

# Data Analysis of New York Times Covid-19 Data

## Working with the New York Times COVID-19 Data

The purpose of this project is to analyze COVID-19 datasets collected by the *New York Times*, which are available in a repository on [Github] (<https://github.com/nytimes/covid-19-data>). These datasets will be utilized to reproduce some of the elements of the *Timegets*'s reporting on COVID-19. Download the files `us-counties.csv`, `us-states.csv`, and `us.csv` by copying and pasting each permalink into [this website] <https://minhaskamal.github.io/DownGit/#/home>. Save the datasets in the same directory as the `.Rmd` file. Adjust the working directory accordingly (<https://bookdown.org/yihui/rmarkdown-cookbook/working-directory.html>).

```
# Set working directory to access datasets
setwd("/Users/siddharth/Documents/UC Davis/STA 141A")
```

## Prerequisite Information

The file `us.csv` contains aggregated data for the entire U.S. In this file, the `cases` and `deaths` columns represent *cumulative* cases and deaths due to covid. The `diff()` function can be used to compute the differences between each consecutive element of a vector, so it could be used to compute the daily numbers of cases and deaths. However, `diff()` returns a vector of length one less than the length of the original vector (e.g., `diff(c(1, 3, 6, 10))` returns 2, 3, 4) and this can make it somewhat inconvenient to use when transforming columns of data frames.

An alternative is to use the more general `filter()` function with an appropriate choice of the `filter` and `sides` arguments. This function can also be utilized for computing running averages and similar quantities.

## Section 1

### Step: 1A

Read the file `us.csv` into R as the data frame `us`.

```
# Read CSV
us <- read.csv("Data/us.csv")

head(us, 10)
```

```
##           date cases deaths
## 1  2020-01-21     1       0
## 2  2020-01-22     1       0
## 3  2020-01-23     1       0
## 4  2020-01-24     2       0
```

```
## 5 2020-01-25 3 0
## 6 2020-01-26 5 0
## 7 2020-01-27 5 0
## 8 2020-01-28 5 0
## 9 2020-01-29 5 0
## 10 2020-01-30 6 0
```

## Step: 1B

Transform the `date` column into a column of class `Date`.

```
# Transform class of 'date' column to 'Date'
us$date <- as.Date(us$date)

# Confirm class transformation worked
data.frame(sapply(us,class))
```

```
##          sapply.us..class.
## date                      Date
## cases                    integer
## deaths                   integer
```

## Step: 1C

Use `filter()` to add a column named `new_cases` containing the number of new cases reported on each date. The first value in this column will be `NA`.

```
# Calculate new cases & add new column
us$new_cases <- abs(stats::filter(us$cases, filter = c(-1, 1), sides = 1))

head(us, 10)
```

```
##          date cases deaths new_cases
## 1 2020-01-21 1 0 NA
## 2 2020-01-22 1 0 0
## 3 2020-01-23 1 0 0
## 4 2020-01-24 2 0 1
## 5 2020-01-25 3 0 1
## 6 2020-01-26 5 0 2
## 7 2020-01-27 5 0 0
## 8 2020-01-28 5 0 0
## 9 2020-01-29 5 0 0
## 10 2020-01-30 6 0 1
```

## Step: 1D

Use `filter()` to add a column named `new_deaths` containing the number of new deaths reported on each date. The first value in this column will be `NA`.

```
# Calculate new deaths & add new column
us$new_deaths <- abs(stats::filter(us$deaths, filter = c(-1, 1), sides = 1))

head(us, 10)
```

```
##           date cases deaths new_cases new_deaths
## 1  2020-01-21     1      0         NA          NA
## 2  2020-01-22     1      0          0          0
## 3  2020-01-23     1      0          0          0
## 4  2020-01-24     2      0          1          0
## 5  2020-01-25     3      0          1          0
## 6  2020-01-26     5      0          2          0
## 7  2020-01-27     5      0          0          0
## 8  2020-01-28     5      0          0          0
## 9  2020-01-29     5      0          0          0
## 10 2020-01-30     6      0          1          0
```

## Step: 1E

Use `filter()` to add a column named `avg_new_cases` where each element represents the mean number of new cases for the previous 7 days (inclusive of the current day). The first 7 values in this column will be NA.

```
# Calculate 7-day running average for new cases & add new column
us$avg_new_cases <- stats::filter(us$new_cases, filter = rep(1/7, 7), sides = 1)

head(us, 10)
```

```
##           date cases deaths new_cases new_deaths avg_new_cases
## 1  2020-01-21     1      0         NA          NA           NA
## 2  2020-01-22     1      0          0          0           NA
## 3  2020-01-23     1      0          0          0           NA
## 4  2020-01-24     2      0          1          0           NA
## 5  2020-01-25     3      0          1          0           NA
## 6  2020-01-26     5      0          2          0           NA
## 7  2020-01-27     5      0          0          0           NA
## 8  2020-01-28     5      0          0          0      0.5714286
## 9  2020-01-29     5      0          0          0      0.5714286
## 10 2020-01-30     6      0          1          0      0.7142857
```

## Step: 1F

Use `filter()` to add a column named `avg_new_deaths` where each element represents the mean number of new deaths for the previous 7 days (inclusive of the current day). The first 7 values in this column will be NA.

```
# Calculate 7-day running average for new deaths & add new column
us$avg_new_deaths <- stats::filter(us$new_deaths, filter = rep(1/7, 7), sides = 1)

head(us, 10)
```

##	date	cases	deaths	new_cases	new_deaths	avg_new_cases	avg_new_deaths
## 1	2020-01-21	1	0	NA	NA	NA	NA
## 2	2020-01-22	1	0	0	0	NA	NA
## 3	2020-01-23	1	0	0	0	NA	NA
## 4	2020-01-24	2	0	1	0	NA	NA
## 5	2020-01-25	3	0	1	0	NA	NA
## 6	2020-01-26	5	0	2	0	NA	NA
## 7	2020-01-27	5	0	0	0	NA	NA
## 8	2020-01-28	5	0	0	0	0.5714286	0
## 9	2020-01-29	5	0	0	0	0.5714286	0
## 10	2020-01-30	6	0	1	0	0.7142857	0

## Note

The `filter()` function used here is `stats::filter()` from the `stats` package, which is loaded by default in R. (The `dplyr` package has a completely different `filter()` function which plays an important role in the `tidyverse`. If you have problems using `filter()`, you should make sure that you do NOT have the `dplyr` package loaded. If you do, then you will need to explicitly type out `stats::filter()` to get the `stats` version.)

## Step: 1G

Create a plot of daily cases similar to the one found at the top of [this page] (<https://www.nytimes.com/interactive/2020/us/coronavirus-cases.html>..) Plot only data beginning from 2020-03-01. (Note that this plot and a similar plot for deaths appear again about 1/3 of the way down the page.)

## Hint

Try to do this using the formula method of the `plot()` function with the optional arguments `type = "h"`, `col = "gray"`, and `data = us` and using the `subset()` argument to plot only the data for dates 2020-03-01 and after. (You may also wish to experiment with the optional argument `lwd`.)

Then use the formula interface to the `lines()` function to add the curve showing the seven-day running average. (Again, you may wish to experiment with the optional argument `lwd`.)

```
# Extract data beginning 2020-03-01
us <- subset(us, date >= as.Date("2020-03-01"))

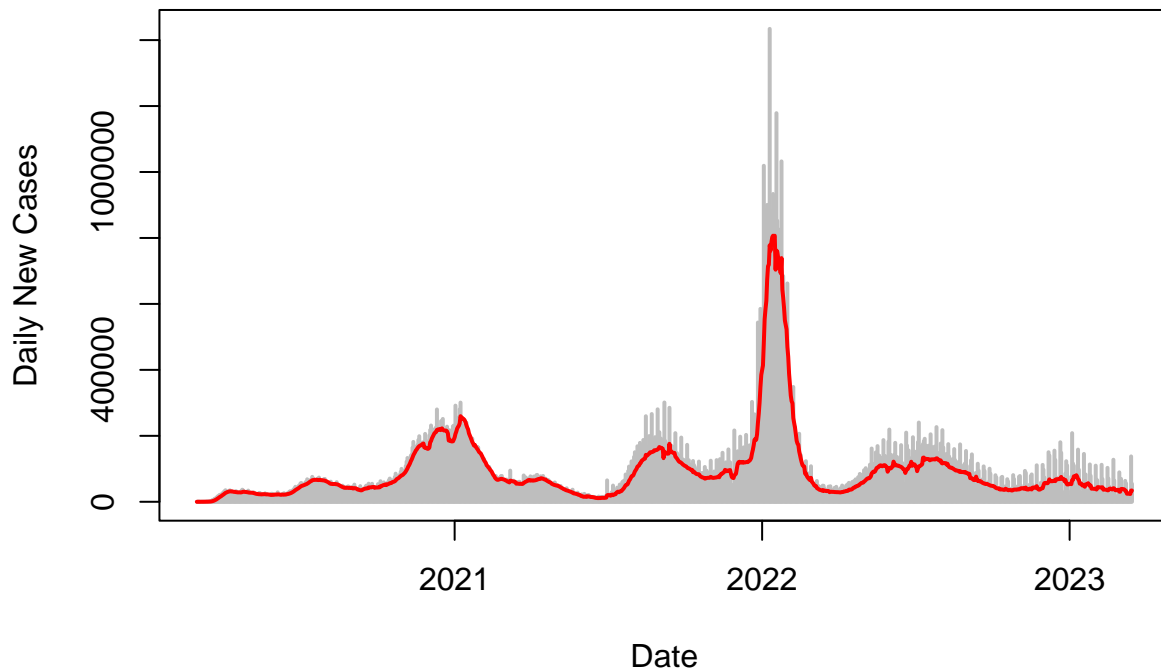
# Set width for knitting PDF
options(width = 60)
matrix(runif(100), ncol = 20)
```

##	[,1]	[,2]	[,3]	[,4]	[,5]
## [1,]	0.8488571	0.51810066	0.6347891	0.04298537	0.1959058
## [2,]	0.7305916	0.74792679	0.9875276	0.63723371	0.4333696
## [3,]	0.5298447	0.97649053	0.6967765	0.72337938	0.1454256
## [4,]	0.7760162	0.98441999	0.1596321	0.71918775	0.9761530
## [5,]	0.4865112	0.07593038	0.2479133	0.05084340	0.3341409
##	[,6]	[,7]	[,8]	[,9]	[,10]

```
## [1,] 0.6450608 0.2147123 0.1663287 0.4100856 0.25776162
## [2,] 0.5341856 0.5365418 0.9698059 0.3583807 0.48638924
## [3,] 0.5205952 0.9164879 0.9132474 0.8468276 0.08683906
## [4,] 0.3999742 0.2715675 0.7061580 0.6001983 0.18629600
## [5,] 0.4349671 0.5236110 0.9265431 0.9113421 0.57480655
##      [,11]      [,12]      [,13]      [,14]      [,15]
## [1,] 0.08342780 0.5523467 0.91532241 0.77502930 0.8622995
## [2,] 0.29508204 0.9677666 0.05721693 0.71025087 0.5375266
## [3,] 0.14475222 0.6521267 0.87844870 0.08718473 0.6340847
## [4,] 0.44424790 0.4264633 0.77909443 0.05613744 0.0537249
## [5,] 0.03082635 0.7725778 0.60618495 0.21979101 0.1597619
##      [,16]      [,17]      [,18]      [,19]      [,20]
## [1,] 0.9951533 0.91958019 0.64081796 0.34071057 0.2376510
## [2,] 0.9013032 0.08947681 0.86577022 0.19491378 0.8732115
## [3,] 0.2626285 0.80972336 0.09278853 0.91268644 0.2455452
## [4,] 0.8272848 0.86360087 0.98043306 0.29908759 0.5591417
## [5,] 0.2634683 0.33399855 0.36214690 0.09914801 0.7046008
```

```
# Plot daily new cases & add line for 7-day running average
{plot(new_cases ~ date, type = "h", col = "gray", data = us, main = 'US Daily New Covid-19 Cases from 20
lines(us$date, us$avg_new_cases, col = "red", lwd= "2")}
```

## US Daily New Covid-19 Cases from 2020-03-01



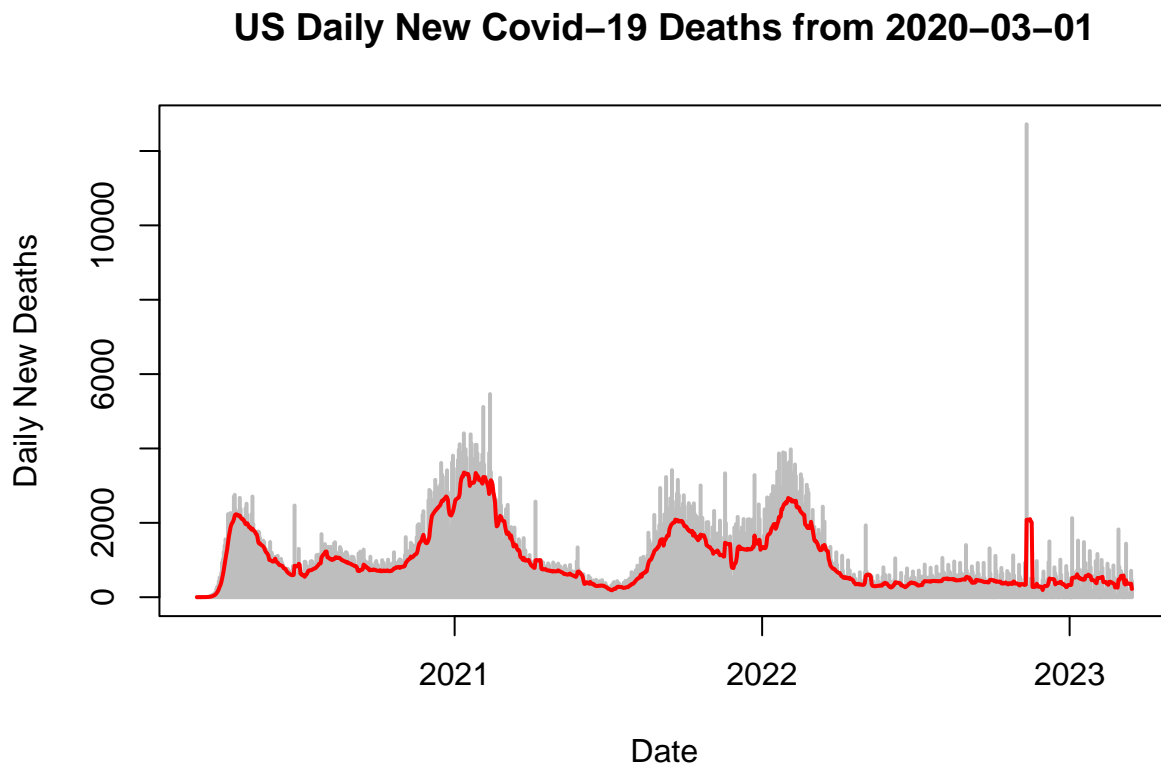
## Analysis

This graph illustrates the number of daily new cases of Covid-19 in the US, from 2020-03-01 to 2023-03-16. The gray portion of the graph indicates the daily new cases, and the red line shows the seven-day running average from the previous seven days. We can see multiple peaks in the number of daily new cases throughout the 3 year period, with the first peak occurring in 2021, and the greatest peak occurring in the beginning of 2022. The occurrence of daily new Covid-19 cases appears to slow down toward the middle of 2021. A probable reason for this pattern is that the Covid-19 vaccine was first released at the end of 2020, so it could have taken a few months for the vaccine to have a nationwide impact. Thus, the US experienced a relative low for the number of new cases in the middle of 2021. Considering that the cases and deaths of Covid-19 were likely under-reported, it is interesting to think about what this graph would look like if the data was completely accurate.

## Step: 1H

Repeat **Step: 1G** for deaths.

```
# Plot daily new deaths & add line for 7-day running average
{plot(new_deaths ~ date, type = "h", col = "gray", data = us, main = 'US Daily New Covid-19 Deaths from
lines(us$date, us$avg_new_deaths, col = "red", lwd= "2")}
```



## Analysis

This graph illustrates the number of daily new deaths due to Covid-19 in the US, from 2020-03-01 to 2023-03-16. The gray portion of the graph indicates the daily new deaths, and the red line shows the seven-day running average from the previous seven days. We can see multiple peaks in the number of daily new deaths throughout the 3 year period, with the main peaks occurring in early 2020, the beginning of 2021, and the few months before and after 2022. The observation which strongly stood out to me was that the number of cases and deaths seem to be strongly correlated, especially during the peak time periods of both graphs. However, the correlation is not perfectly proportional. It is interesting to note that the number of new deaths was significantly large in early 2020, while the number of new cases was relatively low. This can be explained by the fact that a majority of the US was unaware of the presence and severity of Covid-19. In addition, the vaccine had not yet been released, and the growth in ‘herd immunity’ had not occurred yet. Lastly, we can observe that the number of deaths remained consistently low after the middle of 2022. Surprisingly, there does appear to be an outlier peak in deaths near the end of 2022.

## Section 2

The file `us-states.csv` contains state-level data for the U.S.

### Step: 2A

Read `us-states.csv` into R as the data frame `us_states` and transform the `date` column into a column of class `Date`.

```
# Read CSV
us_states <- read.csv("Data/us-states.csv")

# Transform class of 'date' column to 'Date'
us_states$date <- as.Date(us_states$date)

# Confirm class transformation worked
data.frame(sapply(us_states, class))
```

```
##      sapply.us_states..class.
## date                        Date
## state                      character
## fips                       integer
## cases                      integer
## deaths                     integer
```

### Step: 2B

Use `subset()` to extract the data for the state of California and save it as a data frame named `California`. Be sure that the rows are correctly ordered by date, and then repeat **Steps: 1G & 1H** of this assignment for California, i.e., plot the number of daily new cases and deaths, along with their 7-day running averages.

```
# Extract data for California
California <- subset(us_states, state == "California")
```

```
# Confirm correct order of rows by date
California <- California[order(California$date),]

head(California, 10)
```

```
##           date      state fips cases deaths
## 6  2020-01-25 California    6     1      0
## 10 2020-01-26 California    6     2      0
## 14 2020-01-27 California    6     2      0
## 18 2020-01-28 California    6     2      0
## 22 2020-01-29 California    6     2      0
## 26 2020-01-30 California    6     2      0
## 30 2020-01-31 California    6     3      0
## 34 2020-02-01 California    6     3      0
## 39 2020-02-02 California    6     6      0
## 44 2020-02-03 California    6     6      0
```

```
# Calculate new cases & add new column
California$new_cases <- abs(stats::filter(California$cases, filter = c(-1, 1), sides = 1))

# Calculate new deaths & add new column
California$new_deaths <- abs(stats::filter(California$deaths, filter = c(-1, 1), sides = 1))

head(California, 10)
```

```
##           date      state fips cases deaths new_cases
## 6  2020-01-25 California    6     1      0          NA
## 10 2020-01-26 California    6     2      0           1
## 14 2020-01-27 California    6     2      0           0
## 18 2020-01-28 California    6     2      0           0
## 22 2020-01-29 California    6     2      0           0
## 26 2020-01-30 California    6     2      0           0
## 30 2020-01-31 California    6     3      0           1
## 34 2020-02-01 California    6     3      0           0
## 39 2020-02-02 California    6     6      0           3
## 44 2020-02-03 California    6     6      0           0
##      new_deaths
## 6              NA
## 10              0
## 14              0
## 18              0
## 22              0
## 26              0
## 30              0
## 34              0
## 39              0
## 44              0
```

```
# Calculate 7-day running average for new cases & add new column
California$avg_new_cases <- stats::filter(California$new_cases, filter = rep(1/7, 7), sides = 1)

# Calculate 7-day running average for new deaths & add new column
```



```
California$avg_new_deaths <- stats::filter(California$new_deaths, filter = rep(1/7, 7), sides = 1)

head(California, 10)
```

```
##           date      state fips cases deaths new_cases
## 6  2020-01-25 California    6     1      0         NA
## 10 2020-01-26 California    6     2      0          1
## 14 2020-01-27 California    6     2      0          0
## 18 2020-01-28 California    6     2      0          0
## 22 2020-01-29 California    6     2      0          0
## 26 2020-01-30 California    6     2      0          0
## 30 2020-01-31 California    6     3      0          1
## 34 2020-02-01 California    6     3      0          0
## 39 2020-02-02 California    6     6      0          3
## 44 2020-02-03 California    6     6      0          0
##   new_deaths avg_new_cases avg_new_deaths
## 6           NA           NA             NA
## 10           0           NA             NA
## 14           0           NA             NA
## 18           0           NA             NA
## 22           0           NA             NA
## 26           0           NA             NA
## 30           0           NA             NA
## 34           0    0.2857143             0
## 39           0    0.5714286             0
## 44           0    0.5714286             0
```

```
# Extract data beginning 2020-03-01
California <- subset(California, date >= as.Date("2020-03-01"))

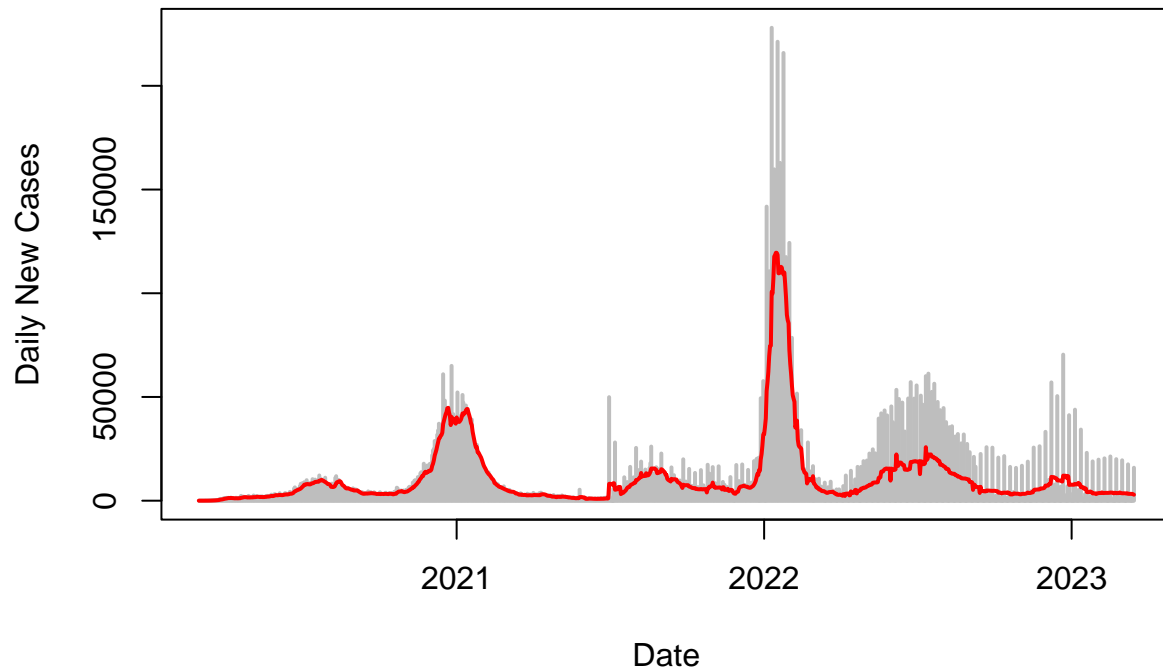
head(California, 10)
```

```
##           date      state fips cases deaths new_cases
## 242 2020-03-01 California    6    33      0          5
## 255 2020-03-02 California    6    38      0          5
## 270 2020-03-03 California    6    45      0          7
## 286 2020-03-04 California    6    55      1         10
## 303 2020-03-05 California    6    67      1         12
## 324 2020-03-06 California    6    81      1         14
## 352 2020-03-07 California    6   100      1         19
## 385 2020-03-08 California    6   112      1         12
## 420 2020-03-09 California    6   172      2         60
## 457 2020-03-10 California    6   179      3          7
##   new_deaths avg_new_cases avg_new_deaths
## 242         0    3.428571  0.0000000
## 255         0    3.857143  0.0000000
## 270         0    4.857143  0.0000000
## 286         1    4.142857  0.1428571
## 303         0    5.857143  0.1428571
## 324         0    7.714286  0.1428571
## 352         0   10.285714  0.1428571
## 385         0   11.285714  0.1428571
## 420         1   19.142857  0.2857143
```

```
## 457          1      19.142857      0.4285714
```

```
# Plot daily new cases & add line for 7-day running average
{plot(new_cases ~ date, type = "h", col = "gray", data = California, main = 'California Daily New Covid
lines(California$date, California$avg_new_cases, col = "red", lwd = "2")}
```

## California Daily New Covid-19 Cases from 2020-03-01

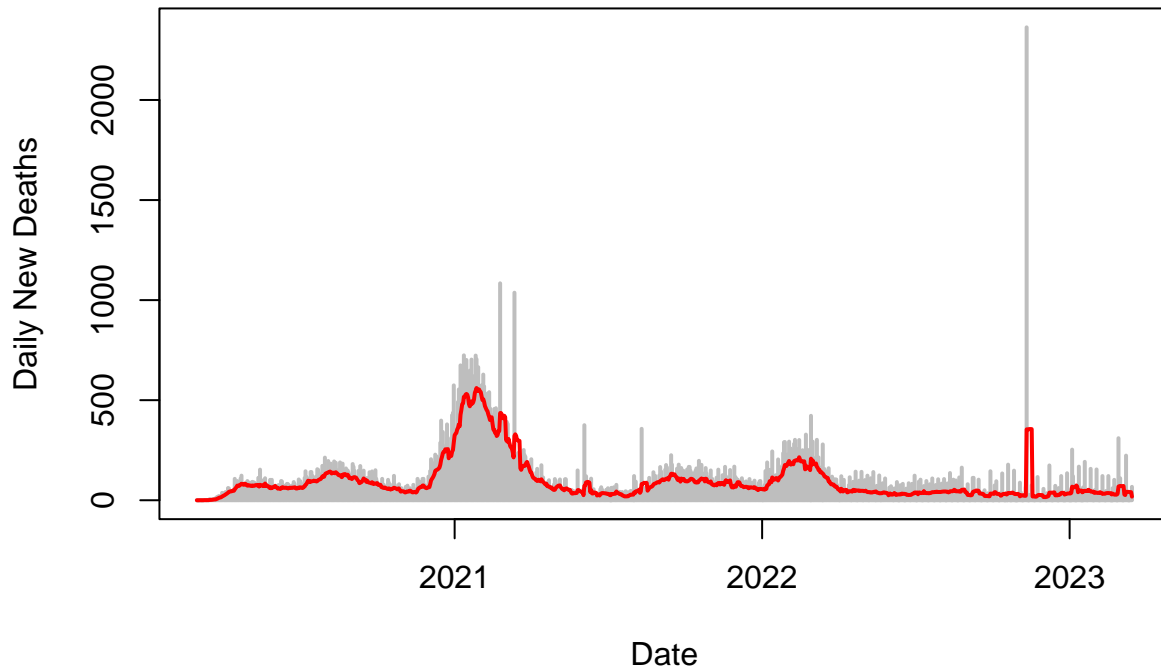


## Analysis

This graph illustrates the number of daily new cases of Covid-19 in California, from 2020-03-01 to 2023-03-16. The gray portion of the graph indicates the daily new cases, and the red line shows the seven-day running average from the previous seven days. We can see multiple peaks in the number of daily new cases throughout the 3 year period, with the main peaks occurring at the end of 2020, and the beginning of 2022. It is interesting to note that a majority of the peaks tend to occur around the beginning and end of each year. This could potentially be explained by the holiday season and celebrations. It is common for friends and families to gather in person during this time, which simultaneously encourages the spread of Covid-19. In addition, we can see that the number of daily new cases in California appears to be relatively similar to the number of daily new cases in the US. This makes sense because California is the nation's most populated state, so it would only be logical for the relationship of daily new cases to be proportional between the US and California.

```
# Plot daily new deaths & add line for 7-day running average
{plot(new_deaths ~ date, type = "h", col = "gray", data = California, main = 'California Daily New Covid
lines(California$date, California$avg_new_deaths, col = "red", lwd = "2")}
```

## California Daily New Covid-19 Deaths from 2020-03-01



### Analysis

This graph illustrates the number of daily new deaths of Covid-19 in California, from 2020-03-01 to 2023-03-16. The gray portion of the graph indicates the daily new deaths, and the red line shows the seven-day running average from the previous seven days. We can see multiple peaks in the number of daily new deaths throughout the 3 year period, with the main peaks occurring at the beginning of 2021 and 2022. The greatest peak occurred in the beginning of 2021, and was followed by a sharp decline. After that time period, the number of daily new deaths remained consistently low, with small peaks near the beginning and end of 2022. It is interesting to note that while the number of deaths in the US significantly increased in early 2020, the number of deaths in California remained relatively low. This indicates that when Covid-19 initially began to spread, other states must have been impacted before California. In addition, we can see a relatively consistent pattern between the US and California where the peaks in the number of new deaths tend to follow a few months after a peak in the number of new cases.

### Section 3

The file `us-counties.csv` contains county-level data for the U.S.

#### Step: 3A

Read `us-counties.csv` into R as the data frame `us_counties` and transform the `date` column into a column of class `Date`.

```

# Read CSV
us_counties <- read.csv("Data/us-counties.csv")

# Transform class of 'date' column to 'Date'
us_counties$date <- as.Date(us_counties$date)

# Confirm class transformation worked
data.frame(sapply(us_counties, class))

##           sapply.us_counties..class.
## date                               Date
## county                          character
## state                          character
## fips                            integer
## cases                          integer
## deaths                         integer

```

## Step: 3B

Use `subset()` to extract the data for Yolo County, California, and save it as a data frame named `Yolo`. Be sure that the rows are correctly ordered by date, and then repeat **Steps: 1G & 1H** for Yolo County, i.e., plot the number of daily new cases and deaths along with their 7-day running average.

```

# Extract data for California counties
State <- subset(us_counties, state == "California")

# Extract data for Yolo County
Yolo <- subset(State, county == "Yolo")

# Confirm correct orders of rows by date
Yolo <- Yolo[order(Yolo$date),]

head(Yolo, 10)

```

```

##           date county      state fips cases deaths
## 599  2020-03-06   Yolo California 6113     1      0
## 677  2020-03-07   Yolo California 6113     1      0
## 772  2020-03-08   Yolo California 6113     1      0
## 887  2020-03-09   Yolo California 6113     1      0
## 1021 2020-03-10   Yolo California 6113     1      0
## 1188 2020-03-11   Yolo California 6113     1      0
## 1395 2020-03-12   Yolo California 6113     1      0
## 1656 2020-03-13   Yolo California 6113     2      0
## 1974 2020-03-14   Yolo California 6113     2      0
## 2355 2020-03-15   Yolo California 6113     2      0

```

```

# Calculate new cases & add new column
Yolo$new_cases <- abs(stats::filter(Yolo$cases, filter = c(-1, 1), sides = 1))

# Calculate new deaths & add new column
Yolo$new_deaths <- abs(stats::filter(Yolo$deaths, filter = c(-1, 1), sides = 1))

```

```
head(Yolo, 10)
```

```
##           date county      state fips cases deaths
## 599  2020-03-06   Yolo California 6113     1      0
## 677  2020-03-07   Yolo California 6113     1      0
## 772  2020-03-08   Yolo California 6113     1      0
## 887  2020-03-09   Yolo California 6113     1      0
## 1021 2020-03-10   Yolo California 6113     1      0
## 1188 2020-03-11   Yolo California 6113     1      0
## 1395 2020-03-12   Yolo California 6113     1      0
## 1656 2020-03-13   Yolo California 6113     2      0
## 1974 2020-03-14   Yolo California 6113     2      0
## 2355 2020-03-15   Yolo California 6113     2      0
##           new_cases new_deaths
## 599             NA          NA
## 677              0            0
## 772              0            0
## 887              0            0
## 1021             0            0
## 1188             0            0
## 1395             0            0
## 1656             1            0
## 1974             0            0
## 2355             0            0
```

```
# Calculate 7-day running average for new cases & add new column
```

```
Yolo$avg_new_cases <- stats::filter(Yolo$new_cases, filter = rep(1/7, 7), sides = 1)
```

```
# Calculate 7-day running average for new deaths & add new column
```

```
Yolo$avg_new_deaths <- stats::filter(Yolo$new_deaths, filter = rep(1/7, 7), sides = 1)
```

```
head(Yolo, 10)
```

```
##           date county      state fips cases deaths
## 599  2020-03-06   Yolo California 6113     1      0
## 677  2020-03-07   Yolo California 6113     1      0
## 772  2020-03-08   Yolo California 6113     1      0
## 887  2020-03-09   Yolo California 6113     1      0
## 1021 2020-03-10   Yolo California 6113     1      0
## 1188 2020-03-11   Yolo California 6113     1      0
## 1395 2020-03-12   Yolo California 6113     1      0
## 1656 2020-03-13   Yolo California 6113     2      0
## 1974 2020-03-14   Yolo California 6113     2      0
## 2355 2020-03-15   Yolo California 6113     2      0
##           new_cases new_deaths avg_new_cases avg_new_deaths
## 599             NA          NA             NA             NA
## 677              0            0             NA             NA
## 772              0            0             NA             NA
## 887              0            0             NA             NA
## 1021             0            0             NA             NA
## 1188             0            0             NA             NA
## 1395             0            0             NA             NA
```

```
## 1656      1      0      0.1428571      0
## 1974      0      0      0.1428571      0
## 2355      0      0      0.1428571      0
```

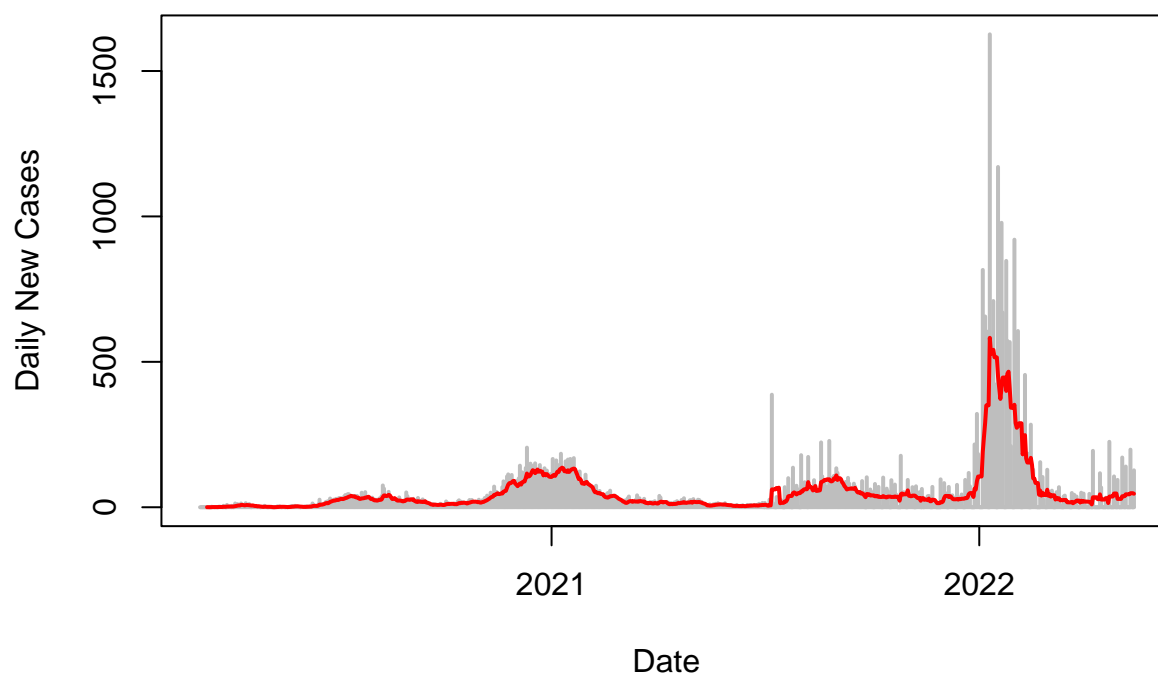
```
# Extract data beginning 2020-03-01
Yolo <- subset(Yolo, date >= as.Date("2020-03-01"))

head(Yolo, 10)
```

```
##      date county      state fips cases deaths
## 599 2020-03-06   Yolo California 6113      1      0
## 677 2020-03-07   Yolo California 6113      1      0
## 772 2020-03-08   Yolo California 6113      1      0
## 887 2020-03-09   Yolo California 6113      1      0
## 1021 2020-03-10   Yolo California 6113      1      0
## 1188 2020-03-11   Yolo California 6113      1      0
## 1395 2020-03-12   Yolo California 6113      1      0
## 1656 2020-03-13   Yolo California 6113      2      0
## 1974 2020-03-14   Yolo California 6113      2      0
## 2355 2020-03-15   Yolo California 6113      2      0
##      new_cases new_deaths avg_new_cases avg_new_deaths
## 599      NA      NA      NA      NA
## 677      0      0      NA      NA
## 772      0      0      NA      NA
## 887      0      0      NA      NA
## 1021     0      0      NA      NA
## 1188     0      0      NA      NA
## 1395     0      0      NA      NA
## 1656     1      0      0.1428571      0
## 1974     0      0      0.1428571      0
## 2355     0      0      0.1428571      0
```

```
# Plot daily new cases & add line for 7-day running average
{plot(new_cases ~ date, type = "h", col = "gray", data = Yolo, main = 'Yolo County Daily New Covid-19 C
lines(Yolo$date, Yolo$avg_new_cases, col = "red", lwd = "2")}
```

## Yolo County Daily New Covid-19 Cases from 2020-03-06

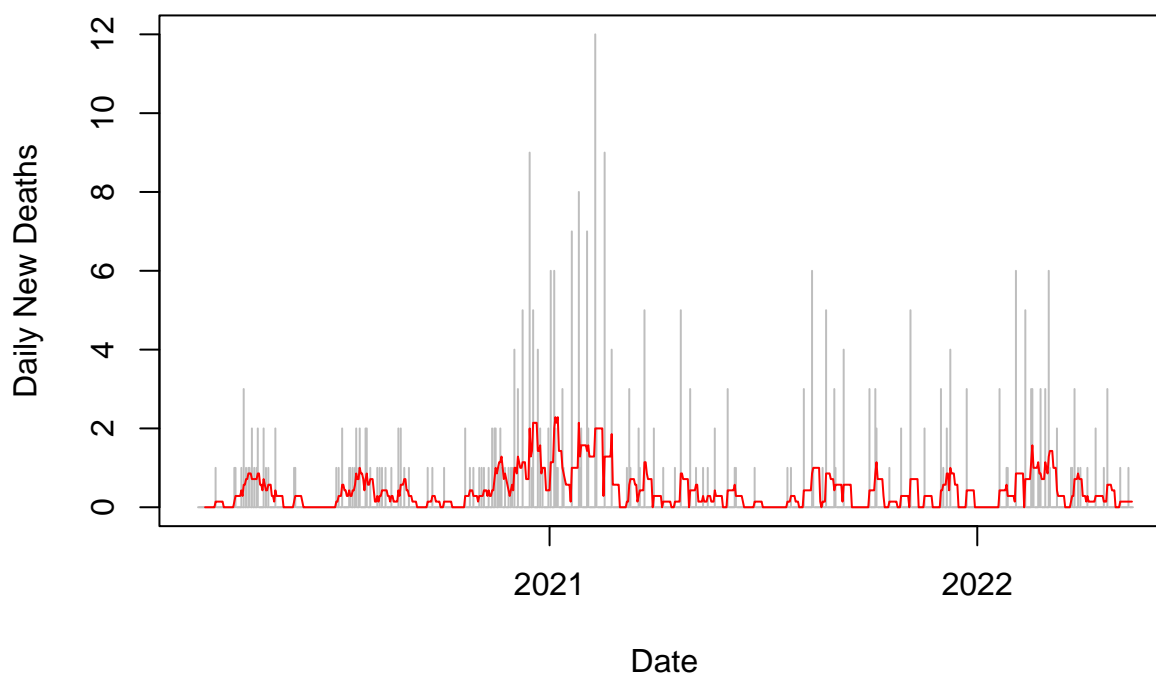


### Analysis

This graph illustrates the number of daily new cases of Covid-19 in Yolo County, from 2020-03-06 to 2022-05-13. The gray portion of the graph indicates the daily new cases, and the red line shows the seven-day running average from the previous seven days. We can see two main peaks in the number of daily new cases, with the main peaks occurring at the end of 2020, and the beginning of 2022. It makes sense that the number of daily new cases is significantly lower than that of California and the US, since Yolo County holds a small fraction of the relative populations. It is interesting to note that the number of daily new cases in Yolo County tends to follow a similar trend to California. The peaks and declines tend to occur around the same time periods at the end of 2020 and the beginning of 2022.

```
# Plot daily new deaths & add line for 7-day running average
{plot(new_deaths ~ date, type = "h", col = "gray", data = Yolo, main = 'Yolo County Daily New Covid-19 I
lines(Yolo$date, Yolo$avg_new_deaths, col = "red")}
```

## Yolo County Daily New Covid-19 Deaths from 2020-03-06



### Analysis

This graph illustrates the number of daily new deaths due to Covid-19 in Yolo County, from 2020-03-06 to 2022-05-13. The gray portion of the graph indicates the daily new deaths, and the red line shows the seven-day running average from the previous seven days. This graph is the most unique, relative to the previous graphs for the number of cases and deaths in the US and California. The key difference in this graph is that the number of daily new deaths appears to be significantly inconsistent. The peaks and declines occur over very short time periods, with fairly unclear patterns. It is worth noting that there is an extended peak in daily new deaths near the end of 2020 and beginning of 2021. Considering that the daily new deaths ranges from 0 to 12 for Yolo County, and 0 to 2,000 for California, it is safe to assume that Yolo County only accounts for a small fraction of the deaths in California. This could be due to the fact that Yolo County has a significantly smaller population, or simply that Covid-19 had a relatively smaller impact on Yolo County.

**What do you notice when comparing the plot of daily new cases in Yolo county to the analogous plot for the state of California as a whole? What might explain what you are seeing?**

When we compare the plot of daily new cases in Yolo county to the analogous plot for the state of California as a whole, both Yolo County and California show similar trends in the time of peaks and declines of daily new cases. The prominent peaks occurred in the 2021 and 2022 for both Yolo County and California. In terms of scale, the number of daily new cases in Yolo County is significantly higher lower than in California as a whole. It is because Yolo County has a smaller population compared to the California. We expect that the trends between Yolo County and California are caused by statewide policies, public health measures, and other factors spreading Covid-19. These causes result in similar patterns of daily new cases.