Covid

Ben Lewis

2025-01-05

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

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

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

urls <- str_c(url_in, file_names)

global_cases <- read_csv(urls[1])
global_deaths <- read_csv(urls[2])
us_cases <- read_csv(urls[3])
us_deaths <- read_csv(urls[4])
```

```
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`, `L

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

```
## 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/U

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() %>%
  filter(cases > 0)
```

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

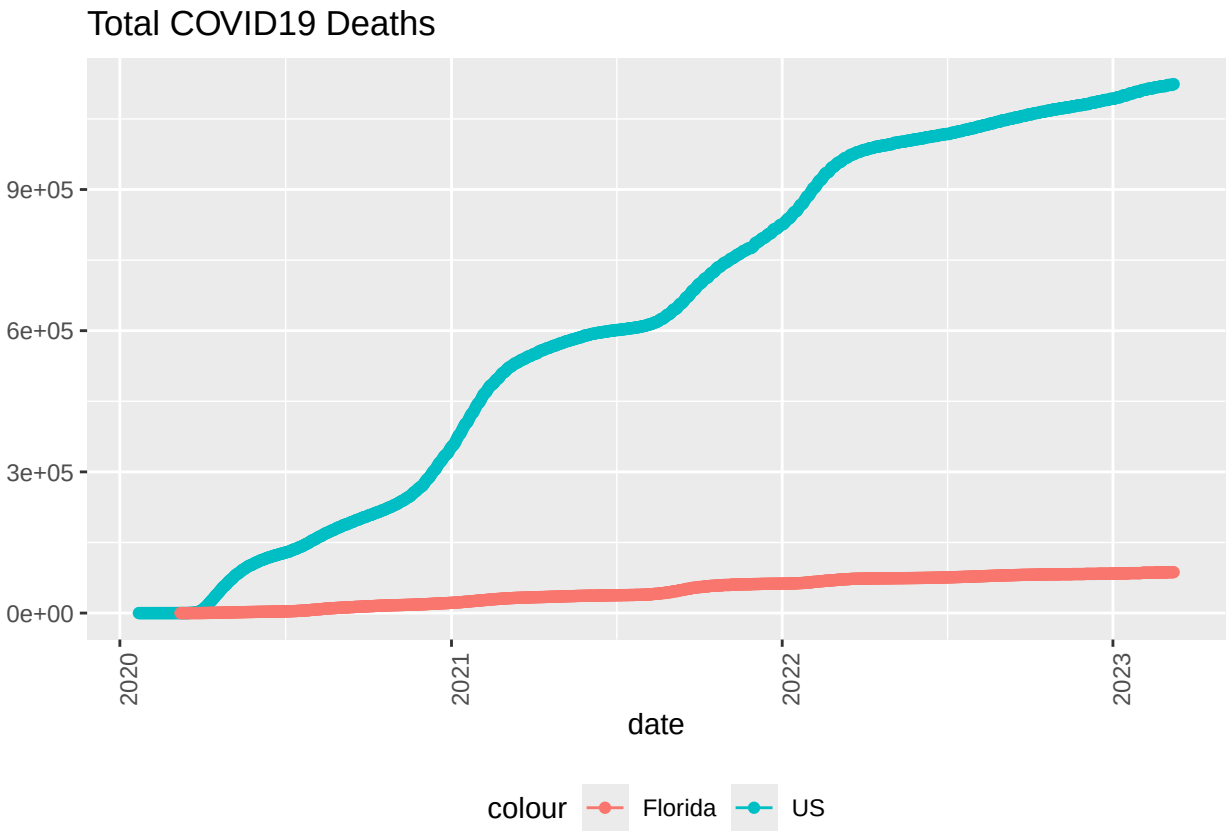
```
tail(us_totals)
```

```
## # 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
```

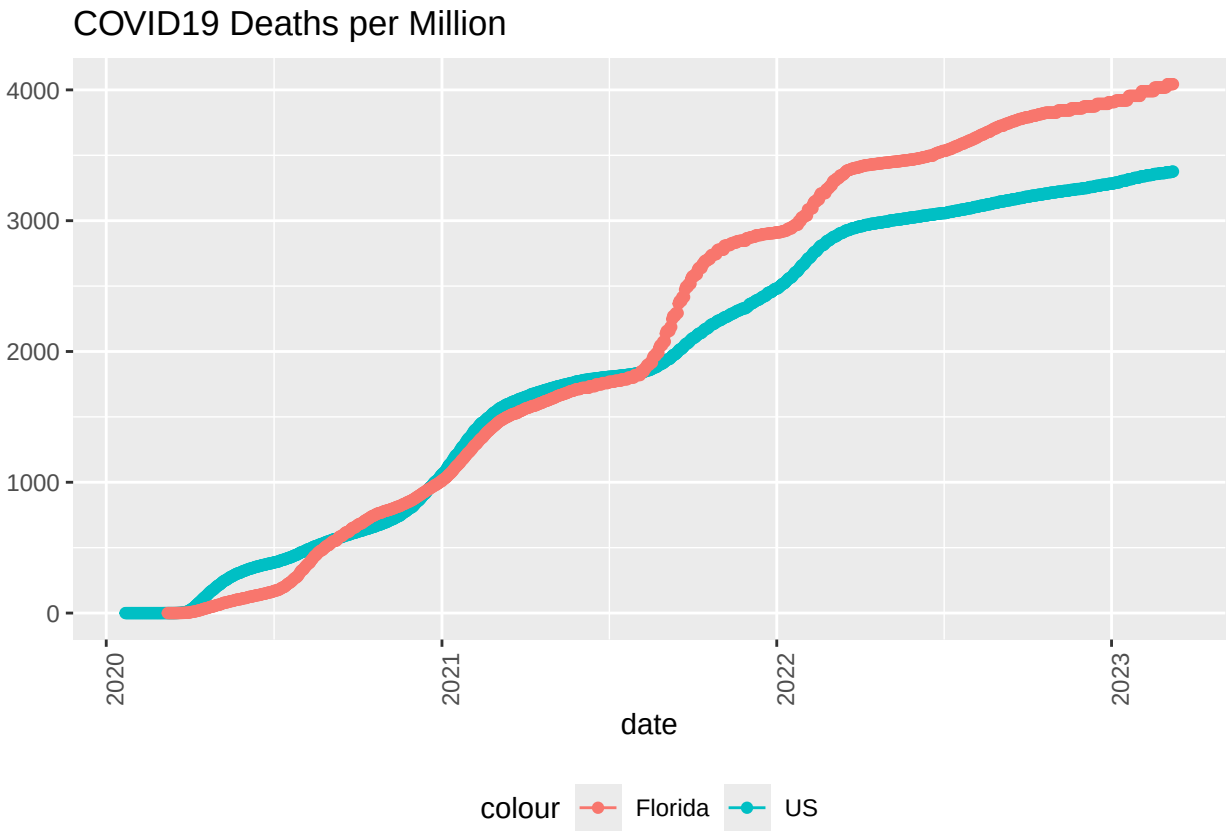
```
state <- "Florida"
state_totals <- us_by_state %>%
  filter(Province_State == state) %>%
  filter(cases > 0 & deaths > 0)

tail(state_totals)
```

```
## # A tibble: 6 x 7
##   Province_State Country_Region date         cases deaths deaths_per_mill
##   <chr>          <chr>          <date>       <dbl>  <dbl>           <dbl>
## 1 Florida        US             2023-03-04 7574590  86850           4044.
## 2 Florida        US             2023-03-05 7574590  86850           4044.
## 3 Florida        US             2023-03-06 7574590  86850           4044.
## 4 Florida        US             2023-03-07 7574590  86850           4044.
## 5 Florida        US             2023-03-08 7574590  86850           4044.
## 6 Florida        US             2023-03-09 7574590  86850           4044.
## # i 1 more variable: Population <dbl>
```

```
  ggplot() +
    geom_line(data = us_totals, aes(x = date, y = deaths, color = "US")) +
    geom_point(data = us_totals, aes(x = date, y = deaths, color = "US")) +
    geom_line(data = state_totals, aes(x = date, y = deaths, color = state)) +
    geom_point(data = state_totals, aes(x = date, y = deaths, color = state)) +
    theme(legend.position = "bottom", axis.text.x = element_text(angle = 90)) +
    labs(title = "Total COVID19 Deaths", y = NULL)
```




```
ggplot() +
  geom_line(data = us_totals, aes(x = date, y = deaths_per_mill, color = "US")) +
  geom_point(data = us_totals, aes(x = date, y = deaths_per_mill, color = "US")) +
  geom_line(data = state_totals, aes(x = date, y = deaths_per_mill, color = state)) +
  geom_point(data = state_totals, aes(x = date, y = deaths_per_mill, color = state)) +
  theme(legend.position = "bottom", axis.text.x = element_text(angle = 90)) +
  labs(title = "COVID19 Deaths per Million", y = NULL)
```



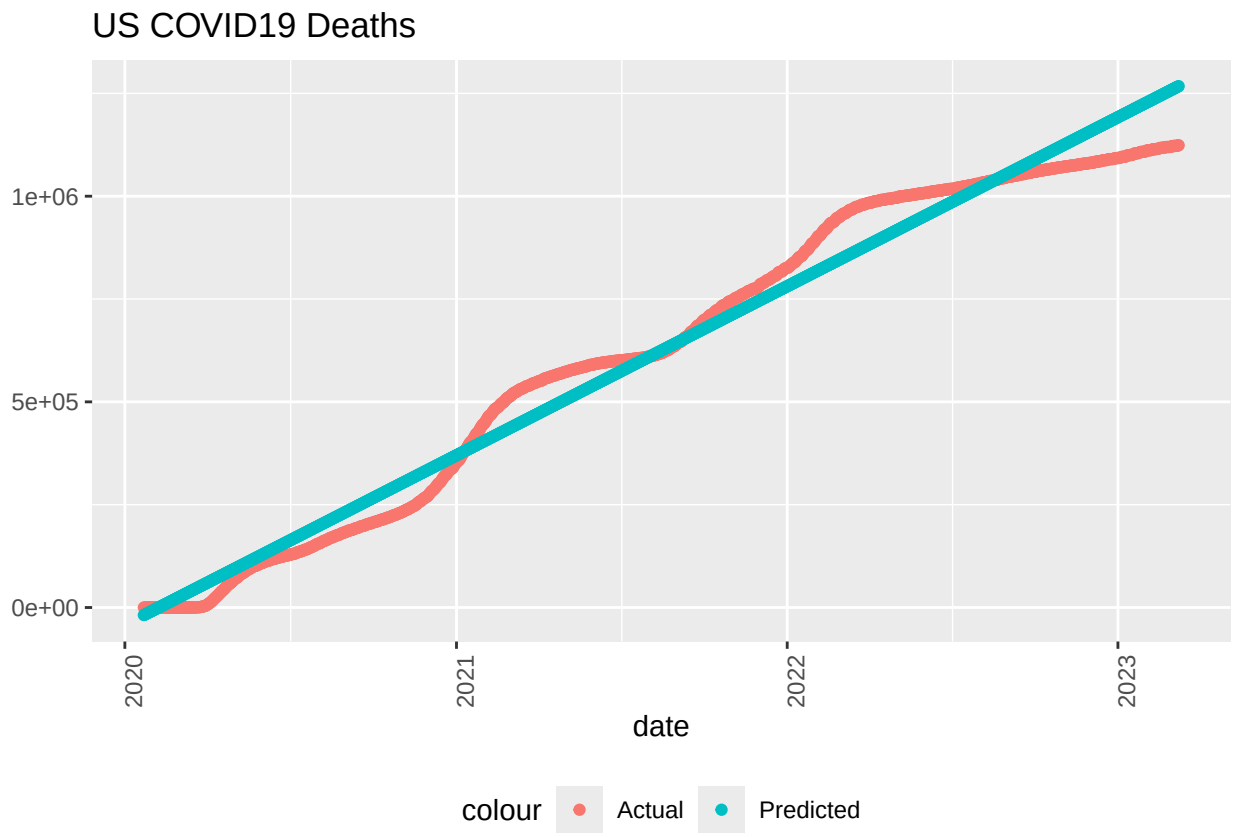

```
us_totals$days_since_start <- as.numeric(us_totals$date - min(us_totals$date))

us_mod <- lm(deaths ~ days_since_start, data = us_totals)
summary(us_mod)
```

```
##
## Call:
## lm(formula = deaths ~ days_since_start, data = us_totals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -143657  -44489    -955   46232  107933
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -18367.12    3521.40  -5.216 2.17e-07 ***
## days_since_start   1125.97       5.34 210.869  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 59570 on 1141 degrees of freedom
## Multiple R-squared:  0.975,  Adjusted R-squared:  0.975
## F-statistic: 4.447e+04 on 1 and 1141 DF,  p-value: < 2.2e-16
```

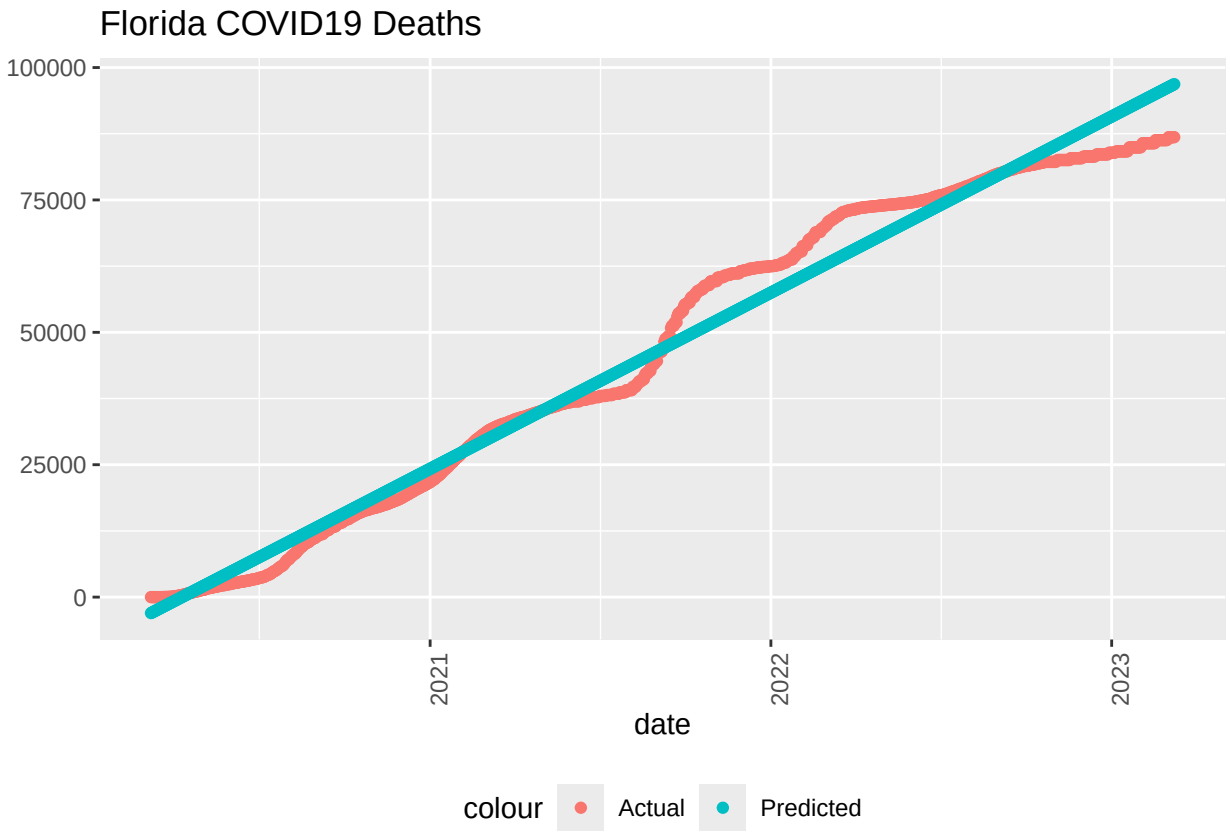
```r
state_totals$days_since_start <- as.numeric(state_totals$date - min(state_totals$date))
state_mod <- lm(deaths ~ days_since_start, data = state_totals)
summary(state_mod)
```

```
##
## Call:
## lm(formula = deaths ~ days_since_start, data = state_totals)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10007.5  -2615.3   -532.2   2630.9   8194.1
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -3010.0300   258.2885  -11.65   <2e-16 ***
## days_since_start    91.1200     0.4081  223.28   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4280 on 1095 degrees of freedom
## Multiple R-squared:  0.9785, Adjusted R-squared:  0.9785
## F-statistic: 4.986e+04 on 1 and 1095 DF,  p-value: < 2.2e-16
```

```r
us_totals$pred_deaths <- predict(us_mod, newdata = us_totals)
ggplot(us_totals, aes(x = date)) +
  geom_point(aes(y = deaths, color = "Actual")) +
  geom_point(aes(y = pred_deaths, color = "Predicted")) +
  theme(legend.position = "bottom", axis.text.x = element_text(angle = 90)) +
  labs(title = "US COVID19 Deaths", y = NULL)
```




```
state_totals$pred_deaths <- predict(state_mod, newdata = state_totals)
ggplot(state_totals, aes(x = date)) +
  geom_point(aes(y = deaths, color = "Actual")) +
  geom_point(aes(y = pred_deaths, color = "Predicted")) +
  theme(legend.position = "bottom", axis.text.x = element_text(angle = 90)) +
  labs(title = paste(state, "COVID19 Deaths"), y = NULL)
```


Florida COVID19 Deaths
100000
75000
50000
25000
0
2021
2022
2023
date
colour
Actual
Predicted