Census 2016

This package contains selected totals from the Community Profiles time series data, released by the Australian Bureau of Statistics as part of the Census 2016 release.

```r
library(knitr)
pkgs_suggested <- c("magrittr",
                     "ggplot2",
                     "scales",
                     "ggrepel")
suggested_packages_usable <-
  all(vapply(pkgs_suggested, requireNamespace, logical(1), quietly = TRUE))

knitr::opts_chunk
  set(eval = suggested_packages_usable,
       # dev = "png",
       fig.width = 8,
       fig.height = 6)
```

```r
library(magrittr)
library(data.table)
library(Census2016)
```

```r
## For additional data packages for the 2016 Census, visit https://github.com/HughParsonage/Census2016.DataPack
library(ggplot2)
library(scales)
library(ggrepel)
```

```r
data.kable <- function(DT) {
current_knitr.kable.NA <- options("knitr.kable.NA")
options(knitr.kable.NA = '...')
on.exit(options(knitr.kable.NA = current_knitr.kable.NA))
if (nrow(DT) > 50) {
middle_row <- as.data.table(matrix(nrow = 1, ncol = ncol(DT)))
setnames(middle_row, seq_along(middle_row), names(DT))
DT_topn <- rbind(head(DT),
                middle_row,
                tail(DT))
kable(DT_topn, format.args = list(big.mark = ","))
} else {
kable(DT, format.args = list(big.mark = ","))
}
}
```

There is one function `see_question()` and 6 data sets.
This is a simple data.table of multiple variables for each statistical area 2 (SA2)-census year combination. The columns are ordered roughly by the order of the questions on the Census form. Not all values are available to satisfy CRAN’s limits on package size.

Both the sa2_code and sa2_name are provided for convenience.

Median mortgage vs median income

```r
Census2016_wide_by_SA2_year %>%
  .[year == 2016] %>%
  .[, .(sa2_name, persons, median_household_income, median_annual_mortgage)] %>%
  .[order(median_annual_mortgage)] %>%
data.kable

<table>
<thead>
<tr>
<th>sa2_name</th>
<th>persons</th>
<th>median_household_income</th>
<th>median_annual_mortgage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deua - Wadbilliga</td>
<td>25</td>
<td>44,148</td>
<td>0</td>
</tr>
<tr>
<td>Port Kembla Industrial</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Illawarra Catchment Reserve</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Prospect Reservoir</td>
<td>40</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Banksmeadow</td>
<td>18</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Port Botany Industrial</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Cottesloe</td>
<td>7,375</td>
<td>138,788</td>
<td>39,000</td>
</tr>
<tr>
<td>Lilli Pilli - Port Hacking - Dolans Bay</td>
<td>3,148</td>
<td>155,220</td>
<td>39,300</td>
</tr>
<tr>
<td>Nedlands - Dalkeith - Crawley</td>
<td>18,534</td>
<td>118,976</td>
<td>39,600</td>
</tr>
<tr>
<td>Rose Bay - Vaucluse - Watsons Bay</td>
<td>11,840</td>
<td>150,228</td>
<td>41,604</td>
</tr>
<tr>
<td>Balgowlah - Clontarf - Seaforth</td>
<td>20,186</td>
<td>145,496</td>
<td>41,604</td>
</tr>
<tr>
<td>Hunters Hill - Woolwich</td>
<td>10,345</td>
<td>142,064</td>
<td>42,000</td>
</tr>
</tbody>
</table>
```

Census2016_wide_by_SA2_year %>%
  .[year == 2016] %>%
  .[, .(sa2_name, persons, median_household_income, median_annual_mortgage)] %>%
  .[median_annual_mortgage > 0] %>%
  .[order(mortgage_less_income := median_annual_mortgage - median_household_income)] %>%
  .[text := NA_character_] %>%
  .[color := "black"] %>%
  .[mortgage_less_income] %>%
  .[N:1 <= 5, text := sa2_name] %>%
  .[N:1 <= 5, color := "red"] %>%
  .[1:.N <= 5, text := sa2_name] %>%
  .[1:.N <= 5, color := "blue"] %>%
ggplot(aes(x = median_household_income,
          y = median_annual_mortgage,
          size = persons,
          alpha = persons,
          color = color)) +
ggplot1d() +
scale_color_identity() +
scale_size(labels = comma) +
scale_alpha_continuous(labels = comma,
Changes from previous years

In addition to the 2016 data, the package also includes 2006 and 2011 census data as part of the time series. The ABS has released these data series to be comparable; even though the SA2 boundaries have changed you may assume that they refer to the same geographic area.

We can see that Mandarin has become much more common

```r
languages_spoken_by_year <- Census2016_languages %>%
  .[, .(persons = sum(persons)), keyby = .(language, year)] %>%
  setorder(-year, -persons) %>%
  .[]
```
languages_spoken_by_year %>%
  # Examine the top six languages,
  # leave the others unlabelled and grey
  .[language %in% languages_spoken_by_year$language[1:6],
    Language := language] %>%
  .[year == 2016, text := Language] %>%
  .[, Language := reorder(Language, -persons)] %>%
  ggplot(aes(x = year,
    y = persons,
    group = language,
    color = Language,
    label = text)) +
  geom_line() +
  scale_y_continuous(label = comma) +
  geom_text_repel(na.rm = TRUE,
    fontface = "bold",
    force = 1.5,
    nudge_x = 0.5)

Although Census2016 is intended as a data-only package, there is one function, see_question.
It is frequently useful to view the actual question that was asked when looking at survey data. `see_question` provides a convenient way to do this without leaving RStudio (or even your keyboard). There are two methods: `see_question.numeric` takes a question number and prints it.

```r
see_question(3)
```

3. Is the person male or female?
   - Mark one box for each person, like this: —
   - Male
   - Female

The other method is dispatched when one of the two-dimensional tables is supplied. This method returns the relevant question to the data set. For example,

```r
see_question(Census2016_ancestries)
```

18. What is the person’s ancestry?
   - Provide up to two ancestries only.
   - Examples of ‘Other’: GREEK, VIETNAMESE, HMONG, KURDISTAN, MAORI, LEBANESE, AUSTRALIAN SOUTH SEA ISLANDER.
   - Remember to mark the box like this: —
   - Go to census.abs.gov.au for more information.

<table>
<thead>
<tr>
<th>English</th>
<th>Irish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scottish</td>
<td>Italian</td>
</tr>
<tr>
<td>German</td>
<td>Chinese</td>
</tr>
<tr>
<td>Australian</td>
<td>Australian</td>
</tr>
<tr>
<td>Other ancestry 1 (please specify)</td>
<td></td>
</tr>
<tr>
<td>Other ancestry 2 (please specify)</td>
<td></td>
</tr>
</tbody>
</table>
(The data input is returned invisibly.)