```r
# load packages
library(tidyverse)
library(lubridate) # for dealing with dates
```

Code to clean daily coronavirus data from California. The data comes from The COVID Tracking Project:

https://covidtracking.com/data/api

The data set was used for a HW problem in STAT 452.

```r
# read in in CSV file directly from website
covid_state <- read_csv("https://api.covidtracking.com/v1/states/ca/daily.csv")
```

```r
# parse date column
covid_ca <- covid_state %>%
  filter(state == "CA") %>%
  select(state, daily_cases = positiveIncrease, date) %>%
  mutate(date = ymd(date)) %>%
  mutate(year = year(date)) %>%
  mutate(month = month(date)) %>%
  mutate(yday = yday(date))
```

```r
# subset data for 2021
covid_ca2021 <- covid_ca %>% filter(year == 2021)
```

```r
# look at first 5 entries
covid_ca2021 %>% slice_head(n=5)
```

```
## # A tibble: 5 x 6
##   state daily_cases date        year month  yday
##   <chr>       <dbl> <date>     <dbl> <dbl> <dbl>
## 1 CA           3816 2021-03-07  2021     3    66
## 2 CA           4452 2021-03-06  2021     3    65
## 3 CA           4659 2021-03-05  2021     3    64
## 4 CA           3504 2021-03-04  2021     3    63
## 5 CA           3352 2021-03-03  2021     3    62
```

```r
# look at last 5 entries
covid_ca2021 %>% slice_tail(n=5)
```

```
## # A tibble: 5 x 6
##   state daily_cases date        year month  yday
##   <chr>       <dbl> <date>     <dbl> <dbl> <dbl>
## 1 CA          31440 2021-01-05  2021     1     5
## 2 CA          29633 2021-01-04  2021     1     4
## 3 CA          45352 2021-01-03  2021     1     3
## 4 CA          53341 2021-01-02  2021     1     2
## 5 CA          47189 2021-01-01  2021     1     1
```

```r
# write data set to file
# saveRDS(covid_ca2021, "covid_ca2021.rds")
```