```
library(tidyverse)
library(tsibble)
library(readabs)
library(raustats)
```

Australian data analysts will know how frustrating it is to work with time series data from the Australian Bureau of Statistics. They are stored as multiple ugly Excel files (each containing multiple sheets) with inconsistent formatting, embedded comments, meta data stored along with the actual data, dates stored in a painful Excel format, and so on.

Fortunately there are now a couple of R packages available to make this a little easier. To illustrate them, I will recreate the `tsibbledata::aus_retail` data, containing monthly Australian retail trade turnover for different combinations of industry and state. Here is the data as provided in the `tsibbledata package`.

```
tsibbledata::aus_retail
```

```
## # A tsibble: 64,532 x 5 [1M]
## # Key:       State, Industry [152]
##    State              Industry                `Series ID`     Month Turnover
##
##  1 Australian Capit… Cafes, restaurants and… A3349849A    1982 Apr      4.4
##  2 Australian Capit… Cafes, restaurants and… A3349849A    1982 May      3.4
##  3 Australian Capit… Cafes, restaurants and… A3349849A    1982 Jun      3.6
##  4 Australian Capit… Cafes, restaurants and… A3349849A    1982 Jul      4
##  5 Australian Capit… Cafes, restaurants and… A3349849A    1982 Aug      3.6
##  6 Australian Capit… Cafes, restaurants and… A3349849A    1982 Sep      4.2
##  7 Australian Capit… Cafes, restaurants and… A3349849A    1982 Oct      4.8
##  8 Australian Capit… Cafes, restaurants and… A3349849A    1982 Nov      5.4
##  9 Australian Capit… Cafes, restaurants and… A3349849A    1982 Dec      6.9
## 10 Australian Capit… Cafes, restaurants and… A3349849A    1983 Jan      3.8
## # … with 64,522 more rows
```

There are 152 combinations of State and Industry (each corresponding to a `Series ID`). The data are from Apr 1982 to Dec 2018.

readabs

The `readabs` package has been around longest, and is maintained by Matt Cowgill from the Grattan Institute — which means it has probably had a very thorough workout!

The main function is `read_abs()` which will download the data, read it into R, and tidy it. While it will do this for all spreadsheets in a given catalogue number, I would not recommend that. Choose the spreadsheet you want. In this case, it is Cat 8501.0, Table 11.

```
system.time(retail1 <- read_abs("8501.0", tables = 11))
```

```
## Finding filenames for tables corresponding to ABS catalogue 8501.0
```

```
## Attempting to download files from catalogue 8501.0, Retail Trade, Australia
```

```
## Extracting data from downloaded spreadsheets
```

```
## Tidying data from imported ABS spreadsheets
```

```
##    user  system elapsed
##   1.362   0.177   3.920
```

```
retail1
```

```
## # A tibble: 85,428 x 12
```

```
##    table_no sheet_no table_title date       series value series_type
##
##  1 8501011  Data1    TABLE 11. … 1982-04-01 Turno…  303. Original
##  2 8501011  Data1    TABLE 11. … 1982-05-01 Turno…  298. Original
##  3 8501011  Data1    TABLE 11. … 1982-06-01 Turno…  298  Original
##  4 8501011  Data1    TABLE 11. … 1982-07-01 Turno…  308. Original
##  5 8501011  Data1    TABLE 11. … 1982-08-01 Turno…  299. Original
##  6 8501011  Data1    TABLE 11. … 1982-09-01 Turno…  305. Original
##  7 8501011  Data1    TABLE 11. … 1982-10-01 Turno…  318  Original
##  8 8501011  Data1    TABLE 11. … 1982-11-01 Turno…  334. Original
##  9 8501011  Data1    TABLE 11. … 1982-12-01 Turno…  390. Original
## 10 8501011  Data1    TABLE 11. … 1983-01-01 Turno…  311. Original
## # … with 85,418 more rows, and 5 more variables: data_type ,
## #   collection_month , frequency , series_id , unit
```

Some of those columns are not particularly useful (containing a single unique value), so we will remove them. We also need to fix the date to be a Month (rather than Day), and we will match the names to `tsibbledata::aus_retail` to make comparisons easier.

```
retail1 <- retail1 %>%
  mutate(Month = yearmonth(date)) %>%
  rename(Turnover = value, `Series ID` = series_id) %>%
  select(Month, `Series ID`, series, Turnover)
retail1
```

```
## # A tibble: 85,428 x 4
##        Month `Series ID` series                                      Turnover
##
##  1  1982 Apr A3349335T   Turnover ;  New South Wales ;  Supermark…     303.
##  2  1982 May A3349335T   Turnover ;  New South Wales ;  Supermark…     298.
##  3  1982 Jun A3349335T   Turnover ;  New South Wales ;  Supermark…     298
##  4  1982 Jul A3349335T   Turnover ;  New South Wales ;  Supermark…     308.
##  5  1982 Aug A3349335T   Turnover ;  New South Wales ;  Supermark…     299.
##  6  1982 Sep A3349335T   Turnover ;  New South Wales ;  Supermark…     305.
##  7  1982 Oct A3349335T   Turnover ;  New South Wales ;  Supermark…     318
##  8  1982 Nov A3349335T   Turnover ;  New South Wales ;  Supermark…     334.
##  9  1982 Dec A3349335T   Turnover ;  New South Wales ;  Supermark…     390.
## 10  1983 Jan A3349335T   Turnover ;  New South Wales ;  Supermark…     311.
## # … with 85,418 more rows
```

The `series` column contains information about the state and industry, so we will need to extract the relevant details. Also, totals are included in addition to the disaggregated data, so let's remove them.

```
retail1 <- retail1 %>%
  separate(series, c("Category", "State", "Industry"), sep = ";", extra =
"drop") %>%
  mutate(
    State = trimws(State),
    Industry = trimws(Industry),
  ) %>%
  select(-Category) %>%
  filter(
    Industry  != "Total (Industry)",
    State != "Total (State)"
  )
```

Next, we turn it into a tsibble by identifying the index and key variables, and removing any missing combinations.

```
retail1 <- retail1 %>%
  as_tsibble(index = Month, key = c(State, Industry)) %>%
  filter(!is.na(Turnover))
retail1
```

```
## # A tsibble: 66,154 x 5 [1M]
## # Key:       State, Industry [152]
##         Month `Series ID` State            Industry                Turnover
##
##  1   1982 Apr A3349849A   Australian Capi… Cafes, restaurants and…      4.4
##  2   1982 May A3349849A   Australian Capi… Cafes, restaurants and…      3.4
##  3   1982 Jun A3349849A   Australian Capi… Cafes, restaurants and…      3.6
##  4   1982 Jul A3349849A   Australian Capi… Cafes, restaurants and…      4
##  5   1982 Aug A3349849A   Australian Capi… Cafes, restaurants and…      3.6
##  6   1982 Sep A3349849A   Australian Capi… Cafes, restaurants and…      4.2
##  7   1982 Oct A3349849A   Australian Capi… Cafes, restaurants and…      4.8
##  8   1982 Nov A3349849A   Australian Capi… Cafes, restaurants and…      5.4
##  9   1982 Dec A3349849A   Australian Capi… Cafes, restaurants and…      6.9
## 10   1983 Jan A3349849A   Australian Capi… Cafes, restaurants and…      3.8
## # … with 66,144 more rows
```

The additional rows here compared to `tsibbledata::aus_retail` are because the data now extend to November 2019.

There's a helpful vignette demonstrating other facilities and features of the `readabs` package.

raustats

The `raustats` package is more recent, and aims to do a little more than `readabs` as it also covers non-time-series data from the ABS as well as data from the Reserve Bank of Australia. It is maintained by David Mitchell.

To download the relevant data, we use the `abs_cat_stats()` function:

```
system.time(retail2 <- abs_cat_stats("8501.0", tables = "11"))
```

```
##    user  system elapsed
##   1.002   0.039   1.863
```

It seems to be about twice as fast as the `readabs` package.

```
retail2 <- as_tibble(retail2)
retail2
```

```
## # A tibble: 79,190 x 16
##    date       series_id value data_item_descr… series_type series_start
##
##  1 1982-04-01 A3349335T  303. Turnover ;  New… Original    1982-04-01
##  2 1982-05-01 A3349335T  298. Turnover ;  New… Original    1982-04-01
##  3 1982-06-01 A3349335T  298  Turnover ;  New… Original    1982-04-01
##  4 1982-07-01 A3349335T  308. Turnover ;  New… Original    1982-04-01
##  5 1982-08-01 A3349335T  299. Turnover ;  New… Original    1982-04-01
##  6 1982-09-01 A3349335T  305. Turnover ;  New… Original    1982-04-01
##  7 1982-10-01 A3349335T  318  Turnover ;  New… Original    1982-04-01
##  8 1982-11-01 A3349335T  334. Turnover ;  New… Original    1982-04-01
##  9 1982-12-01 A3349335T  390. Turnover ;  New… Original    1982-04-01
## 10 1983-01-01 A3349335T  311. Turnover ;  New… Original    1982-04-01
## # … with 79,180 more rows, and 10 more variables: series_end ,
## #   no_obs , unit , data_type , freq ,
## #   collection_month , catalogue_no , publication_title ,
```

```
## #   table_no , table_title
```

Then we repeat the steps above to create a tsibble.

```
retail2 <- retail2 %>%
  mutate(Month = yearmonth(date)) %>%
  rename(Turnover = value) %>%
  separate(data_item_description,
           c("Category", "State", "Industry"), sep = ";", extra = "drop") %>%
  mutate(
    State = trimws(State),
    Industry = trimws(Industry),
  ) %>%
  filter(
    Industry  != "Total (Industry)",
    State != "Total (State)"
  ) %>%
  as_tsibble(index = Month, key = c(State, Industry)) %>%
  filter(!is.na(Turnover)) %>%
  rename(`Series ID` = series_id) %>%
  select(Month, `Series ID`, State, Industry, Turnover)
```

Finally we check that the result is the same as that obtained with the `readabs` package.

```
identical(retail2,retail1)
```

```
## [1] TRUE
```