

Covid

Ben Lewis

2025-01-05

```
library(stringr)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v purrr      1.0.2
## v forcats    1.0.0      v readr      2.1.5
## v ggplot2    3.5.1      v tibble     3.2.1
## v lubridate  1.9.4      v tidyr      1.3.1
## -- 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
```

```
url_in <- "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_data"

file_names <- c("time_series_covid19_confirmed_global.csv", "time_series_covid19_deaths_global.csv", "time_series_covid19_recovered_global.csv")
```

```
urls <- str_c(url_in, file_names)
```

```
global_cases <- read_csv(urls[1])
```

```
## Rows: 289 Columns: 1147
## -- Column specification -----
## Delimiter: ","
## chr      (2): Province/State, Country/Region
## dbl (1145): Lat, Long, 1/22/20, 1/23/20, 1/24/20, 1/25/20, 1/26/20, 1/27/20,...
##
## 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.
```

```
global_deaths <- read_csv(urls[2])
```

```
## Rows: 289 Columns: 1147
## -- Column specification -----
## Delimiter: ","
## chr      (2): Province/State, Country/Region
## dbl (1145): Lat, Long, 1/22/20, 1/23/20, 1/24/20, 1/25/20, 1/26/20, 1/27/20,...
##
## 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.
```

```
us_cases <- read_csv(urls[3])
```

```
## Rows: 3342 Columns: 1154
## -- Column specification -----
## Delimiter: ","
## chr      (6): iso2, iso3, Admin2, Province_State, Country_Region, Combined_Key
## dbl (1148): UID, code3, FIPS, Lat, Long_, 1/22/20, 1/23/20, 1/24/20, 1/25/20...
##
## 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.
```

```
us_deaths <- read_csv(urls[4])
```

```
## Rows: 3342 Columns: 1155
## -- Column specification -----
## Delimiter: ","
## chr      (6): iso2, iso3, Admin2, Province_State, Country_Region, Combined_Key
## dbl (1149): UID, code3, FIPS, Lat, Long_, Population, 1/22/20, 1/23/20, 1/24/20...
##
## 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.
```

```
global_cases <- global_cases %>% pivot_longer(cols = -c(`Province/State`, `Country/Region`, `Lat`, `Long_`))
```

```
global_deaths <- global_deaths %>% pivot_longer(cols = -c(`Province/State`, `Country/Region`, `Lat`, `Long_`))
```

```
global <- global_cases %>% full_join(global_deaths) %>% rename(Country_Region = `Country/Region`, Province_State = `Province/State`)
```

```
## Joining with 'by = join_by('Province/State', 'Country/Region', date)'
```

```
global <- global %>% filter(cases > 0)
```

```
us_cases <- us_cases %>% pivot_longer(cols = -(UID:Combined_Key),
                                     names_to = "date",
                                     values_to = "cases") %>%
```

```
  select(Admin2:cases) %>%
  mutate(date = mdy(date)) %>%
  select(-c(Lat, Long_))
```

```
us_deaths <- us_deaths %>%
  pivot_longer(cols = -(UID:Population),
               names_to = "date",
               values_to = "deaths") %>%
```

```
  select(Admin2:deaths) %>%
  mutate(date = mdy(date)) %>%
  select(-c(Lat, Long_))
```

```
us <- us_cases %>%
  full_join(us_deaths)
```

```
## Joining with 'by = join_by(Admin2, Province_State, Country_Region,
## Combined_Key, date)'
```

```

global <- global %>%
  unite("Combined_Key",
        c(Province_State, Country_Region),
        sep = ", ",
        na.rm = TRUE,
        remove = FALSE)

uid_lookup_url <- "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/"

uid <- read_csv(uid_lookup_url) %>%
  select(-c(Lat, Long_, Combined_Key, code3, iso2, iso3, Admin2))

```

```

## Rows: 4321 Columns: 12
## -- Column specification -----
## Delimiter: ","
## chr (7): iso2, iso3, FIPS, Admin2, Province_State, Country_Region, Combined_Key
## dbl (5): UID, code3, Lat, Long_, Population
##
## 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.

```

```

global <- global %>%
  left_join(uid, by = c("Province_State", "Country_Region")) %>%
  select(-c(UID, FIPS)) %>%
  select(Province_State, Country_Region, date,
        cases, deaths, Population,
        Combined_Key)

us_by_state <- us %>%
  group_by(Province_State, Country_Region, date) %>%
  summarize(cases = sum(cases), deaths = sum(deaths),
            Population = sum(Population)) %>%
  mutate(deaths_per_mill = deaths * 1000000 / Population) %>%
  select(Province_State, Country_Region, date,
        cases, deaths, deaths_per_mill, Population) %>%
  ungroup()

```

```

## 'summarise()' has grouped output by 'Province_State', 'Country_Region'. You can
## override using the '.groups' argument.

```

```

us_totals <- us_by_state %>%
  group_by(Country_Region, date) %>%
  summarize(cases = sum(cases), deaths = sum(deaths),
            Population = sum(Population)) %>%
  mutate(deaths_per_mill = deaths * 1000000 / Population) %>%
  select(Country_Region, date,
        cases, deaths, deaths_per_mill, Population) %>%
  ungroup()

```

```

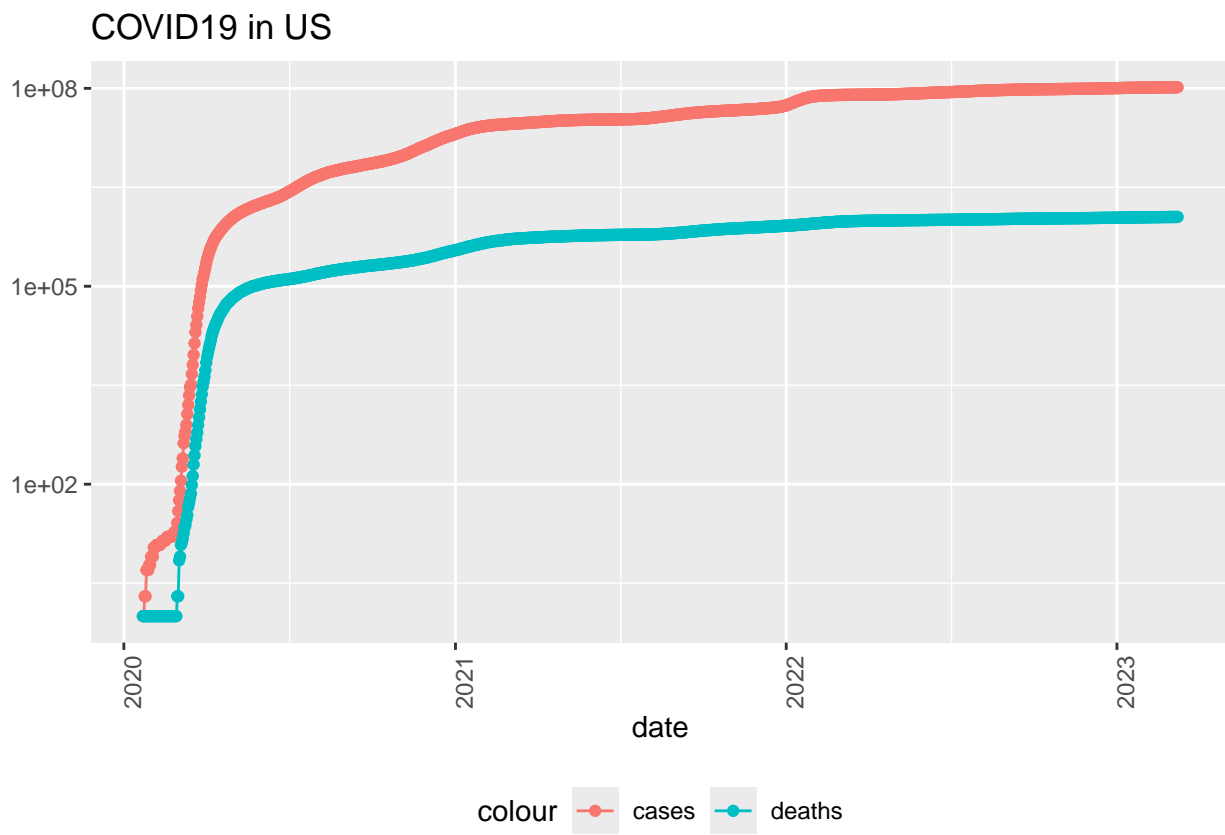
## 'summarise()' has grouped output by 'Country_Region'. You can override using
## the '.groups' argument.

```

```
tail(us_totals)
```

```
## # A tibble: 6 x 6
##   Country_Region date           cases  deaths deaths_per_mill Population
##   <chr>          <date>       <dbl>   <dbl>      <dbl>      <dbl>
## 1 US            2023-03-04 103650837 1122172      3371.  332875137
## 2 US            2023-03-05 103646975 1122134      3371.  332875137
## 3 US            2023-03-06 103655539 1122181      3371.  332875137
## 4 US            2023-03-07 103690910 1122516      3372.  332875137
## 5 US            2023-03-08 103755771 1123246      3374.  332875137
## 6 US            2023-03-09 103802702 1123836      3376.  332875137
```

```
us_totals %>%
  filter(cases > 0) %>%
  ggplot(aes(x = date, y = cases)) +
  geom_line(aes(color = "cases")) +
  geom_point(aes(color = "cases")) +
  geom_line(aes(y = deaths, color = "deaths")) +
  geom_point(aes(y = deaths, color = "deaths")) +
  scale_y_log10() +
  theme(legend.position = "bottom",
        axis.text.x = element_text(angle = 90)) +
  labs(title = "COVID19 in US", y = NULL)
```



```

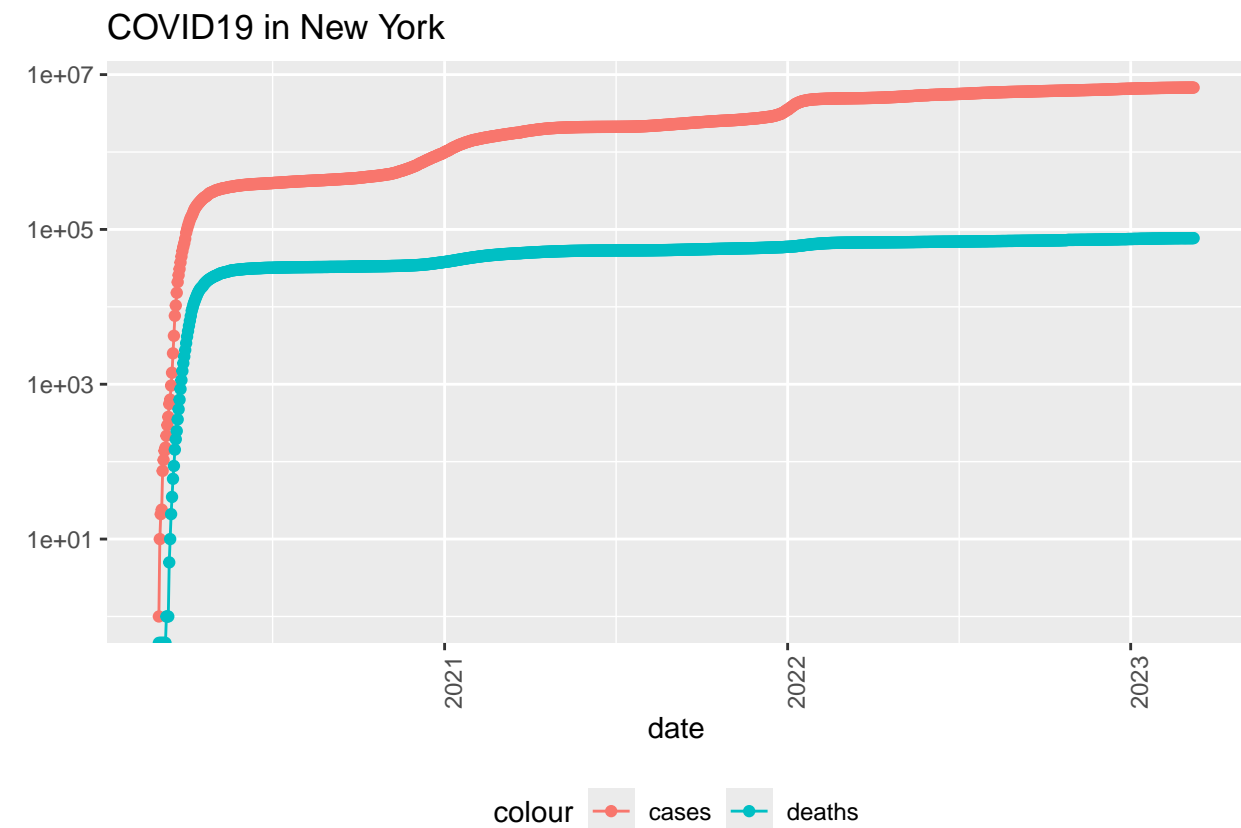
state <- "New York"
us_by_state %>%
  filter(Province_State == state) %>%
  filter(cases > 0) %>%
  ggplot(aes(x = date, y = cases)) +
  geom_line(aes(color = "cases")) +
  geom_point(aes(color = "cases")) +
  geom_line(aes(y = deaths, color = "deaths")) +
  geom_point(aes(y = deaths, color = "deaths")) +
  scale_y_log10() +
  theme(legend.position = "bottom", axis.text.x = element_text(angle = 90)) +
  labs(title = str_c("COVID19 in ", state), y = NULL)

```

```

## Warning in scale_y_log10(): log-10 transformation introduced infinite values.
## log-10 transformation introduced infinite values.

```



```
max(us_totals$date)
```

```
## [1] "2023-03-09"
```

```
max(us_totals$deaths)
```

```
## [1] 1123836
```

```
us_by_state <- us_by_state %>%
  mutate(new_cases = cases - lag(cases),
         new_deaths = deaths - lag(deaths))

us_totals <- us_totals %>%
  mutate(new_cases = cases - lag(cases),
         new_deaths = deaths - lag(deaths))

tail(us_totals %>% select(new_cases, new_deaths, everything()))
```

```
## # A tibble: 6 x 8
##   new_cases new_deaths Country_Region date      cases deaths deaths_per_mill
##   <dbl>     <dbl> <chr>          <date>    <dbl> <dbl>      <dbl>
## 1      2147         7 US            2023-03-04 1.04e8 1.12e6    3371.
## 2     -3862        -38 US            2023-03-05 1.04e8 1.12e6    3371.
## 3      8564         47 US            2023-03-06 1.04e8 1.12e6    3371.
## 4     35371        335 US            2023-03-07 1.04e8 1.12e6    3372.
## 5     64861        730 US            2023-03-08 1.04e8 1.12e6    3374.
## 6     46931        590 US            2023-03-09 1.04e8 1.12e6    3376.
## # i 1 more variable: Population <dbl>
```

```
us_totals %>%
  ggplot(aes(x = date, y = new_cases)) +
  geom_line(aes(color = "new_cases")) +
  geom_point(aes(color = "new_cases")) +
  geom_line(aes(y = new_deaths, color = "new_deaths")) +
  geom_point(aes(y = new_deaths, color = "new_deaths")) +
  scale_y_log10() +
  theme(legend.position = "bottom",
        axis.text.x = element_text(angle = 90)) +
  labs(title = "COVID19 in US", y = NULL)
```

```
## Warning in transformation$transform(x): NaNs produced
## Warning in transformation$transform(x): log-10 transformation introduced
## infinite values.

## Warning in transformation$transform(x): NaNs produced

## Warning in scale_y_log10(): log-10 transformation introduced infinite values.

## Warning in transformation$transform(x): NaNs produced

## Warning in scale_y_log10(): log-10 transformation introduced infinite values.

## Warning in transformation$transform(x): NaNs produced

## Warning in scale_y_log10(): log-10 transformation introduced infinite values.

## Warning: Removed 1 row containing missing values or values outside the scale range
## ('geom_line()').
```

```
## Warning: Removed 2 rows containing missing values or values outside the scale range
## ('geom_point()').
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## ('geom_line()').
```

```
## Warning: Removed 4 rows containing missing values or values outside the scale range
## ('geom_point()').
```

COVID19 in US



```
## R Markdown
```

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

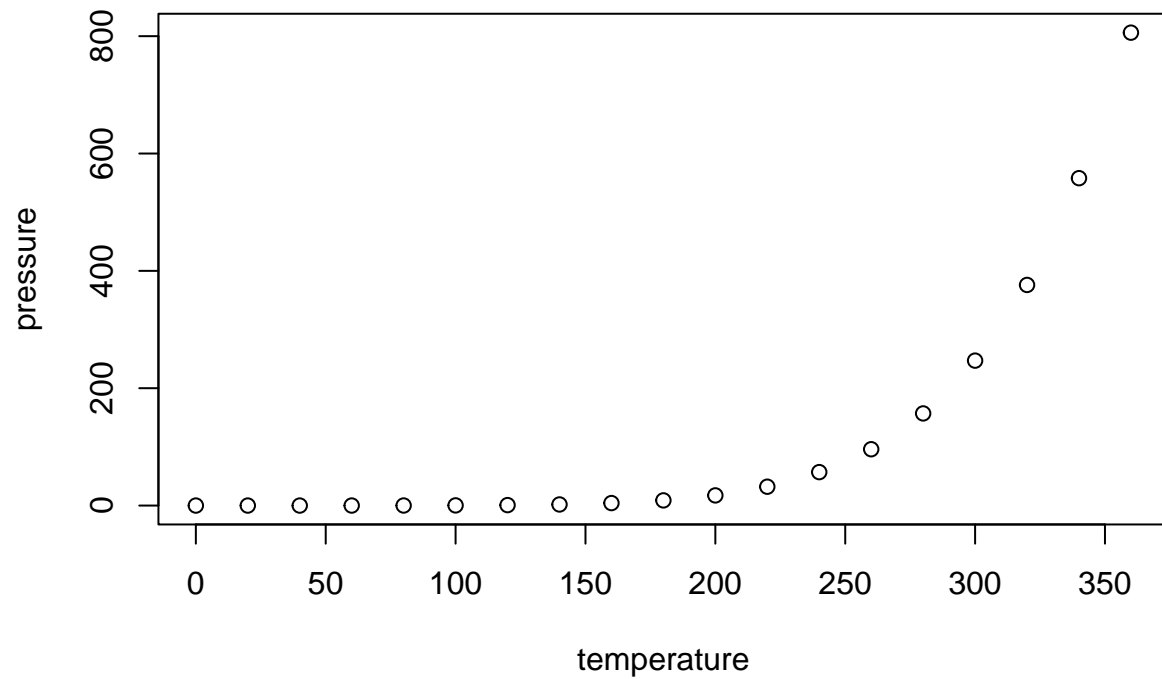
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   : 2.00
##  1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##  Mean   :15.4    Mean   : 42.98
##  3rd Qu.:19.0    3rd Qu.: 56.00
##  Max.   :25.0    Max.   :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.