

COVID19

```
url_in <- "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series.csv"
file_names <- c("time_series_covid19_confirmed_US.csv", "time_series_covid19_confirmed_global.csv", "time_series_covid19_deaths_US.csv")
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(UID:Combined_Key),
               names_to = "date",
               values_to = "cases") %>%
  select(Admin2:cases) %>%
  select(-c(Lat,Long_))

global_deaths <- global_deaths %>%
  pivot_longer(cols = -c(`Province/State`,
                        `Country/Region`,
                        Lat, Long),
               names_to = "date",
               values_to = "cases") %>%
  select(-c(Lat,Long))

US_cases <- US_cases %>%
  pivot_longer(cols = -c(UID:Combined_Key),
               names_to = "date",
               values_to = "cases") %>%
  select(Admin2:cases) %>%
  select(-c(Lat,Long_))

US_deaths <- US_deaths %>%
  pivot_longer(cols = -c(`Province/State`,
                        `Country/Region`,
                        Lat, Long),
               names_to = "date",
               values_to = "cases") %>%
  select(-c(Lat,Long))

US_cases$date <- as.Date(US_cases$date, format = "%m/%d/%y")

US_cases_2021_by_states <- US_cases %>%
  filter(year(date) == 2021) %>%
  group_by(Province_State, date) %>%
```

```

summarize(cases = sum(cases)) %>%
select(Province_State, date, cases) %>%
ungroup()

US_cases_2022_by_states <- US_cases %>%
filter(year(date) == 2022) %>%
group_by(Province_State, date) %>%
summarize(cases = sum(cases)) %>%
select(Province_State, date, cases) %>%
ungroup()

US_cases_by_states <- US_cases %>%
filter(year(date) != "Population") %>%
group_by(Province_State, date) %>%
summarize(cases = sum(cases)) %>%
select(Province_State, date, cases) %>%
ungroup()

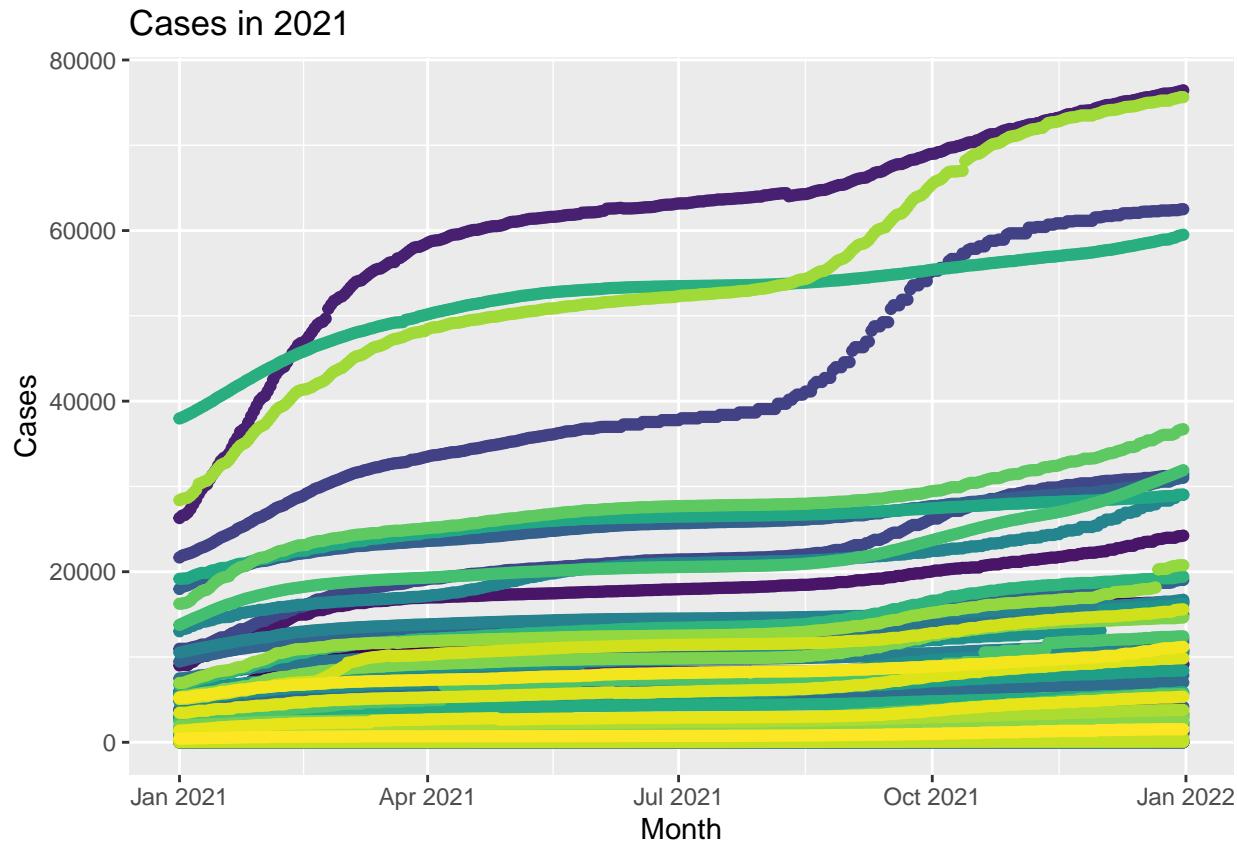
```

I analyzed COVID-19 cases in the United States. When processing the data, I organized it into three groups: 2021, 2022, and data excluding population information.

```

US_cases_2021_by_states %>%
ggplot(aes(x = date, y = cases, color = Province_State)) +
geom_point() +
labs(title = "Cases in 2021", x = "Month", y = "Cases") +
scale_color_viridis_d(guide = "none")

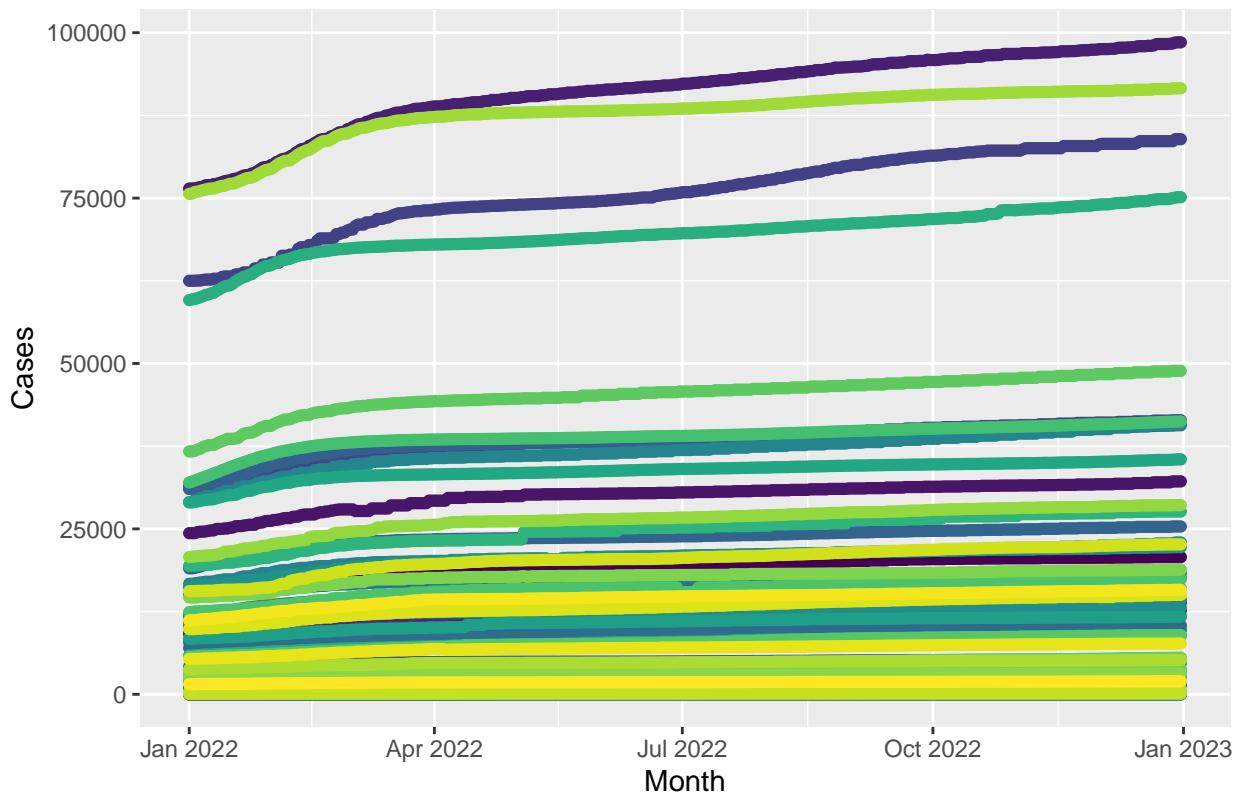
```



This plot displays how cases changed in each state during 2021. We can observe a significant increase over time.

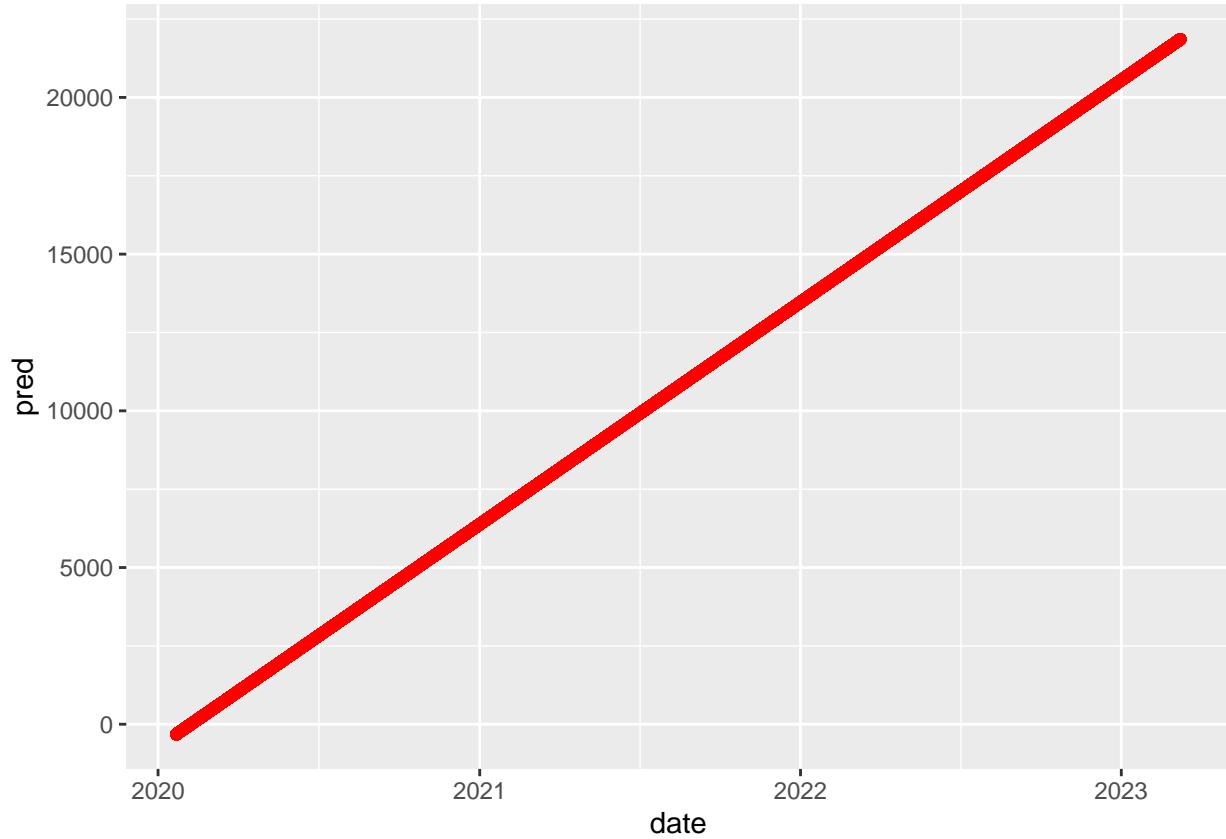
```
US_cases_2022_by_states %>%
  ggplot(aes(x = date, y = cases, color = Province_State)) +
  geom_point() +
  labs(title = "Cases in 2022", x = "Month", y = "Cases") +
  scale_color_viridis_d(guide = "none")
```

Cases in 2022



This plot illustrates the changes in cases for each state in 2022. We can see that the increase was slower compared to 2021.

```
mod <- lm(cases ~ date, data = US_cases_by_states)
US_cases_pred <- US_cases_by_states %>% mutate(pred = predict(mod))
US_cases_pred %>% ggplot() +
  geom_point(aes(x = date, y = pred), color = "red")
```



This model shows the general COVID-19 case trend in the US. We observe a swift rise, possibly influenced by the notable increases in 2020 and 2021.