Sydney_water

Seshadhri S

2025-05-13

Sydney Water Quality

This week we're exploring the water quality of Sydney's iconic beaches. The data is available at the New South Wales State Government Beachwatch website.

Beachwatch and our partners monitor water quality at swim sites to ensure that recreational water environments are managed as safely as possible so that as many people as possible can benefit from using the water.

Sydney beaches were in the news this summer with high rainfall causing concerns about the safety of the water.

The dataset this week includes both water quality and historical weather data from 1991 until 2025.

Has the water quality declined over this period? How does rainfall impact E-coli bacteria levels? Are some swimming sites particularly prone to high bacteria levels following rain? Thank you to Jen Richmond (R-Ladies Sydney) for curating this week's dataset.

```
# Using R
# Option 1: tidytuesdayR R package
## install.packages("tidytuesdayR")

tuesdata <- tidytuesdayR::tt_load('2025-05-20')
```

```
## ---- Compiling #TidyTuesday Information for 2025-05-20 ----
## --- There are 2 files available ---
##
##
## -- Downloading files --------------------------------------------------------
##
##   1 of 2: "water_quality.csv"
##   2 of 2: "weather.csv"
```

```
## OR
tuesdata <- tidytuesdayR::tt_load(2025, week = 20)
```

```
## ---- Compiling #TidyTuesday Information for 2025-05-20 ----
## --- There are 2 files available ---
##
##
## -- Downloading files --------------------------------------------------------
##
##   1 of 2: "water_quality.csv"
##   2 of 2: "weather.csv"
```

```
water_quality <- tuesdata$water_quality
weather <- tuesdata$weather

# Option 2: Read directly from GitHub

water_quality <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/main/da
```

```
## Rows: 123530 Columns: 10
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (3): region, council, swim_site
## dbl  (5): enterococci_cfu_100ml, water_temperature_c, conductivity_ms_cm, la...
## date (1): date
## time (1): time
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
weather <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/main/data/2025
```

```
## Rows: 12538 Columns: 6
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl  (5): max_temp_C, min_temp_C, precipitation_mm, latitude, longitude
## date (1): date
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.2     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.4
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(lubridate)
library(tidygeocoder)
```

Has the water quality declined over this period?

```
table(water_quality$region)
```

```
##
## Northern Sydney Southern Sydney     Sydney City  Sydney Harbour  Western Sydney
##           43430            18770           18606           41770             954
```

```
table(water_quality$council)
```

```
##
##                      Blue Mountains City Council
##                                              656
##                       City of Canada Bay Council
##                                             2663
##                          Hawkesbury City Council
##                                              247
##                               Inner West Council
##                                             1805
##                              Lane Cove Council
##                                             4594
##                         Mosman Municipal Council
##                                             6779
##                             North Sydney Council
##                                             1538
##                         Northern Beaches Council
##                                            54569
##                             Penrith City Council
##                                               51
##                            Randwick City Council
##                                            12137
##                         Sutherland Shire Council
##                                            18770
##                               The City of Sydney
##                                             1426
## The Council of the Municipality of Hunters Hill
##                                             2350
##                                  Waverley Council
##                                             6469
##                            Willoughby City Council
##                                             1553
##                      Woollahra Municipal Council
##                                             7923
```

```
table(water_quality$swim_site)
```

```
##
##                        Avalon Beach                      Balmoral Baths
##                                2064                                1552
##                   Bilarong Reserve                       Bilgola Beach
##                                 643                                2080
##                       Boat Harbour                         Bondi Beach
##                                2495                                2128
##                       Bronte Beach                       Bungan Beach
##                                2200                                2080
##                     Cabarita Beach                 Callan Park Seawall
##                                1405                                 257
```

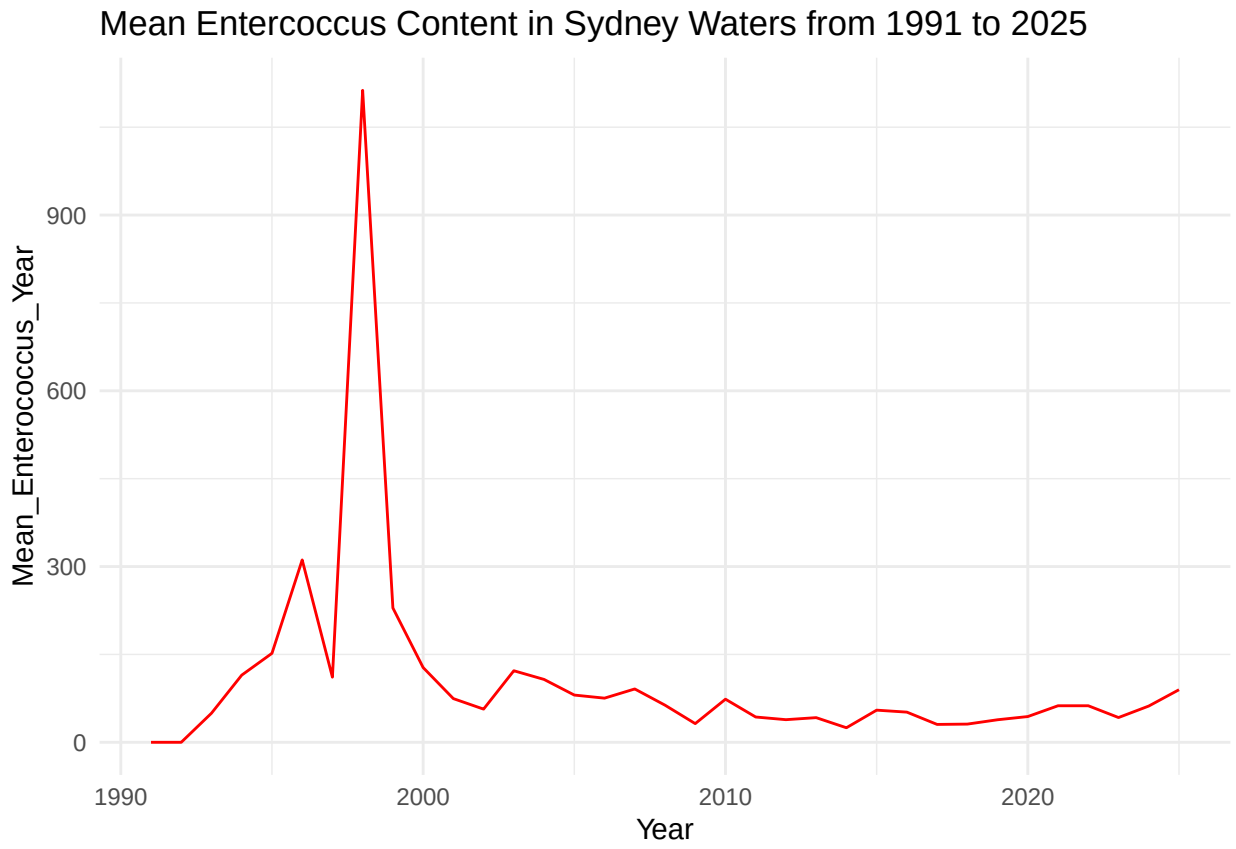
```
##                         Camp Cove                   Chinamans Beach
##                               398                              1258
##                     Chiswick Baths                 Clifton Gardens
##                               1258                             1499
##                      Clontarf Pool                  Clovelly Beach
##                               1554                             2151
##                     Collaroy Beach                    Coogee Beach
##                               2078                             2207
##                    Darling Harbour                Davidson Reserve
##                               1426                             1492
##                   Dawn Fraser Pool                   Dee Why Beach
##                               1548                             2077
##                     Edwards Beach                    Elouera Beach
##                               1502                             2508
##                    Fairlight Beach               Forty Baskets Pool
##                               1378                             1503
##                   Freshwater Beach               Gordons Bay (East)
##                               2058                              698
##                   Greenhills Beach                  Greenwich Baths
##                               2485                             1558
##              Gurney Crescent Baths               Hayes Street Beach
##                               1373                             1538
## Henley Baths (Kelly Street Baths)                 Little Bay Beach
##                                803                             1333
##                  Little Manly Cove               Little Sirius Cove
##                               1501                              968
##                   Long Reef Beach                    Malabar Beach
##                               2048                             2142
##                        Manly Cove                   Maroubra Beach
##                               1551                             2135
##                    Megalong Creek                   Mona Vale Beach
##                                172                             2063
##                   Murray Rose Pool  Narrabeen Lagoon (Birdwood Park)
##                               1500                             1152
##                      Newport Beach                      Nielsen Park
##                               2081                             1501
##               North Cronulla Beach             North Curl Curl Beach
##                               2500                             2146
##              North Narrabeen Beach                North Steyne Beach
##                               2076                             2146
##                  Northbridge Baths                    Oak Park Beach
##                               1553                             1886
##                       Palm Beach                       Parsley Bay
##                               2065                             1501
##                     Penrith Beach                  Queenscliff Beach
##                                 51                             2212
##                    Rose Bay Beach                   Sangrado Baths
##                               1522                              787
##               Shelly Beach (Manly)          Shelly Beach (Sutherland)
##                               2142                             1884
##               South Cronulla Beach              South Curl Curl Beach
##                               2505                             2059
##               South Maroubra Beach            South Maroubra Rockpool
##                                722                              749
```

```
##                South Steyne Beach                    Tamarama Beach
##                              2186                              2141
##                   Tambourine Bay                   Turimetta Beach
##                              1499                              1815
##                      Wanda Beach                  Warriewood Beach
##                              2507                              2077
##                      Watsons Bay     Wentworth Falls Lake - Beach
##                              1501                               183
##     Wentworth Falls Lake - Jetty                       Whale Beach
##                               191                              2082
##                    Windsor Beach                     Woodford Bay
##                               124                              1537
##                    Woolwich Baths                Yarramundi Reserve
##                              1547                               123
##  Yosemite Creek - Minnehaha Falls
##                               110
```

```
water_quality_clean <- water_quality %>% filter(!is.na(enterococci_cfu_100ml))
class(water_quality$date)
```

```
## [1] "Date"
```

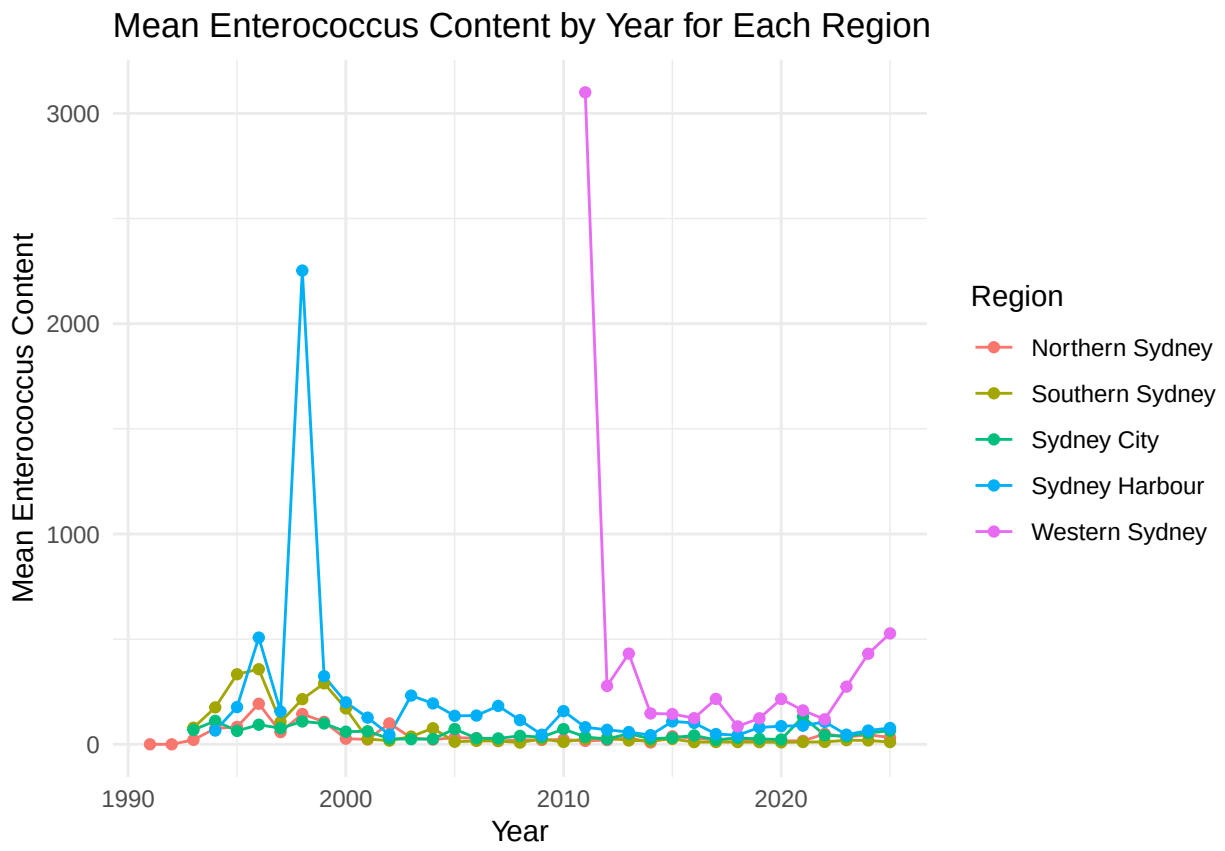
```
water_quality_clean <- water_quality_clean %>% mutate(Year = year(date))
water_quality_summarised <- water_quality_clean %>% group_by(Year) %>% summarise(Mean_Enterococcus_Year

ggplot(water_quality_summarised, aes(x = Year, y = Mean_Enterococcus_Year)) +
  geom_line(color = "red") +
  labs(title = "Mean Entercoccus Content in Sydney Waters from 1991 to 2025", xlab = "Year", ylab = "Mea
  theme_minimal()
```




```
#Water quality by Swim site, council and region
# Region
mean_content_region <- water_quality_clean %>%
  group_by(region, Year) %>%
  summarise(mean_enterococcus = mean(enterococci_cfu_100ml, na.rm = TRUE)) %>%
  ungroup()
```

```
## `summarise()` has grouped output by 'region'. You can override using the
## `.groups` argument.
```

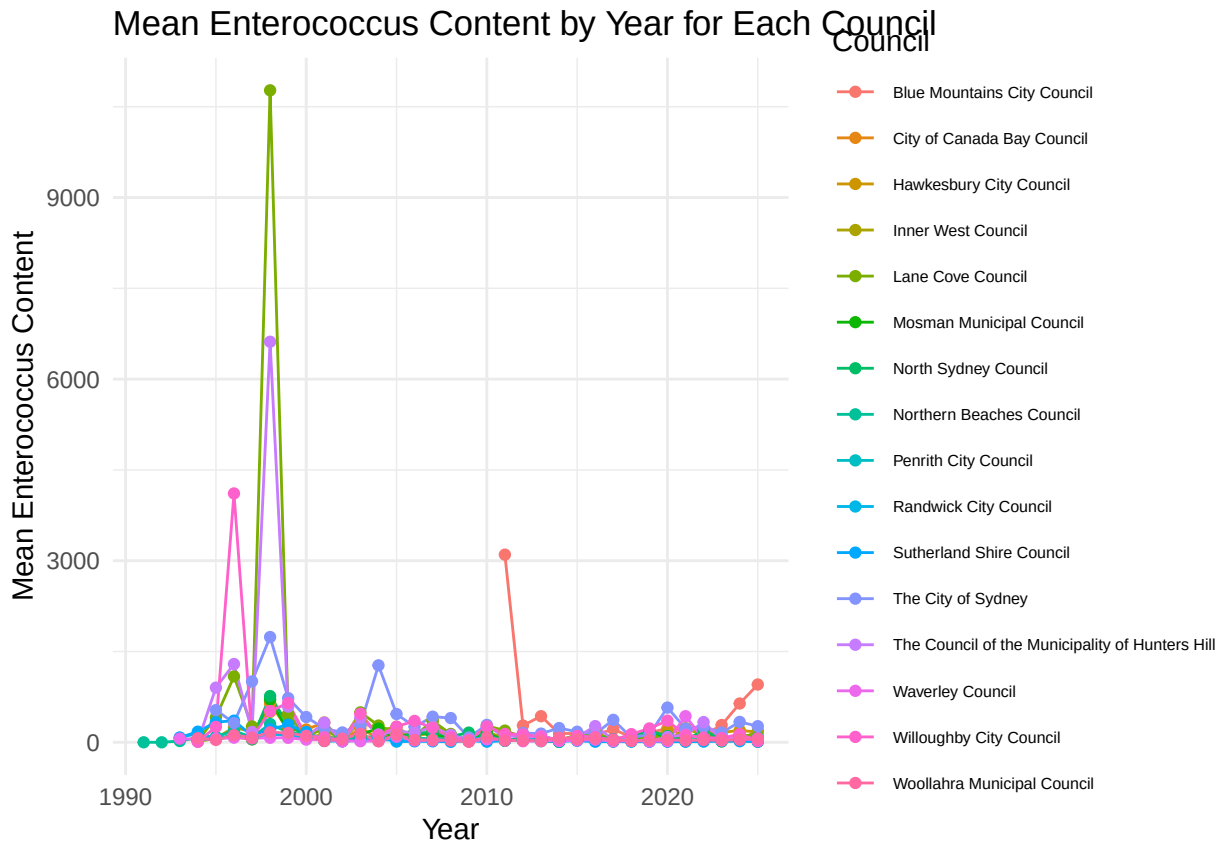
```
ggplot(mean_content_region, aes(x = Year, y = mean_enterococcus, color = region)) +
  geom_line() +
  geom_point() +
  labs(title = "Mean Enterococcus Content by Year for Each Region",
       x = "Year",
       y = "Mean Enterococcus Content",
       color = "Region") +
  theme_minimal()
```




```
# Council

mean_content_council <- water_quality_clean %>%
  group_by(council, Year) %>%
  summarise(mean_enterococcus = mean(enterococci_cfu_100ml, na.rm = TRUE)) %>%
  ungroup()
```

```
## `summarise()` has grouped output by 'council'. You can override using the
## `.groups` argument.
```

```
ggplot(mean_content_council, aes(x = Year, y = mean_enterococcus, color = council)) +
  geom_line() +
  geom_point() +
  labs(title = "Mean Enterococcus Content by Year for Each Council",
       x = "Year",
       y = "Mean Enterococcus Content",
       color = "Council") +
  theme_minimal() +
  theme(legend.text = element_text(size = 6))
```



How does rainfall impact E-coli bacteria levels?

```
rainfall <- weather %>% select(date, precipitation_mm)
rainfall <- rainfall %>% mutate(Year = year(date))
rainfall <- rainfall %>% group_by(Year) %>% summarise(Mean_Rainfall = mean(precipitation_mm))

rainfall_ecoli_merged <- merge(rainfall, water_quality_summarised, by = "Year")
cor(rainfall_ecoli_merged$Mean_Rainfall, rainfall_ecoli_merged$Mean_Enterococcus_Year, method = "spearma
```

```
## [1] -0.03039429
```

```
weather_places <- weather %>%
  reverse_geocode(lat = latitude, long = longitude, method = "osm")
```

```
## Passing 1 coordinate to the Nominatim single coordinate geocoder
```

```
## Query completed in: 1.1 seconds
```

I thought the latitude and longitude were different for each measurement. Turns out they aren't so couldn't go further. At least I learnt about tidygeocoder

There's also no real correlation vbetween rainfall and bacteria levels. I didn't really feel like doing another line plot if there isn't a massive correlation over time.

Are some swimming sites particularly prone to high bacteria levels following rain?

Didn't have time to do this