Appendix to cubble: An R Package for Organizing and Wrangling Multivariate Spatio-temporal Data

Sherry Zhang

2022-06-14

1 Linking pipeline

Figure 1 illustrates the linking pipeline from a point selection on the time series to the map.

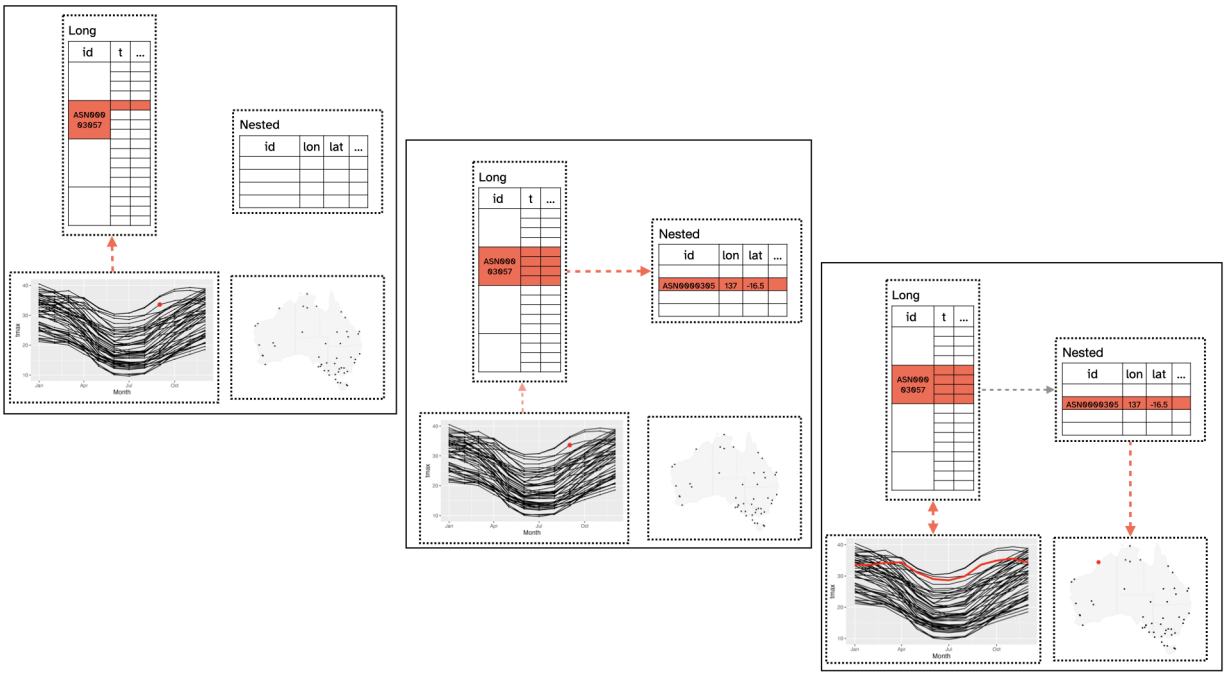

Figure 1: An illustration of the data model under interactive graphics with cubble. When a point on the time series is selected, the corresponding row in the long cubble will be activated. This will link to all the rows with the same id in the long cubble and the row in the nested cubble with the same id (middle). Both plots will be updated with the full line selected and the point highlighted on the map (right).

2 Script to obtain example data

The script below presents the codes required to obtain the data, `historical_tmax`, used in the example of Section 5.2 *Australian historical maximum temperature*. The function `rnoaa::meteo_pull_monitors()` may take a while to query a large number of stations in the first time. A copy of the data is provided in the data folder of the paper repository at: https://github.com/huizezhang-sherry/paper-cubble.

```
library(tidyverse)
library(cubble)
all_stations <- rnoaa::ghcnd_stations() %>%
  filter(str_starts(id, "ASN")) %>% # Australian stations start wiht "ASN"
  filter(last_year >= 2020) %>%
  mutate(wmo_id = as.numeric(wmo_id), name = str_to_lower(name)) %>%
  select(-state, -gsn_flag) %>%
  select(id, longitude, latitude, elevation, name,
         wmo_id, element, first_year, last_year) %>%
  rename(long = longitude, lat = latitude, elev = elevation)

tmax_stations <- all_stations %>%
  filter(element == "TMAX", first_year < 1970, !is.na(wmo_id))

raw_tmax <- all_stations %>%
  rowwise() %>%
  mutate(ts = list(rnoaa::meteo_pull_monitors(
    monitors = id, var = "TMAX",
    date_min = glue::glue("{first_year}-01-01"),
    date_max = glue::glue("{last_year}-12-31")
    ) %>%
      select(-id)
    )
  )

historical_tmax <- raw_tmax %>%
  select(-element) %>%
  unnest(ts) %>%
  mutate(tmax = tmax/10) %>%
  filter(lubridate::year(date) %in% c(1971: 1975, 2016:2020)) %>%
  as_cubble(index = date, key = id, coords = c(long, lat))

save(historical_tmax, file = here::here("data/historical_tmax.rda"))
```

The data `climate_full` used in Section 5.3, 5.4, and 5.6 can be obtained in a similar fashion with a change of variable and date parameter in `rnoaa::meteo_pull_monitors()`. The full script is provided below and a copy of the data is also available in the data folder of the paper repository linked above.

```
# all the Australian stations have all of the three PRCP, TMAX, and TMIN recorded
aus_stations <- all_stations %>%
  filter(element %in% c("PRCP", "TMAX", "TMIN")) %>%
  nest(element: last_year) %>%
  rowwise() %>%
  filter(nrow(data) == 3) %>%
  select(-data)

aus_climate_raw <- aus_stations %>%
  rowwise() %>%
  mutate(ts = list(
    rnoaa::meteo_pull_monitors(
      monitors = id, var = c("PRCP", "TMAX", "TMIN"),
      date_min = "2016-01-01", date_max = "2020-12-31"
      ) %>%
      select(-id)
```

```
    )
  )

climate_full <- aus_climate_raw %>%
  unnest(ts) %>%
  mutate(tmax = tmax/10, tmin = tmin/10) %>%
  cubble::as_cubble(key = id, index = date, coords = c(long, lat))

save(climate_full, file = here::here("data/climate_full.rda"))
```