from the ABS website instead.

```
# 2022
geopath_2021_G02 <- here::here("data/Geopackage_2021_G02_VIC_GDA2020/G02_VIC_GDA2020.gpkg")
st_layers(geopath_2021_G02)

## Driver: GPKG
## Available layers:
##            layer_name geometry_type features fields crs_name
## 1     G02_UCL_2021_VIC Multi Polygon      347     11  GDA2020
## 2     G02_SUA_2021_VIC Multi Polygon       22     11  GDA2020
## 3     G02_STE_2021_VIC Multi Polygon        1     11  GDA2020
## 4    G02_SOSR_2021_VIC Multi Polygon       12     11  GDA2020
## 5     G02_SOS_2021_VIC Multi Polygon        6     11  GDA2020
```

ABS GeoPackage (2016)

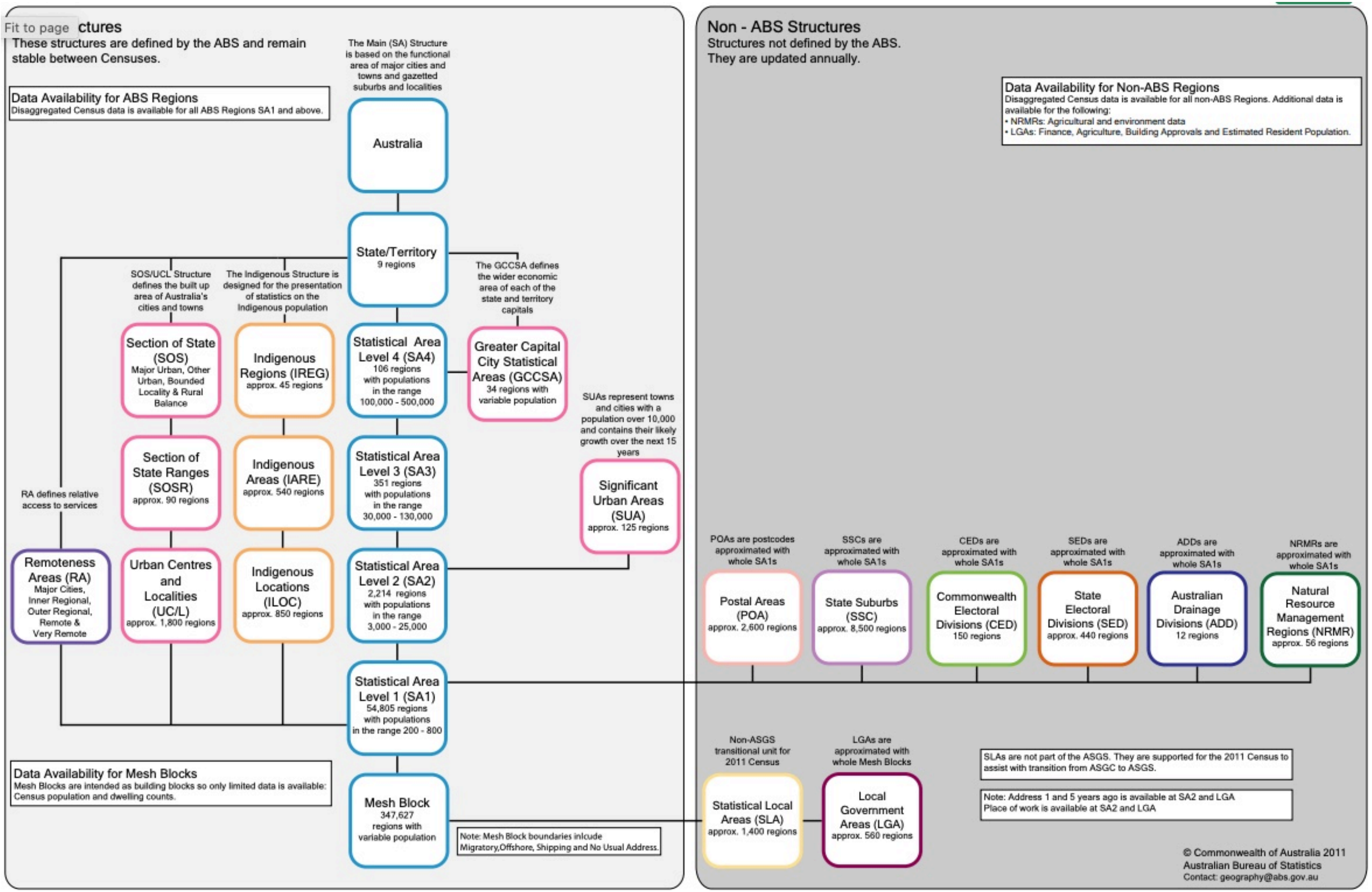
https://datapacks.censusdata.abs.gov.au/geopackages/

1. Victoria
2. Employment, Income and Unpaid Work (EIUW)
3. EIUW GeoPackage A

- **Or use the `strayr` package!** We'll use the one from the ABS website instead.

```
geopath_2016_eiuwa <- here::here("data/Geopackage_2016_EIUWA_for_VIC/census2016_eiuwa_vic_short.gpkg")
st_layers(geopath_2016_eiuwa)

## Driver: GPKG
## Available layers:
##                         layer_name geometry_type features fields crs_name
## 1    census2016_eiuwa_vic_ced_short                     39    489    GDA94
## 2  census2016_eiuwa_vic_gccsa_short                      4    489    GDA94
## 3    census2016_eiuwa_vic_lga_short                     82    489    GDA94
## 4    census2016_eiuwa_vic_poa_short                    698    489    GDA94
## 5     census2016_eiuwa_vic_ra_short                      6    489    GDA94
## 6    census2016_eiuwa_vic_sa1_short                  14073    489    GDA94
## 7    census2016_eiuwa_vic_sa2_short                    464    489    GDA94
```

The Australian Statistical Geography Standard (ASGS)



Fit to page ctures

These structures are defined by the ABS and remain stable between Censuses.

Data Availability for ABS Regions

Disaggregated Census data is available for all ABS Regions SA1 and above.

The Main (SA) Structure is based on the functional area of major cities and towns and gazetted suburbs and localities

Australia

State/Territory 9 regions

SOS/UCL Structure defines the built up area of Australia's cities and towns

The Indigenous Structure is designed for the presentation of statistics on the Indigenous population

The GCCSA defines the wider economic area of each of the state and territory capitals

Section of State (SOS) Major Urban, Other Urban, Bounded Locality & Rural Balance

Indigenous Regions (IREG) approx. 45 regions

Statistical Area Level 4 (SA4) 106 regions with populations in the range 100,000 - 500,000

Greater Capital City Statistical Areas (GCCSA) 34 regions with variable population

SUAs represent towns and cities with a population over 10,000 and contains their likely growth over the next 15 years

Section of State Ranges (SOSR) approx. 90 regions

Indigenous Areas (IARE) approx. 540 regions

Statistical Area Level 3 (SA3) 351 regions with populations in the range 30,000 - 130,000

Significant Urban Areas (SUA) approx. 125 regions

RA defines relative access to services

Remoteness Areas (RA) Major Cities, Inner Regional, Outer Regional, Remote & Very Remote

Urban Centres and Localities (UC/L) approx. 1,800 regions

Indigenous Locations (ILOC) approx. 850 regions

Statistical Area Level 2 (SA2) 2,214 regions with populations in the range 3,000 - 25,000

Statistical Area Level 1 (SA1) 54,805 regions with populations in the range 200 - 800

Data Availability for Mesh Blocks

Mesh Blocks are intended as building blocks so only limited data is available: Census population and dwelling counts.

Mesh Block 347,627 regions with variable population

Note: Mesh Block boundaries inlcude Migratory,Offshore, Shipping and No Usual Address.

Non - ABS Structures

Structures not defined by the ABS. They are updated annually.

Data Availability for Non-ABS Regions

Disaggregated Census data is available for all non-ABS Regions. Additional data is available for the following:

- NRMRs: Agricultural and environment data
- LGAs: Finance, Agriculture, Building Approvals and Estimated Resident Population.

POAs are postcodes approximated with whole SA1s

Postal Areas (POA) approx. 2,600 regions

SSCs are approximated with whole SA1s

State Suburbs (SSC) approx. 8,500 regions

CEDs are approximated with whole SA1s

Commonwealth Electoral Divisions (CED) 150 regions

SEDs are approximated with whole SA1s

State Electoral Divisions (SED) approx. 440 regions

ADDs are approximated with whole SA1s

Australian Drainage Divisions (ADD) 12 regions

NRMRs are approximated with whole SA1s

Natural Resource Management Regions (NRMR) approx. 56 regions

Non-ASGS transitional unit for 2011 Census

Statistical Local Areas (SLA) approx. 1,400 regions

LGAs are approximated with whole Mesh Blocks

Local Government Areas (LGA) approx. 560 regions

SLAs are not part of the ASGS. They are supported for the 2011 Census to assist with transition from ASGC to ASGS.

Note: Address 1 and 5 years ago is available at SA2 and LGA
Place of work is available at SA2 and LGA

© Commonwealth of Australia 2011
Australian Bureau of Statistics
Contact: geography@abs.gov.au

The number of regions for each layer

```
st_layers(geopath_2021_G02) %>%
  # make it into a data.frame first
  tibble(!!!.) %>%
  # then you can the dplyr operations
  dplyr::arrange(features)
```

```
## # A tibble: 16 × 6
##    name                geomtype  driver features fields crs
##    <chr>               <list>    <chr>     <dbl>  <dbl> <list>
##  1 G02_STE_2021_VIC    <chr [1]> GPKG          1     11 <crs>
##  2 G02_GCCSA_2021_VIC <chr [1]> GPKG          4     11 <crs>
##  3 G02_SOS_2021_VIC    <chr [1]> GPKG          6     11 <crs>
##  4 G02_RA_2021_VIC     <chr [1]> GPKG          6     11 <crs>
##  5 G02_SOSR_2021_VIC   <chr [1]> GPKG         12     11 <crs>
##  6 G02_SA4_2021_VIC    <chr [1]> GPKG         19     11 <crs>
##  7 G02_SUA_2021_VIC    <chr [1]> GPKG         22     11 <crs>
##  8 G02_CED_2021_VIC    <chr [1]> GPKG         41     11 <crs>
##  9 G02_SA3_2021_VIC    <chr [1]> GPKG         68     11 <crs>
## 10 G02_LGA_2021_VIC    <chr [1]> GPKG         82     11 <crs>
## 11 G02_SED_2021_VIC    <chr [1]> GPKG         90     11 <crs>
## 12 G02_UCL_2021_VIC    <chr [1]> GPKG        347     11 <crs>
## 13 G02_SA2_2021_VIC    <chr [1]> GPKG        524     11 <crs>
## 14 G02_POA_2021_VIC    <chr [1]> GPKG        694     11 <crs>
## 15 G02_SAL_2021_VIC    <chr [1]> GPKG       2946     11 <crs>
## 16 G02_SA1_2021_VIC    <chr [1]> GPKG      15482     11 <crs>
```

Data in the layer

```r
vicmap_ste_G02 <- read_sf(geopath_2021_G02, layer = "G02_STE_2021_VIC")
vicmap_ste_G02$geom
```

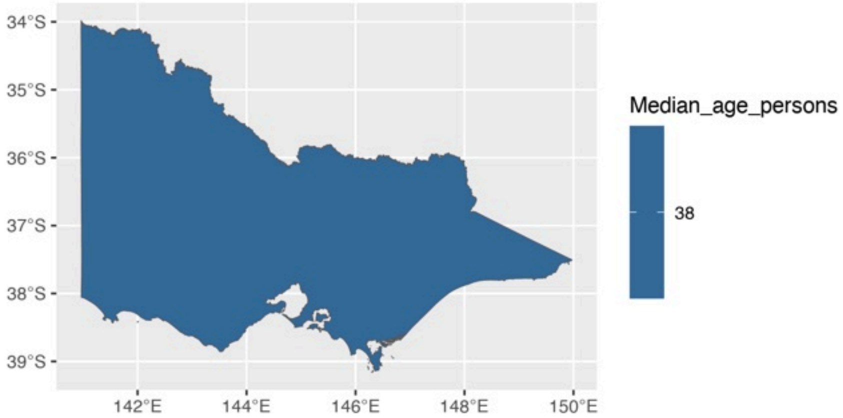
```
## Geometry set for 1 feature
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: 140.9619 ymin: -39.15918 xmax: 149.9762 ymax: -33.98064
## Geodetic CRS:  GDA2020
```

```r
str(vicmap_ste_G02)
```

```
## sf [1 × 12] (S3: sf/tbl_df/tbl/data.frame)
##  $ STE_CODE_2021                : chr "2"
##  $ STE_NAME_2021                : chr "Victoria"
##  $ Median_age_persons           : num 38
##  $ Median_mortgage_repay_monthly: num 1859
##  $ Median_tot_prsnl_inc_weekly  : num 803
##  $ Median_rent_weekly           : num 370
##  $ Median_tot_fam_inc_weekly    : num 2136
##  $ Average_num_psns_per_bedroom : num 0.8
##  $ Median_tot_hhd_inc_weekly    : num 1759
##  $ Average_household_size       : num 2.5
##  $ AREA_ALBERS_SQKM             : num 227496
##  $ geom                         :sfc_MULTIPOLYGON of length 1; first list element: List of 157
##   ..$ :List of 1
##   .. ..$ : num [1:39, 1:2] 146 146 146 146 146 ...
```

State or Territory (STE)

```
vicmap_ste_G02 <- read_sf(geopath_2021_G02, layer = "G02_STE_2021_VIC")

ggplot(vicmap_ste_G02) +
  geom_sf(aes(geometry = geom, fill = Median_age_persons))
```



```
nrow(vicmap_ste)

## [1] 1
```

Breakout Session

Try it yourself time:

- Download the geopackage for the 2021 census and variable G02
- Look at how that spatial data is organised
- Look at the different layers

What are the differences between the regionalisations?

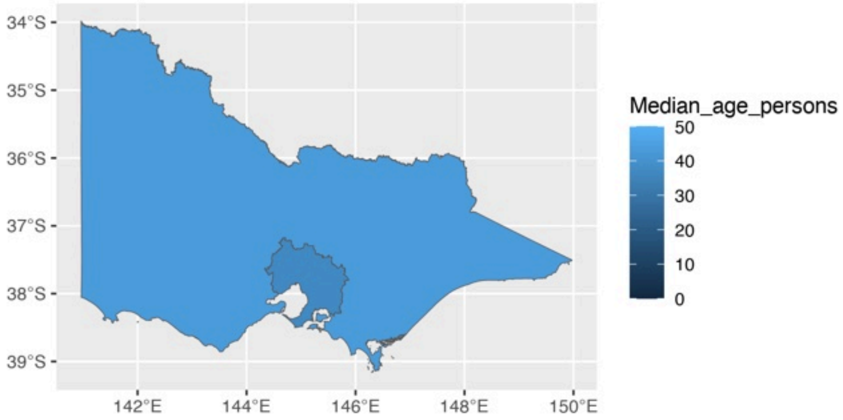
What do you notice?

Want a challenge visualise the 2016 data!

Greater Capital City Statistical Areas (GCCSA)

- Each region with variable population

```
vicmap_gccsa_G02 <- read_sf(geopath_2021_G02, layer = "G02_GCCSA_2021_VIC")
ggplot(vicmap_gccsa_G02) +
  geom_sf(aes(geometry = geom, fill = Median_age_persons))
```

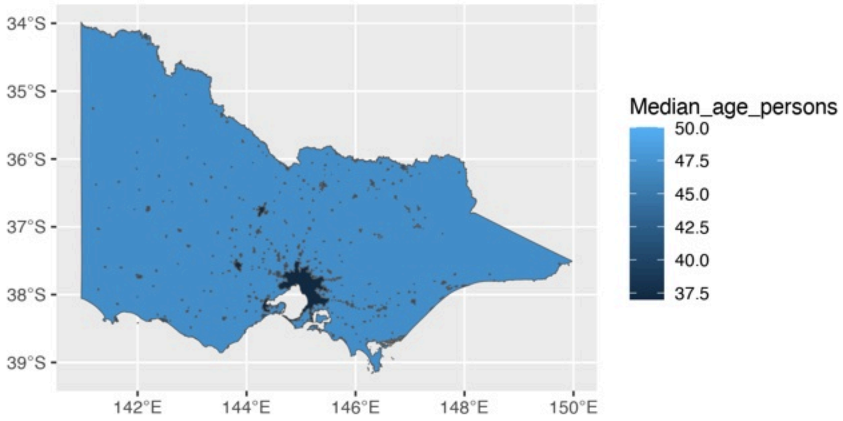


```
nrow(vicmap_gccsa_G02)

## [1] 4
```

Section of State (SOS)

- Major urban, other urban, bounded locally & rural balance

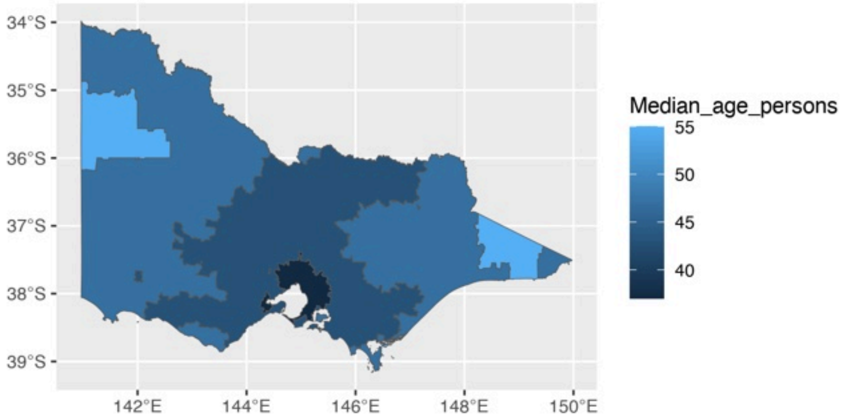
```
vicmap_sos_G02 <- read_sf(geopath_2021_G02, layer = "G02_SOS_2021_VIC")
ggplot(vicmap_sos_G02) +
  geom_sf(aes(geometry = geom, fill = Median_age_persons))
```



```
nrow(vicmap_sos_G02)

## [1] 6
```

Remoteness Areas (RA)

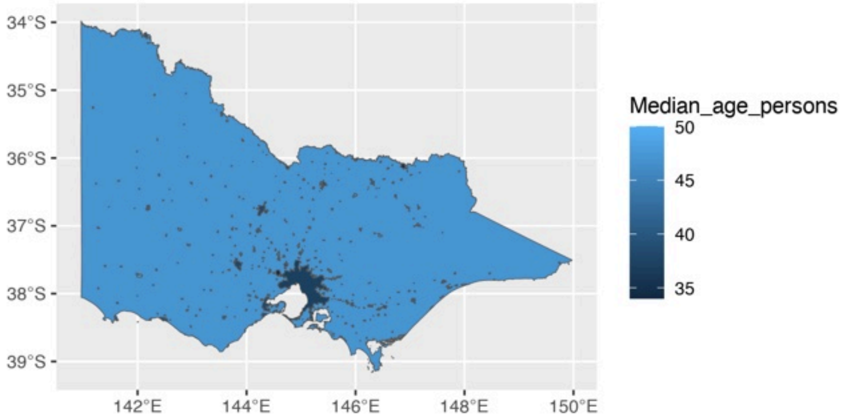
```r
vicmap_ra_G02 <- read_sf(geopath_2021_G02, layer = "G02_RA_2021_VIC")
ggplot(vicmap_ra_G02) +
  geom_sf(aes(geometry = geom, fill = Median_age_persons))
```



```r
nrow(vicmap_ra_G02)

## [1] 6
```

Section of State Ranges (SOSR)

```
vicmap_sosr_G02 <- read_sf(geopath_2021_G02, layer = "G02_SOSR_2021_VIC")
ggplot(vicmap_sosr_G02) +
  geom_sf(aes(geometry = geom, fill = Median_age_persons))
```

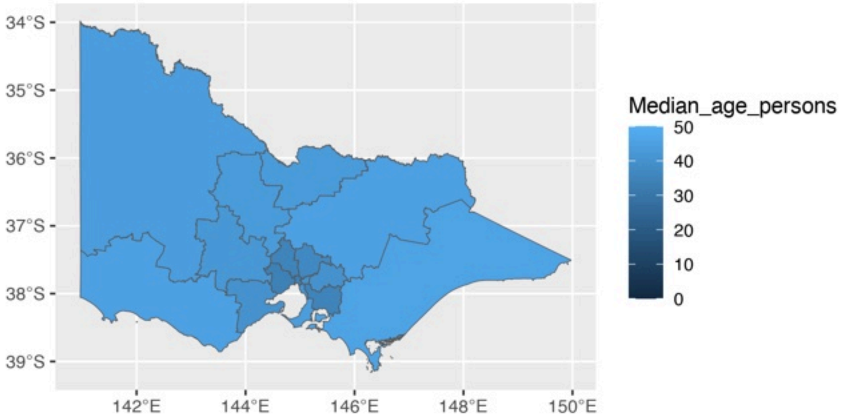


```
nrow(vicmap_sosr_G02)

## [1] 12
```

Statistical Area Level 4 (SA4)

- Each region with population of 100,000 - 500,000

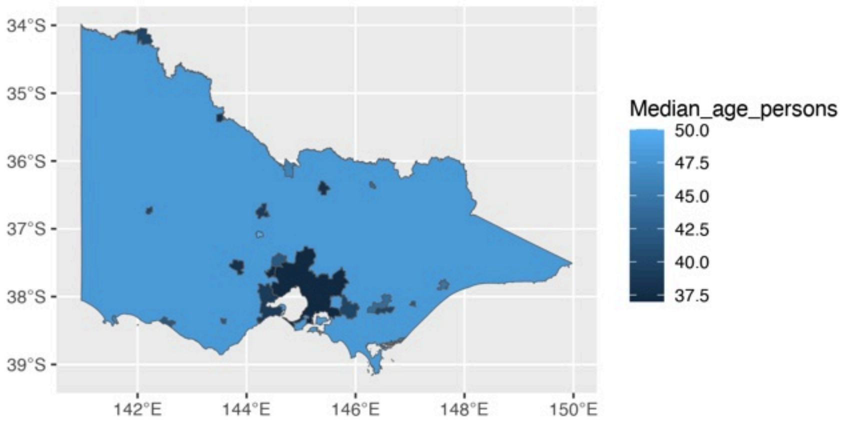
```
vicmap_sa4_G02 <- read_sf(geopath_2021_G02, layer = "G02_SA4_2021_VIC")
ggplot(vicmap_sa4_G02) +
  geom_sf(aes(geometry = geom, fill = Median_age_persons))
```



```
nrow(vicmap_sa4_G02)

## [1] 19
```

Significant Urban Areas (SUA)

```
vicmap_sua_G02 <- read_sf(geopath_2021_G02, layer = "G02_SUA_2021_VIC")
ggplot(vicmap_sua_G02) +
  geom_sf(aes(geometry = geom, fill = Median_age_persons))
```

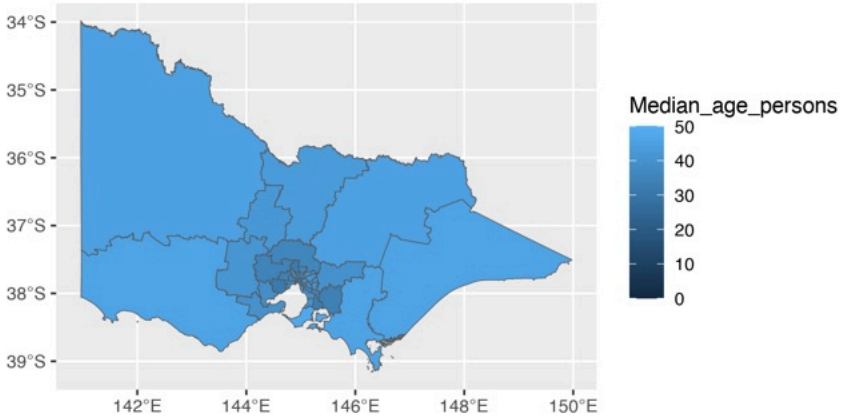


```
nrow(vicmap_sua_G02)
```

```
## [1] 22
```

Commonwealth Electoral Division (CED)

```
vicmap_ced_G02 <- read_sf(geopath_2021_G02, layer = "G02_CED_2021_VIC")
ggplot(vicmap_ced_G02) +
  geom_sf(aes(geometry = geom, fill = Median_age_persons))
```

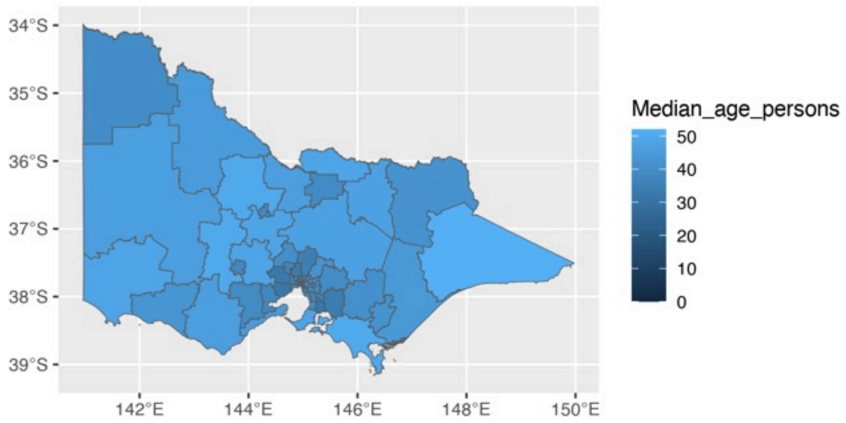


```
nrow(vicmap_ced_G02)

## [1] 41
```

Statistical Area Level 3 (SA3)

- Each region with population of 30,000 - 130,000

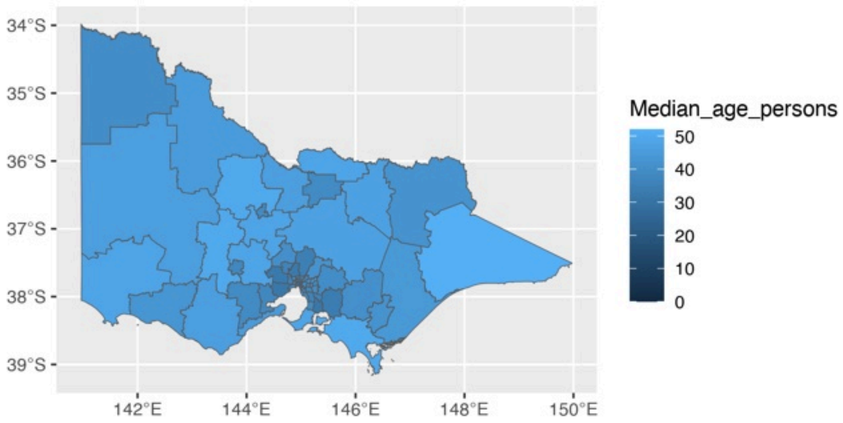
```
vicmap_sa3_G02 <- read_sf(geopath_2021_G02, layer = "G02_SA3_2021_VIC")
ggplot(vicmap_sa3_G02) +
  geom_sf(aes(geometry = geom, fill = Median_age_persons))
```



```
nrow(vicmap_sa3_G02)

## [1] 68
```

Local Government Area (LGA)

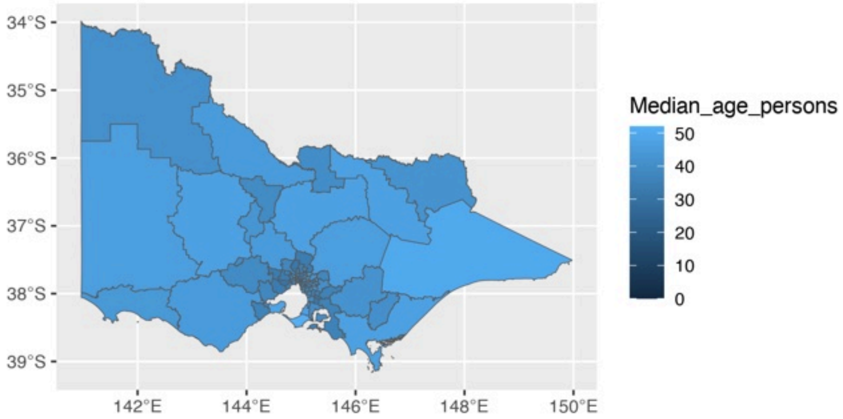
```
vicmap_lga_G02 <- read_sf(geopath_2021_G02, layer = "G02_SA3_2021_VIC")
ggplot(vicmap_lga_G02) +
  geom_sf(aes(geometry = geom, fill = Median_age_persons))
```



```
nrow(vicmap_lga_G02)

## [1] 68
```

State Electoral Division (SED)

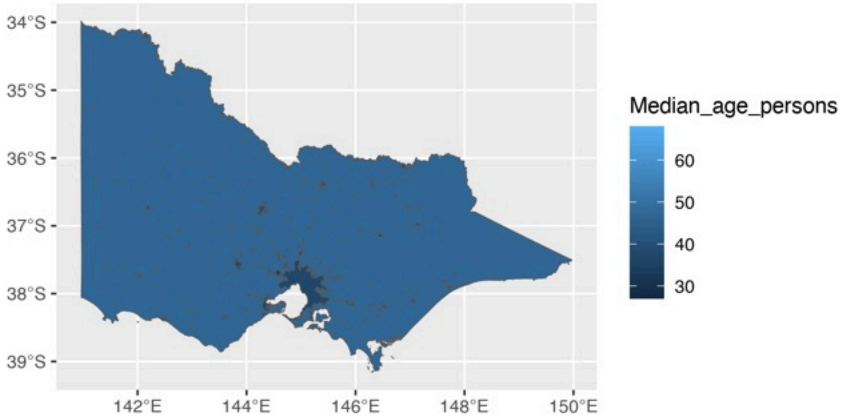
```r
vicmap_sed_G02 <- read_sf(geopath_2021_G02, layer = "G02_SED_2021_VIC")
ggplot(vicmap_sed_G02) +
  geom_sf(aes(geometry = geom, fill = Median_age_persons))
```



```r
nrow(vicmap_sed_G02)

## [1] 90
```

Urban Centres and Localities (UCL)

```
vicmap_ucl_G02 <- read_sf(geopath_2021_G02, layer = "G02_UCL_2021_VIC")
ggplot(vicmap_ucl_G02) +
  geom_sf(aes(geometry = geom, fill = Median_age_persons))
```

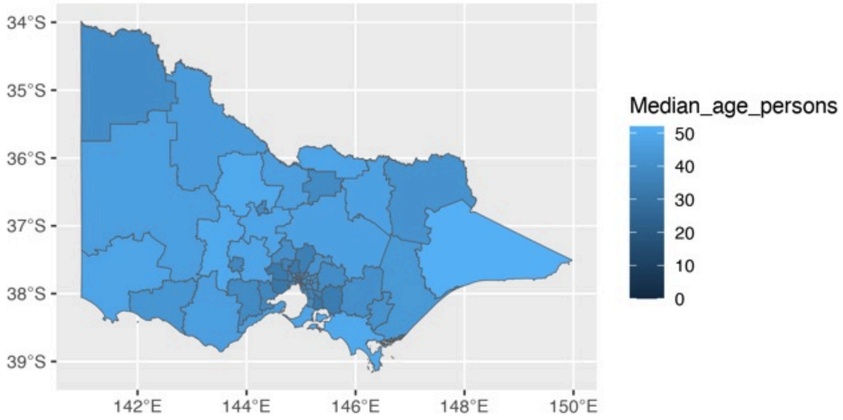


```
nrow(vicmap_ucl_G02)

## [1] 347
```

Statistical Area Level 2 (SA2)

- Each region with populations in the range of 3,000-25,000

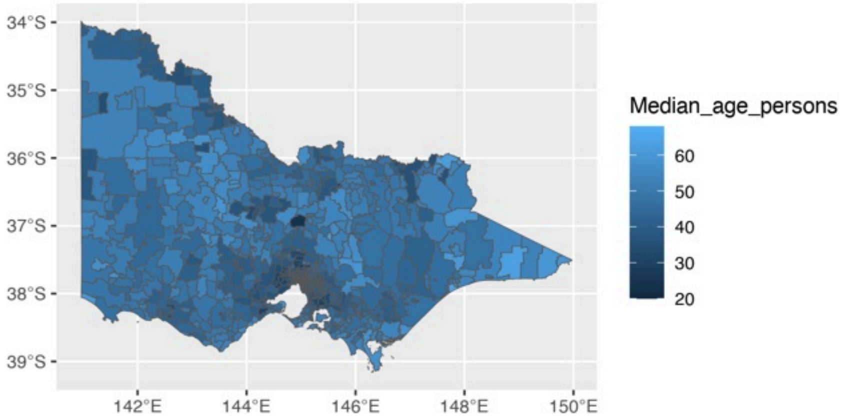
```
vicmap_sa2_G02 <- read_sf(geopath_2021_G02, layer = "G02_SA3_2021_VIC")
ggplot(vicmap_sa2_G02) +
  geom_sf(aes(geometry = geom, fill = Median_age_persons))
```



```
nrow(vicmap_sa2_G02)

## [1] 68
```

Postal Areas (POA)

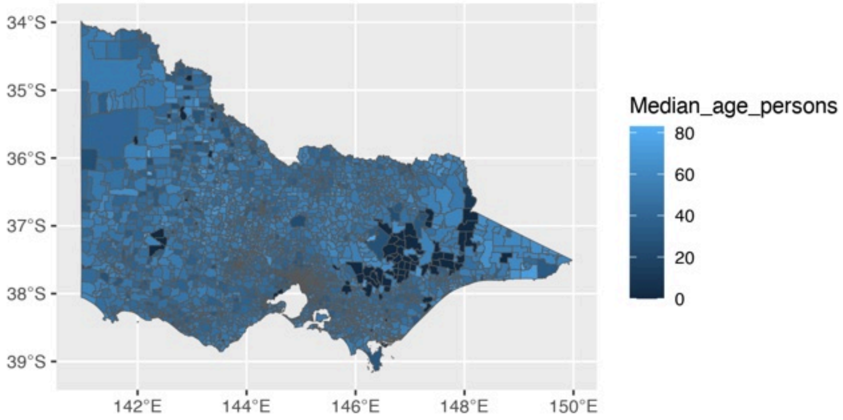
```
vicmap_poa_G02 <- read_sf(geopath_2021_G02, layer = "G02_POA_2021_VIC")
ggplot(vicmap_poa_G02) +
  geom_sf(aes(geometry = geom, fill = Median_age_persons))
```



```
nrow(vicmap_poa_G02)

## [1] 694
```

State Area Localitites (SAL) (Formerly SSC)

```
vicmap_sal_G02 <- read_sf(geopath_2021_G02, layer = "G02_SAL_2021_VIC")
ggplot(vicmap_sal_G02) +
  geom_sf(aes(geometry = geom, fill = Median_age_persons))
```



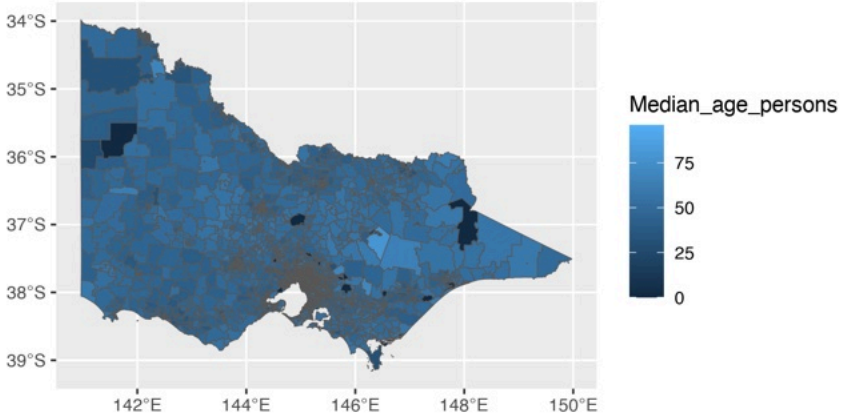
```
nrow(vicmap_sal_G02)
```

```
## [1] 2946
```

Statistical Area Level 1 (SA1)

- Each region with a population of range 200-800

```
vicmap_sa1_G02 <- read_sf(geopath_2021_G02, layer = "G02_SA1_2021_VIC")
ggplot(vicmap_sa1_G02) +
  geom_sf(aes(geometry = geom, fill = Median_age_persons))
```



```
nrow(vicmap_sa1_G02)

## [1] 15482
```

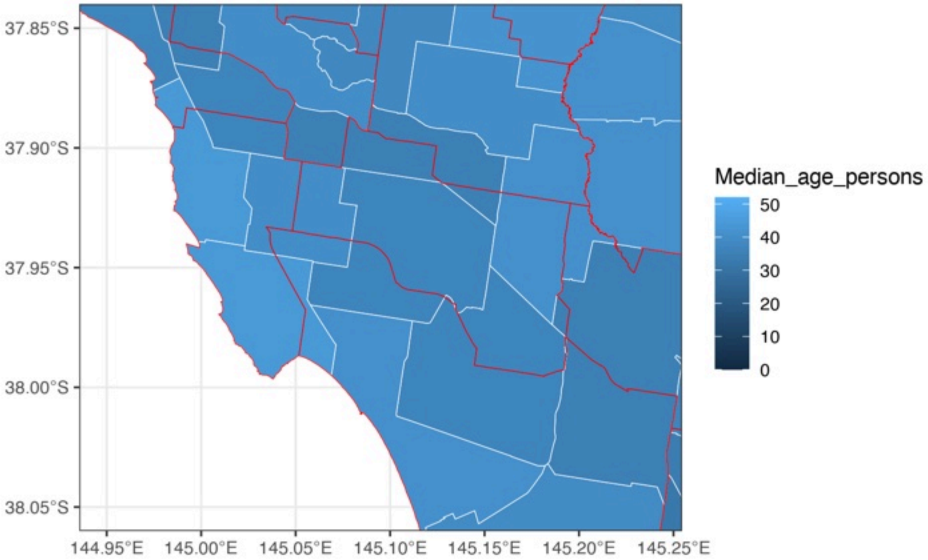
Electorate boundary
vs
Census boundary

🎯 Estimate a median age for an electorate

Comparing SED 2021 and electorates divisions 2022

See here for `electoral_winners` data was.

```
ggplot() +
  geom_sf(data = vicmap_sed_G02,
          aes(geometry = geom, fill = Median_age_persons),
    alpha = 1, color = "white", size = 2) +
  geom_sf(data = winners_with_map, aes(geometry = geometry),
    fill = "transparent", color = "red", size = 2) +
  coord_sf(xlim = c(144.95, 145.24), ylim = c(-38.05, -37.85)) +
  theme_bw()
```



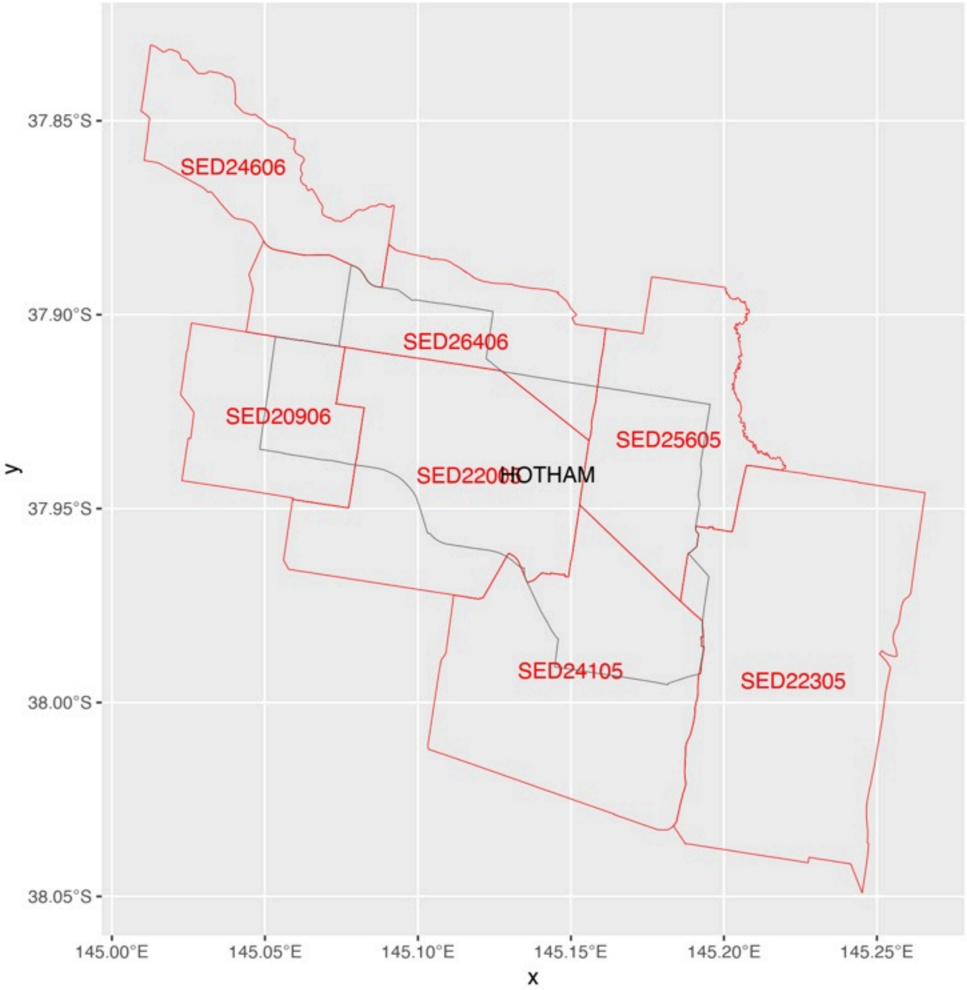
Closer look Hotham electorate ❶

```
electorate <- winners_with_map |>
  filter(DivisionNm == "HOTHAM")

# Set projection to GDA1994 using EPSG:4283
st_crs(electorate$geometry,4283)

# Transform projection from GDA1994 to GDA2020 using EPSG:7844
electorate$geometry = st_transform(electorate$geometry, 7844)

sed_intersect <- vicmap_sed_G02 |>
  filter(st_intersects(geom,
    electorate$geometry,
    sparse = FALSE
  )[, 1])

ggplot() +
  geom_sf(data = sed_intersect,
    aes(geometry = geom), color = "red", fill = "transparent") +
  geom_sf_text(data = sed_intersect,
    aes(label = SED_CODE_2021, geometry = geom), color = "red") +
  geom_sf(data = electorate, aes(geometry = geometry), fill = "transparent") +
  geom_sf_text(data = electorate, aes(geometry = geometry, label = DivisionNm))
```

```
## Coordinate Reference System:
##   User input: GDA94
##   wkt:
```

Closer look Hotham electorate ❶

There are 7 SED regions that intersect with Hotham electorate.

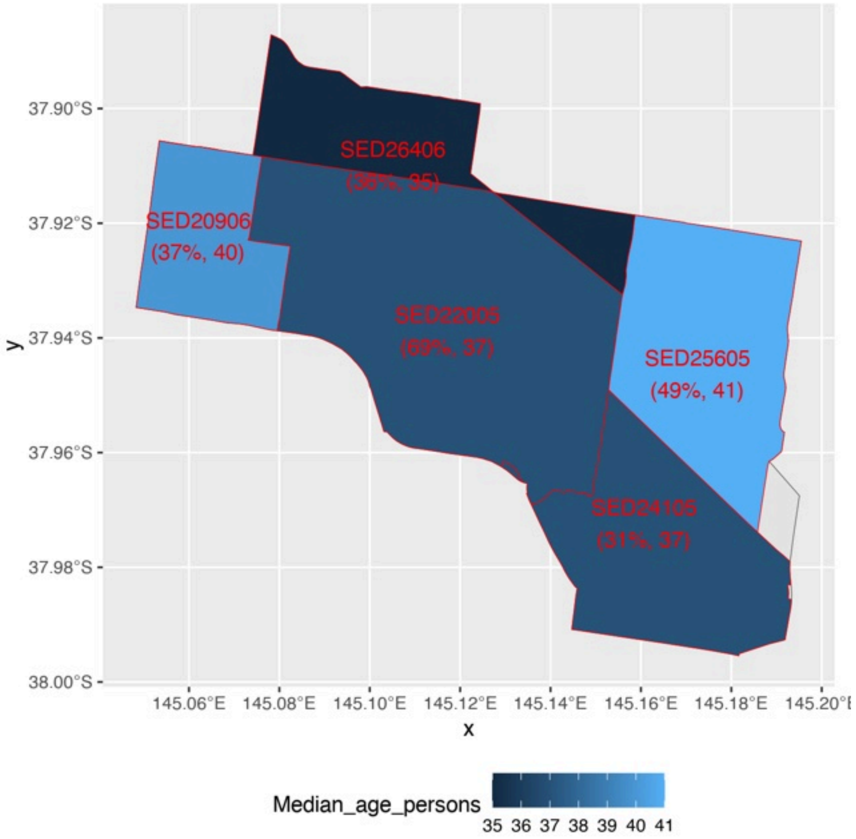


Closer look   Hotham electorate 2

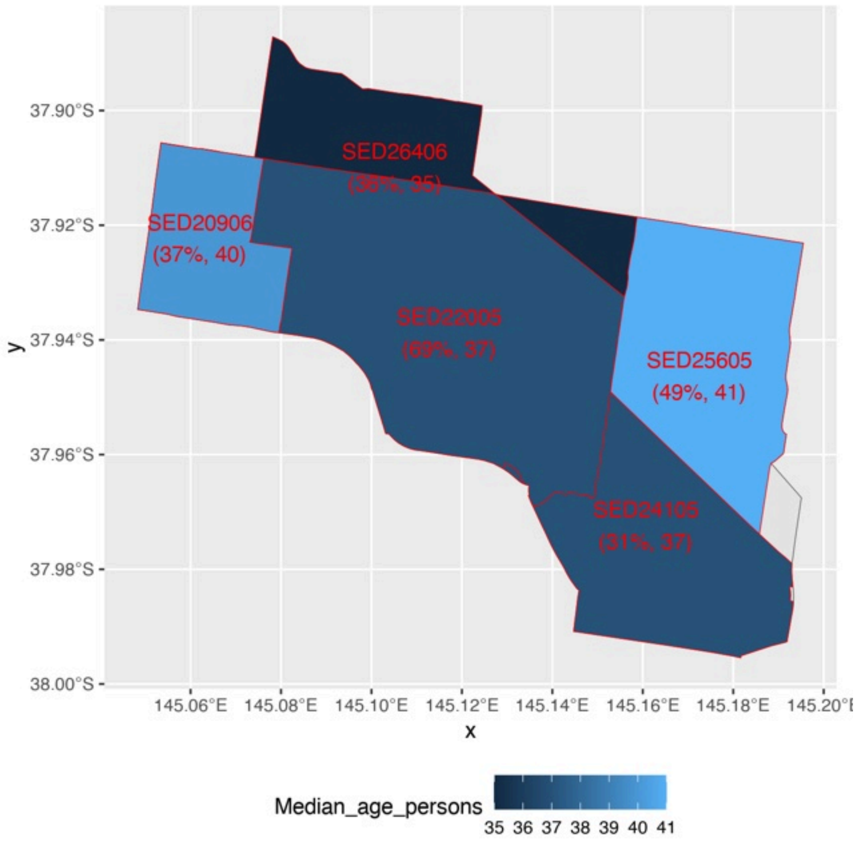
```
sed_intersect2 <- sed_intersect |>
  mutate(
    geometry = st_intersection(geom, electorate$geometry),
    perc_area = 100 * st_area(geometry) / st_area(geom),
    perc_area = as.numeric(perc_area)
  ) |>
  filter(perc_area > 5)

ggplot(sed_intersect2, aes(geometry = geometry)) +
  geom_sf(data = electorate) +
  geom_sf_text(
    data = electorate,
    aes(label = DivisionNm)
  ) +
  geom_sf(color = "red", aes(fill = Median_age_persons)) +
  geom_sf_text(
    aes(
      label = glue::glue("{SED_CODE_2021}
                        ({scales::comma(perc_area, 1)}%, {Median_age_persons})")
    ),
    color = "red"
  ) +
  theme(legend.position = "bottom")
```

Closer look Hotham electorate ❷

- There are 5 SED areas with at least 5% intersection with the electoral area.
- **How would you characterise the median age for Hotham?**



Closer look Hotham electorate ❸



Strategy 1

```
sort(sed_intersect2$Median_age_persons)

## [1] 35 37 37 40 41
```

Strategy 2

```
mean(sed_intersect2$Median_age_persons)

## [1] 38
```

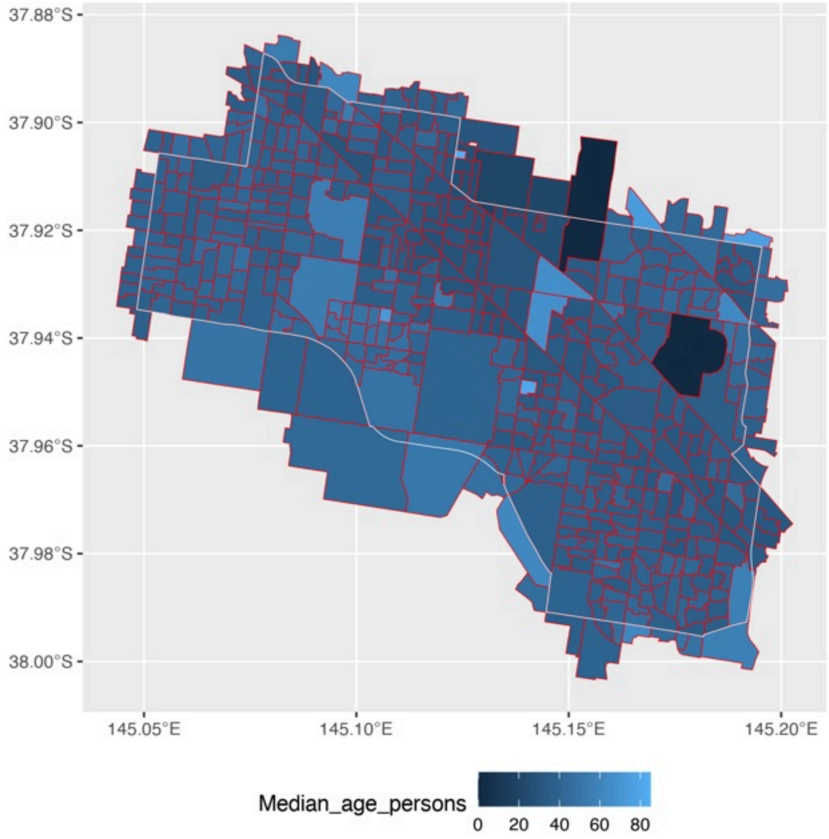
Strategy 3

```
weighted.mean(
  sed_intersect2$Median_age_persons,
  sed_intersect2$perc_area
)

## [1] 38.06118
```

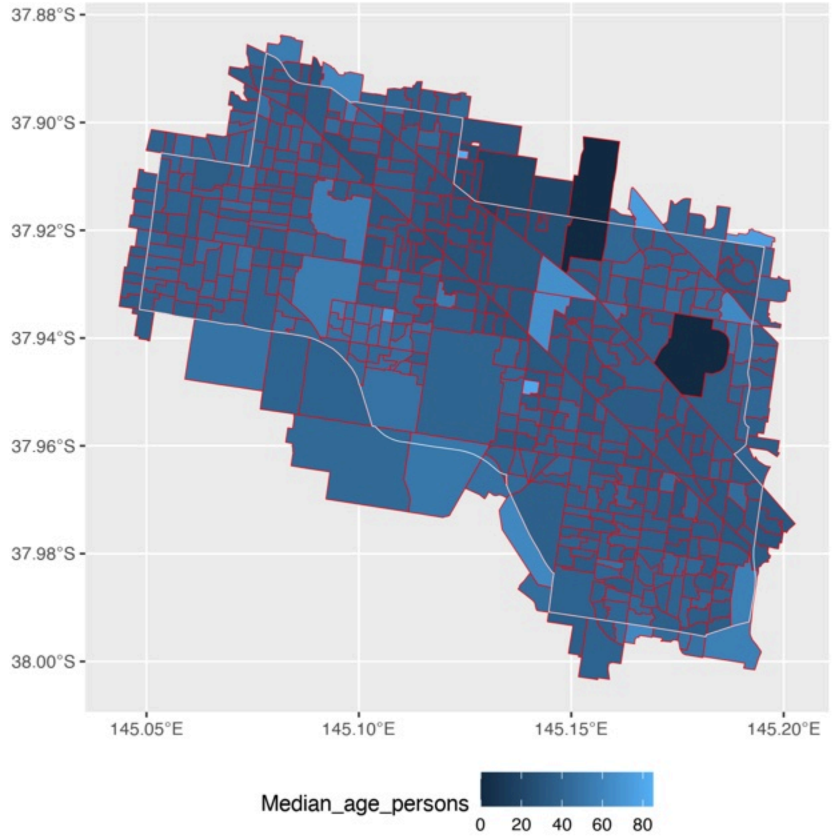
Closer look

Hotham electorate 4

```r
sa1_intersect <- vicmap_sa1_G02 %>%
  filter(st_intersects(geom,
    electorate$geometry,
    sparse = FALSE
  )[, 1])

sa1_intersect2 <- sa1_intersect %>%
  mutate(
    geometry = st_intersection(geom, electorate$geometry),
    perc_area = 100 * st_area(geometry) / st_area(geom),
    perc_area = as.numeric(perc_area)
  ) %>%
  filter(perc_area > 5)

ggplot(sa1_intersect) +
  geom_sf(color = "red",
          aes( fill = Median_age_persons, geometry = geom)) +
  geom_sf(data = electorate,
          color = "white", size = 2, fill = "transparent",
          aes(geometry = geometry)) +
  theme(legend.position = "bottom")
```



Closer look   Hotham electorate 5



Strategy 1

```
fivenum(sa1_intersect2$Median_age_persons)

## [1]  0 35 38 41 81
```

Strategy 2

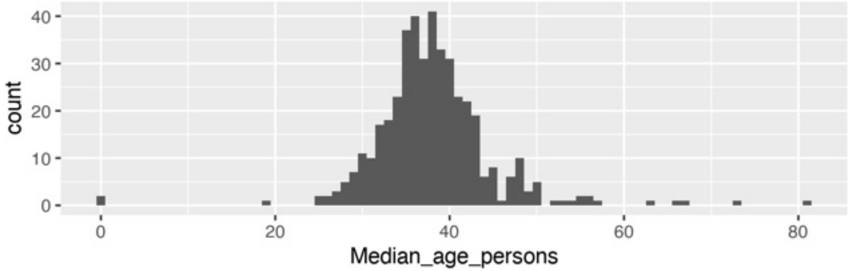
```
mean(sa1_intersect2$Median_age_persons)

## [1] 38.07674
```

Strategy 3

```
weighted.mean(sa1_intersect2$Median_age_persons, sa1_intersect2$perc_are

## [1] 38.02792
```

Strategy 4

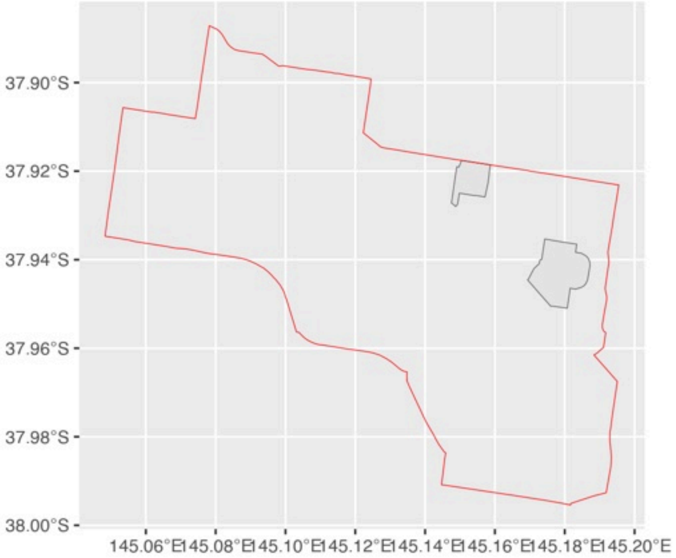
```
ggplot(sa1_intersect2, aes(x = Median_age_persons)) +
  geom_histogram(binwidth = 1)
```

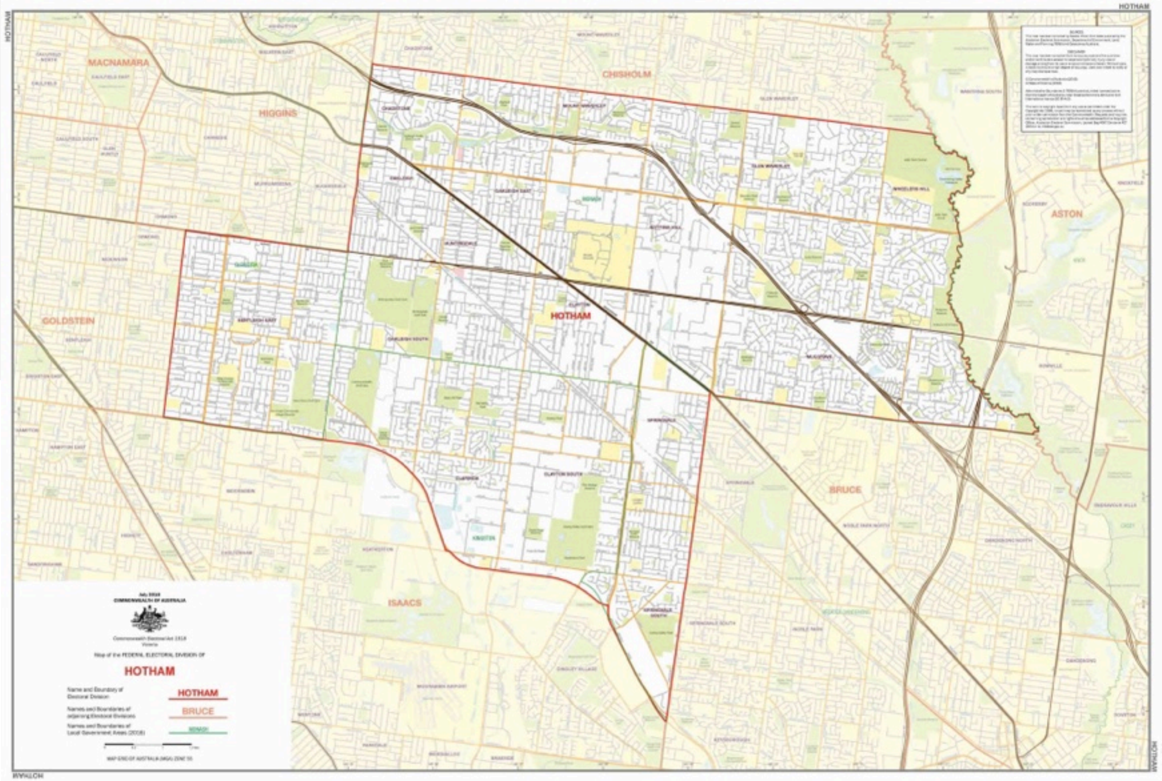


Closer look Zero median age

(Hotham 2022)[https://www.aec.gov.au/profiles/vic/hotham.htm]

```r
sa1_intersect2 %>%
  filter(Median_age_persons == 0) %>%
  ggplot() +
  geom_sf() +
  geom_sf(
    data = electorate, color = "red",
    fill = "transparent",
    aes(geometry = geometry)
  )
```





Closer look Hotham electorate 6

Before

Strategy 1

```
fivenum(sa1_intersect2$Median_age_persons)

## [1]  0 35 38 41 81
```

Strategy 2

```
mean(sa1_intersect2$Median_age_persons)

## [1] 38.07674
```

Strategy 3

```
weighted.mean(sa1_intersect2$Median_age_persons, sa1_interse

## [1] 38.02792
```

After

```
sa1_intersect3 <- sa1_intersect2 %>%
  filter(Median_age_persons != 0)
```

Strategy 1

```
fivenum(sa1_intersect3$Median_age_persons)

## [1] 19 35 38 41 81
```

Strategy 2

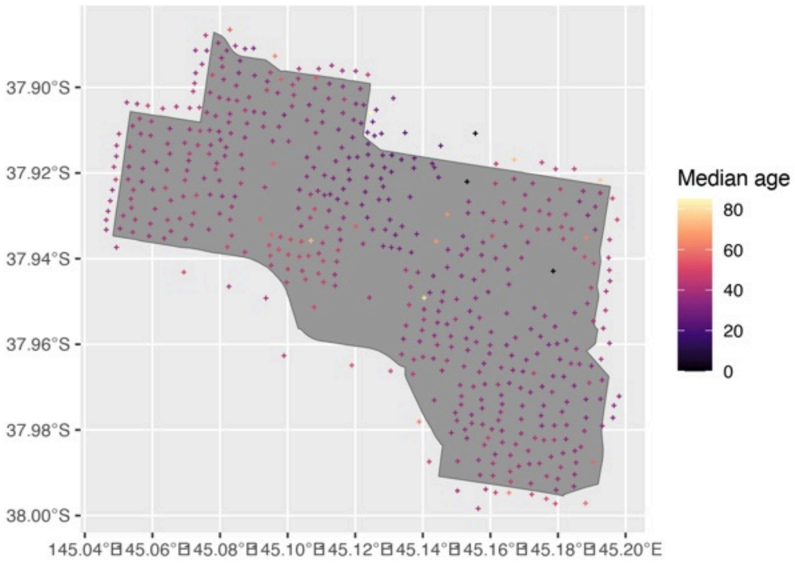
```
mean(sa1_intersect3$Median_age_persons)

## [1] 38.25467
```

Strategy 3

```
weighted.mean(sa1_intersect3$Median_age_persons, sa1_inter

## [1] 38.20661
```

Dorling Cartogram

```
sa1_intersect4 <- sa1_intersect %>%
  mutate(centroid = st_centroid(geom))

dorling_plot <- ggplot(sa1_intersect4) +
  geom_sf(
    data = electorate,
    aes(geometry = geometry), size = 4, fill = "grey60"
  ) +
  geom_sf(aes(geometry = centroid, color = Median_age_persons),
    size = 0.5, shape = 3
  ) +
  scale_color_viridis_c(name = "Median age", option = "magma")
```



Closer look   Hotham electorate 7

```
sa1_intersect5 <- sa1_intersect4 %>%
  filter(st_intersects(centroid, electorate$geometry, sparse = FALSE)[, 1],
    Median_age_persons != 0)
```

Strategy 1

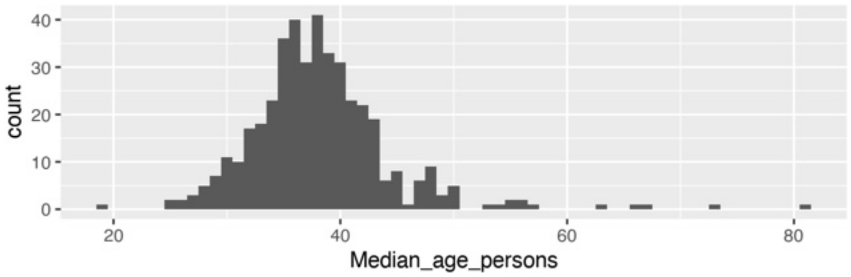
```
fivenum(sa1_intersect5$Median_age_persons)

## [1] 19 35 38 41 81
```

Strategy 2

```
mean(sa1_intersect5$Median_age_persons)

## [1] 38.20706
```

Strategy 4

```
ggplot(sa1_intersect5, aes(x = Median_age_persons)) +
  geom_histogram(binwidth = 1)
```



Summary

- There are many ways to characterise an electorate.
- Estimates of median age of an electorate is more consistent using SA1 map data than SED map data.

- We looked at mapping the 2021 census boundaries and projected a summary of the census variable (i.e. median age) onto a 2022 electoral district
- Discovered some of the challenges with matching two different types of data

Read Forbes, Cook & Hyndman (2020) Spatial modelling of the two-party preferred vote in Australian federal elections: 2001–2016. *Australian & New Zealand Journal of Statisitcs*. for a more sophisticated approach to studying the census variables and election results together.

Slides updated and maintained by Dr. Kate Saunders. Slides originally developed by Dr. Emi Tanaka

Lecturer: *Kate Saunders*

Department of Econometrics and Business Statistics

✉ ETC5512.Clayton-x@monash.edu

Week 6