Covid

Ben Lewis

2025-01-05

Overview

Florida gained a reputation for having relatively lax COVID-19 rules compared to other states, particularly in contrast to more restrictive states like California and New York. Under Governor Ron DeSantis, Florida prioritized keeping businesses, schools, and public spaces open while resisting strict mask mandates, vaccine requirements, and prolonged lockdowns. We will explore the outcomes of COVID-19 relative the country as a whole.

Set up

Import libraries

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

Download data

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

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

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])
```

Clean and format data

```
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

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)

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))

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()

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)

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

```
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
```

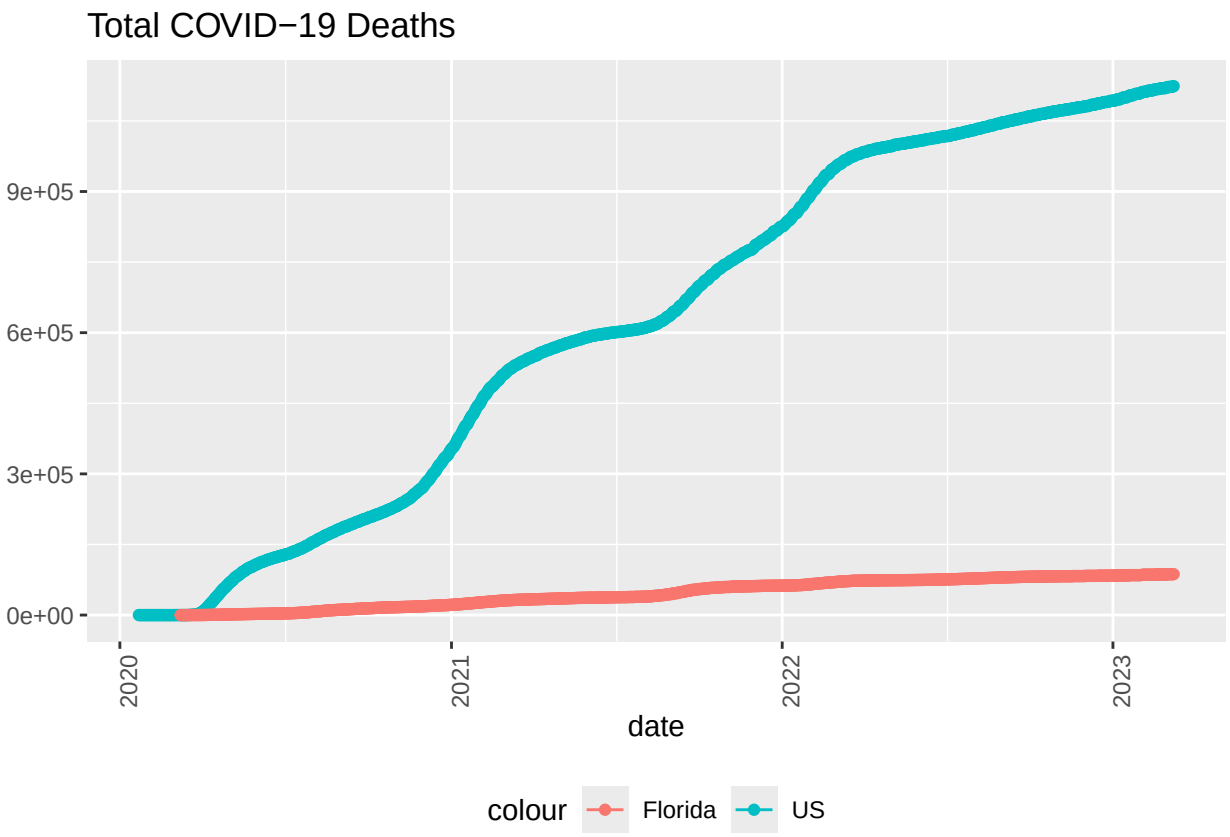
```
tail(state_totals)
```

```
## # 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>
```

Data Analysis

Let's visualize the data by viewing the total number of COVID-19 deaths in the US and Florida.

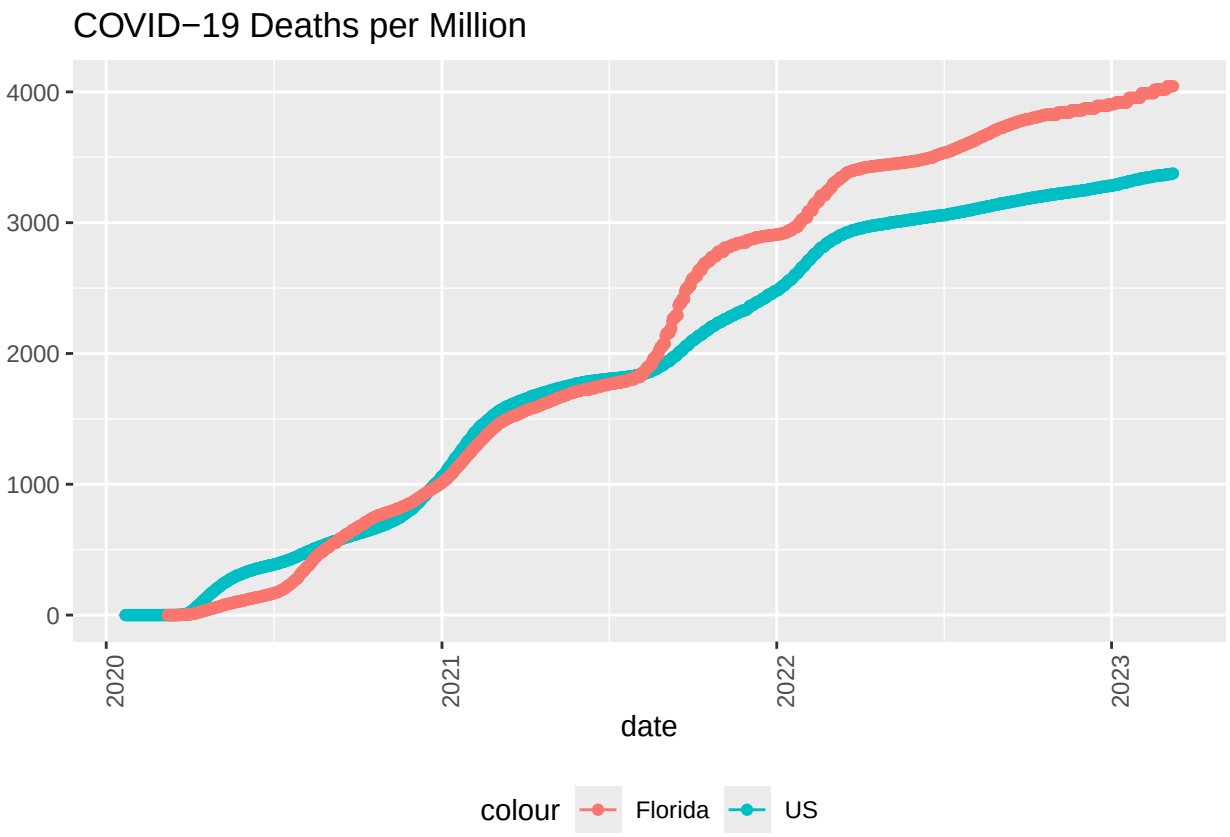
```
  ggplot() +
    geom_line(data = us_totals, aes(x = date, y = deaths, color = "US")) +
    geom_point(data = us_totals, aes(x = date, y = deaths, color = "US")) +
    geom_line(data = state_totals, aes(x = date, y = deaths, color = state)) +
    geom_point(data = state_totals, aes(x = date, y = deaths, color = state)) +
    theme(legend.position = "bottom", axis.text.x = element_text(angle = 90)) +
    labs(title = "Total COVID-19 Deaths", y = NULL)
```



We see that Florida makes up a relatively small percentage of the total deaths in the US. Also note that both Florida and the US deaths grew at an approximately linear rate.

Let's now normalize the data by comparing the deaths per million.

```
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 = "COVID-19 Deaths per Million", y = NULL)
```



The graph shows that deaths were pretty similar between the summers of 2020 and 2021. When the Delta variant surged in August-September 2021 the deaths grew at a higher rate than the national average, and the total deaths remained elevated to the end of time period captured in the graph (2023).

Since it appears the deaths grew at an approximately linear rate let us develop linear regression models for both the US and Florida and and compare them.

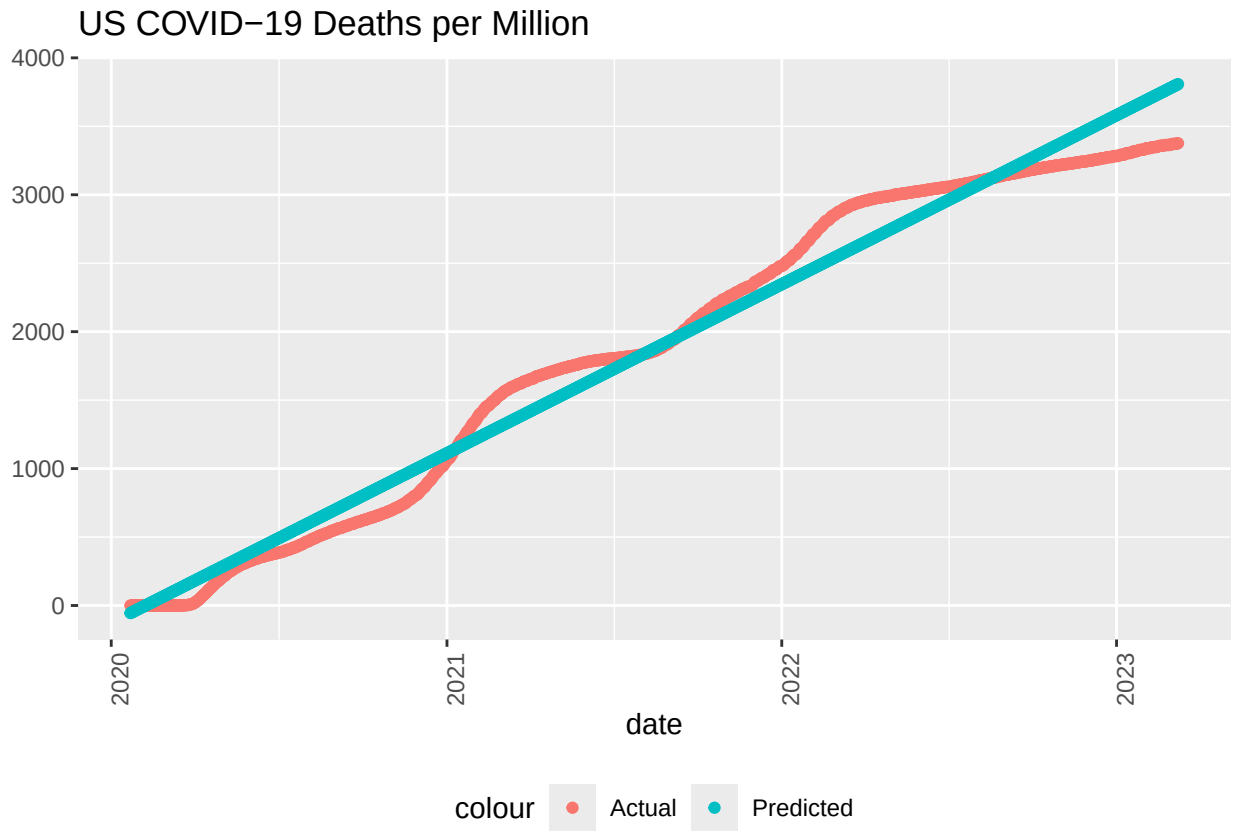
```
us_totals$days_since_start <- as.numeric(us_totals$date - min(us_totals$date))

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

```
##
## Call:
## lm(formula = deaths_per_mill ~ days_since_start, data = us_totals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -431.56 -133.65   -2.87  138.89  324.25
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -55.17722   10.57874  -5.216 2.17e-07 ***
## days_since_start   3.38256    0.01604 210.869  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 178.9 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
```

```
us_totals$pred_deaths <- predict(us_mod, newdata = us_totals)
ggplot(us_totals, aes(x = date)) +
  geom_point(aes(y = deaths_per_mill, color = "Actual")) +
  geom_point(aes(y = pred_deaths, color = "Predicted")) +
  theme(legend.position = "bottom", axis.text.x = element_text(angle = 90)) +
  labs(title = "US COVID-19 Deaths per Million", y = NULL)
```

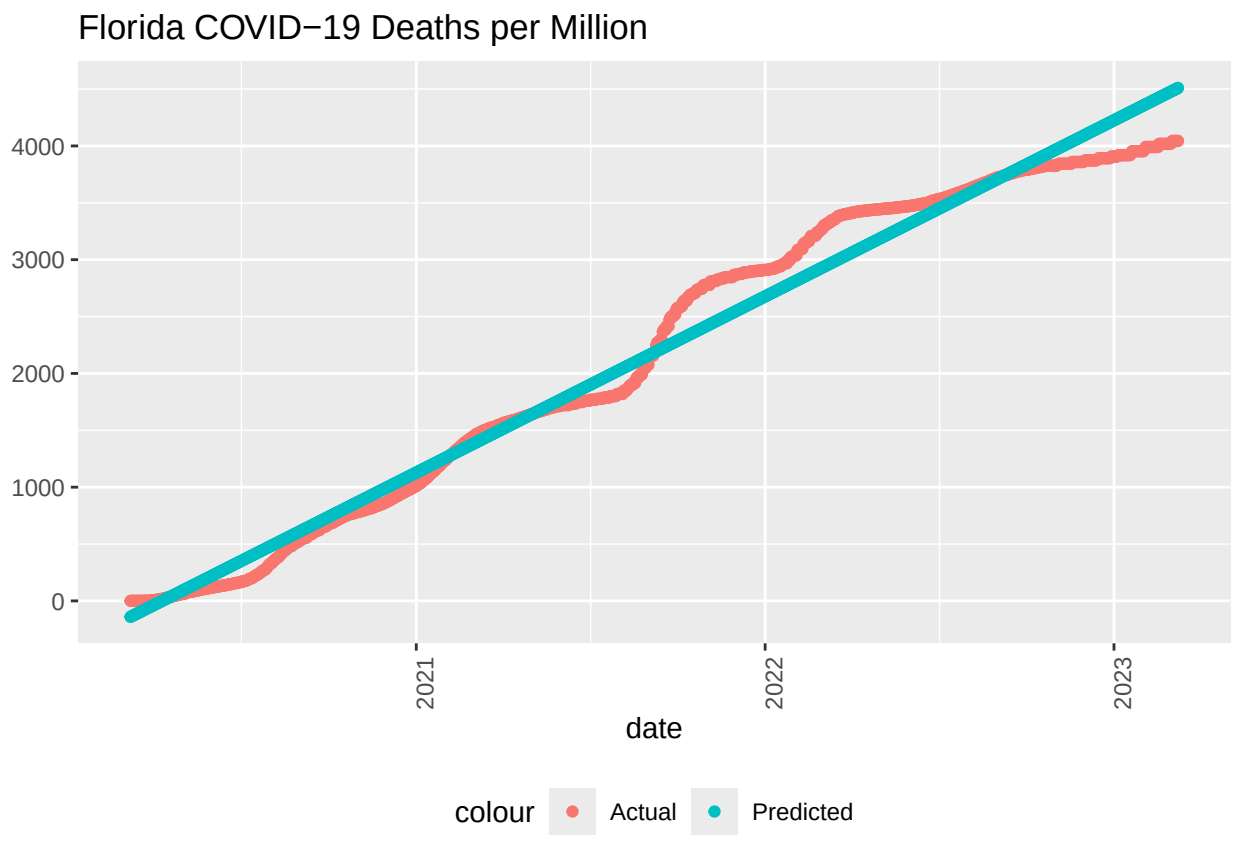



```
state_totals$days_since_start <- as.numeric(state_totals$date - min(state_totals$date))
state_mod <- lm(deaths_per_mill ~ days_since_start, data = state_totals)
summary(state_mod)
```

```
##
## Call:
## lm(formula = deaths_per_mill ~ days_since_start, data = state_totals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -465.95 -121.77  -24.78  122.49  381.52
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
```

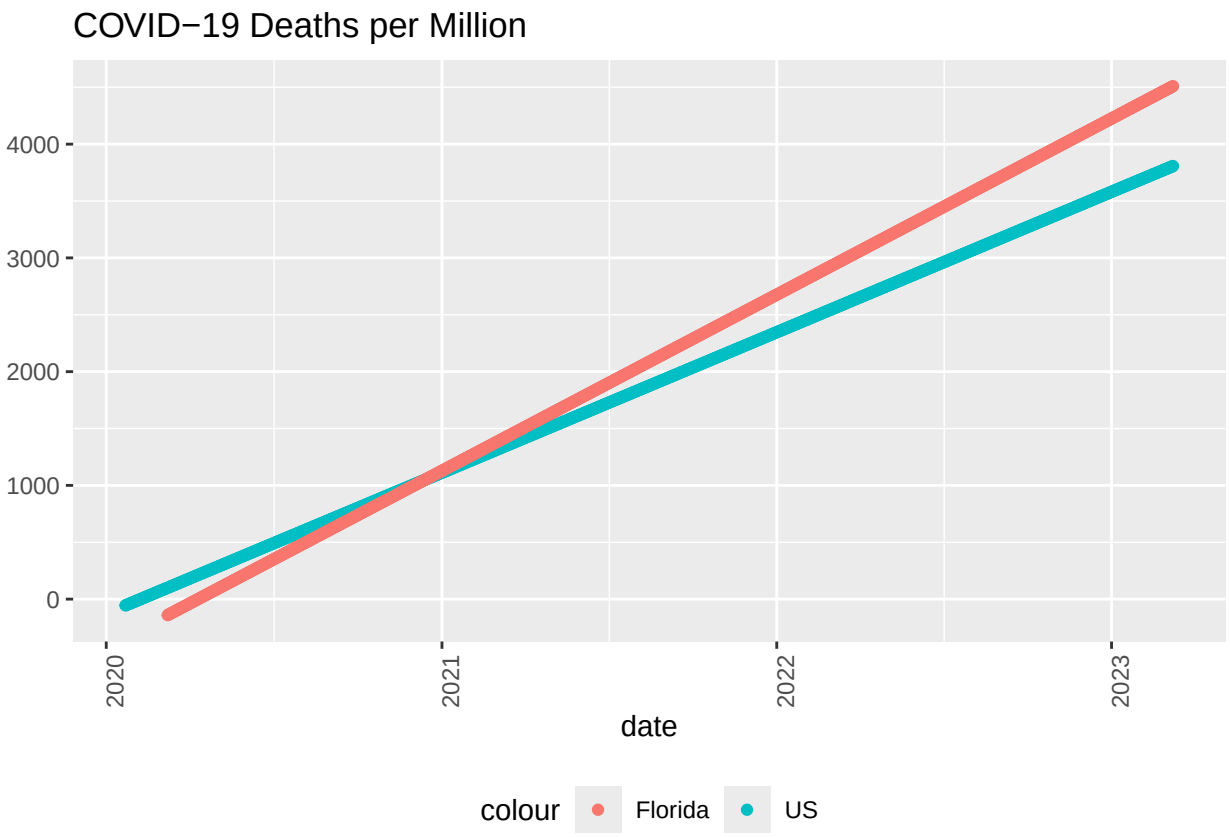
```
## (Intercept)      -140.147     12.026  -11.65   <2e-16 ***
## days_since_start    4.243      0.019  223.28   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 199.3 on 1095 degrees of freedom
## Multiple R-squared:  0.9785, Adjusted R-squared:  0.9785
## F-statistic: 4.986e+04 on 1 and 1095 DF,  p-value: < 2.2e-16
```

```r
state_totals$pred_deaths <- predict(state_mod, newdata = state_totals)
ggplot(state_totals, aes(x = date)) +
  geom_point(aes(y = deaths_per_mill, 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, "COVID-19 Deaths per Million"), y = NULL)
```



Now let us compare the linear regressions.

```r
ggplot() +
  geom_point(data = us_totals, aes(x = date, y = pred_deaths, color = "US")) +
  geom_point(data = state_totals, aes(x = date, y = pred_deaths, color = state)) +
  theme(legend.position = "bottom", axis.text.x = element_text(angle = 90)) +
  labs(title = "COVID-19 Deaths per Million", y = NULL)
```



The linear regression models indicate that while the first deaths in the US occurred outside of Florida, deaths grew at a higher rate in Florida. It is important to note that in actuality deaths did not grow at an perfectly linear rate. There were surges in January 2021, the Delta Variant surge in August-September 2021, and the Omicron wave in January 2022.

Sources of Possible Bias

Florida's older population means it naturally had a higher risk of severe outcomes, making direct comparisons to the national average potentially misleading. Using linear regression assumes a steady increase in deaths, whereas real-world trends were shaped by variant-driven surges, vaccination rollouts, and seasonal effects.

Overall, while the findings indicate a higher rate of death growth in Florida, these results should be interpreted with caution. Adjusting for demographic differences, incorporating non-linear modeling, and exploring additional data sources would provide a more comprehensive understanding of Florida's COVID-19 outcomes.