

COVID-19 Case Data

G. E.

Introduction

John Hopkins University (JHU) Center for Systems Science and Engineering (CSSE) Last Updated: March 10, 2023

This document analyzes COVID mortality.

Dataset Description: This is the data repository for the 2019 Novel Coronavirus Visual Dashboard operated by the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE). Also, Supported by ESRI Living Atlas Team and the Johns Hopkins University Applied Physics Lab (JHU APL).

Setup the R Environment

To setup the R environment, libraries `lubridate` and `tidyverse` were imported.

Import the Data

The dataset was imported via URL, made publicly available on Github by the JHU CSSE.

Dataset URL: “https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/”

```
url_in <- "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/"
file_names <- c("time_series_covid19_confirmed_global.csv",
                "time_series_covid19_deaths_global.csv",
                "time_series_covid19_confirmed_US.csv",
                "time_series_covid19_deaths_US.csv")
urls <- str_c(url_in, file_names) # string concatenate
global_cases <- read_csv(urls[1])
global_deaths <- read_csv(urls[2])
us_cases <- read_csv(urls[3])
us_deaths <- read_csv(urls[4])
uid_lookup_url <- "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/refs/heads/master/csse_covid_19_data/csse_covid_19_historical_data/csse_covid_19_historical_data/"
uid <- read_csv(uid_lookup_url) %>% # uid lookup
  select(-c(Lat, Long_, Combined_Key, code3, iso2, iso3, Admin2))
```

Data Cleaning

```

global_cases <- global_cases %>% # global cases
  pivot_longer(cols = -c("Province/State",
                        "Country/Region", Lat, Long),
              names_to = "date",
              values_to = "cases") %>%
  select(-c(Lat, Long))
global_deaths <- global_deaths %>% # global deaths
  pivot_longer(cols = -c("Province/State",
                        "Country/Region", Lat, Long),
              names_to = "date",
              values_to = "deaths") %>%
  select(-c(Lat, Long))
global <- global_cases %>% # global all
  full_join(global_deaths) %>%
  rename(Country_Region = "Country/Region",
         Province_State = "Province/State") %>%
  mutate(date = mdy(date))
global <- global %>% filter(cases > 0)
global <- global %>%
  unite("Combined_Key",
        c(Province_State, Country_Region),
        sep = ", ",
        na.rm = TRUE,
        remove = FALSE)
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_cases %>% # us cases
  pivot_longer(cols = -(UID:Combined_Key),
              names_to = "date",
              values_to = "cases")

```

```

## # A tibble: 3,819,906 x 13
##       UID iso2 iso3 code3 FIPS Admin2 Province_State Country_Region Lat
##       <dbl> <chr> <chr> <dbl> <dbl> <chr> <chr> <chr> <dbl>
## 1 84001001 US USA 840 1001 Autauga Alabama US 32.5
## 2 84001001 US USA 840 1001 Autauga Alabama US 32.5
## 3 84001001 US USA 840 1001 Autauga Alabama US 32.5
## 4 84001001 US USA 840 1001 Autauga Alabama US 32.5
## 5 84001001 US USA 840 1001 Autauga Alabama US 32.5
## 6 84001001 US USA 840 1001 Autauga Alabama US 32.5
## 7 84001001 US USA 840 1001 Autauga Alabama US 32.5
## 8 84001001 US USA 840 1001 Autauga Alabama US 32.5
## 9 84001001 US USA 840 1001 Autauga Alabama US 32.5
## 10 84001001 US USA 840 1001 Autauga Alabama US 32.5
## # i 3,819,896 more rows
## # i 4 more variables: Long_ <dbl>, Combined_Key <chr>, date <chr>, cases <dbl>

```

```

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 %>% # 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 %>% # us all
  full_join(us_deaths)

```

Visualize and Analyze

Visualizations

```

us_by_state <- us %>% # us state
  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 %>% # us totals
  group_by(Country_Region, date) %>%
  summarise(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()

```

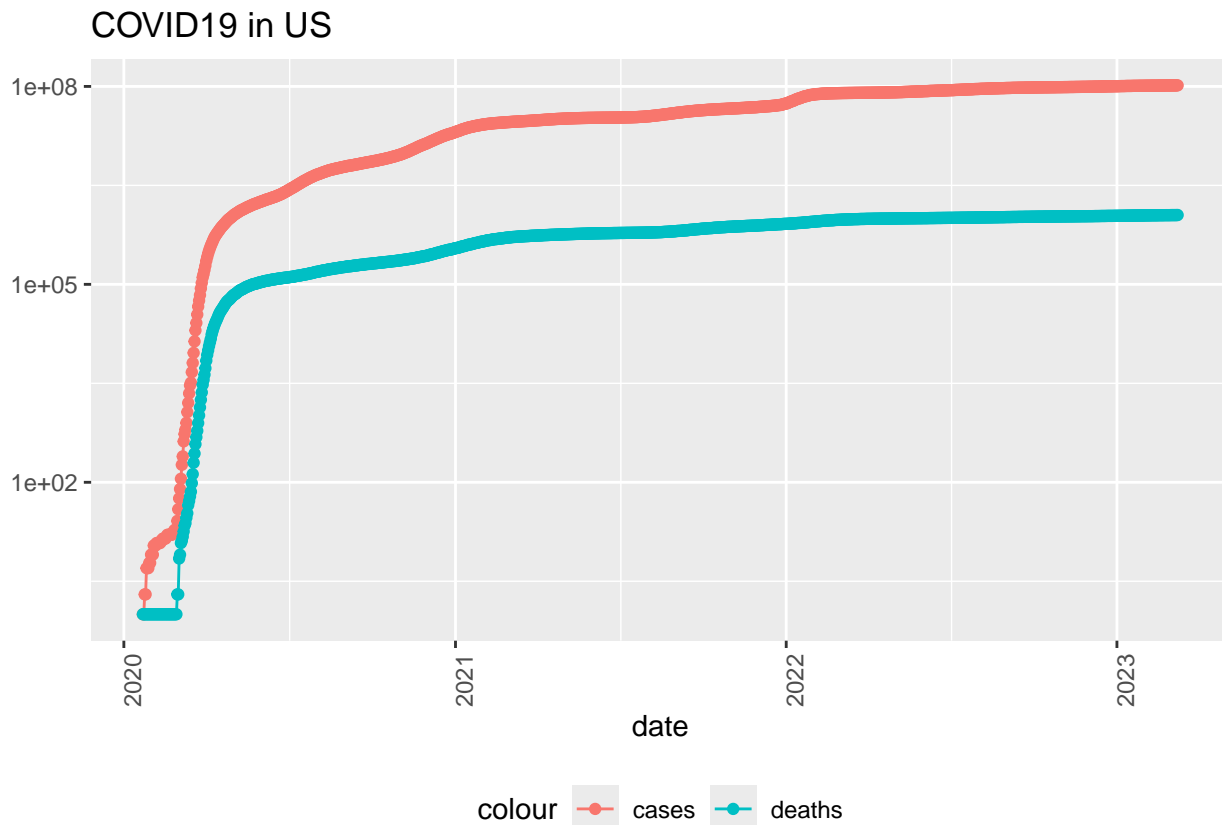
'summarise()' has grouped output by 'Country_Region'. You can override using
the '.groups' argument.

```

us_totals %>% # line graph of us totals
  filter(cases > 0) %>%
  ggplot(aes(x = date, y = cases)) +
  geom_line(aes(color = "cases")) +
  geom_point(aes(color = "cases")) +

```

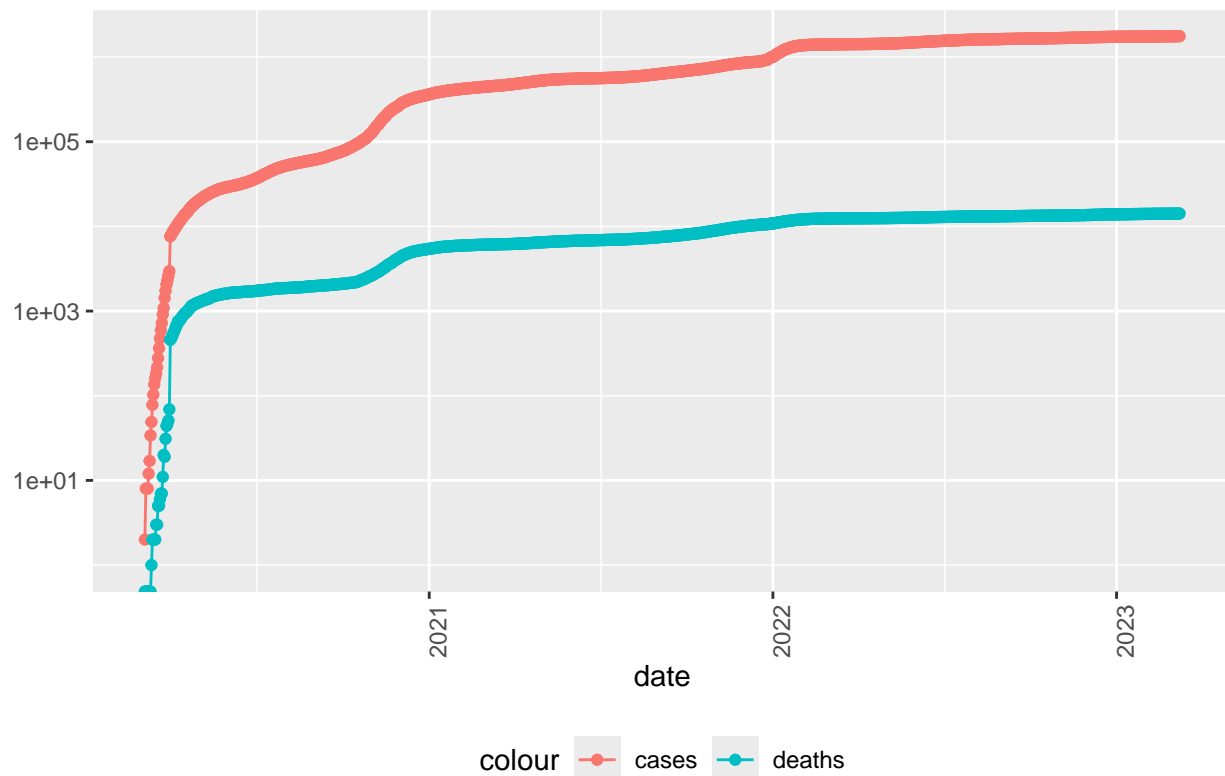
```
geom_line(aes(y = deaths, color = "deaths")) +
geom_point(aes(y = deaths, color = "deaths")) +
scale_y_log10() +
theme(legend.position = "bottom",
      axis.text.x = element_text(angle = 90)) +
labs(title = "COVID19 in US", y = NULL)
```



```
state <- "Colorado"
colo <- us_by_state %>%
  filter(Province_State == state) %>%
  filter(cases > 0)
colo %>%
  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.
```

COVID19 in Colorado



Analysis

```
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))
us_by_state %>%
  filter(Province_State == state) %>%
  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 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 in transformation$transform(x): NaNs produced

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

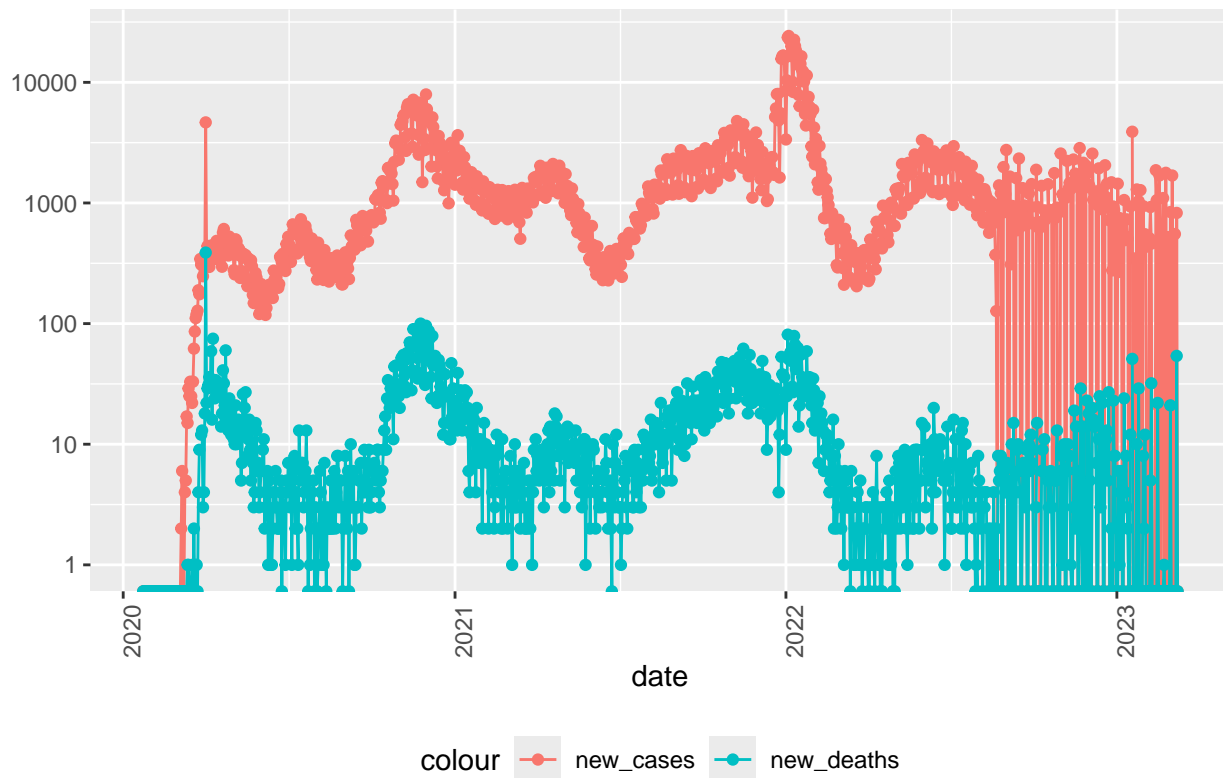
## Warning: Removed 1 row containing missing values or values outside the scale range
## ('geom_line()').

## 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 5 rows containing missing values or values outside the scale range
## ('geom_point()').
```

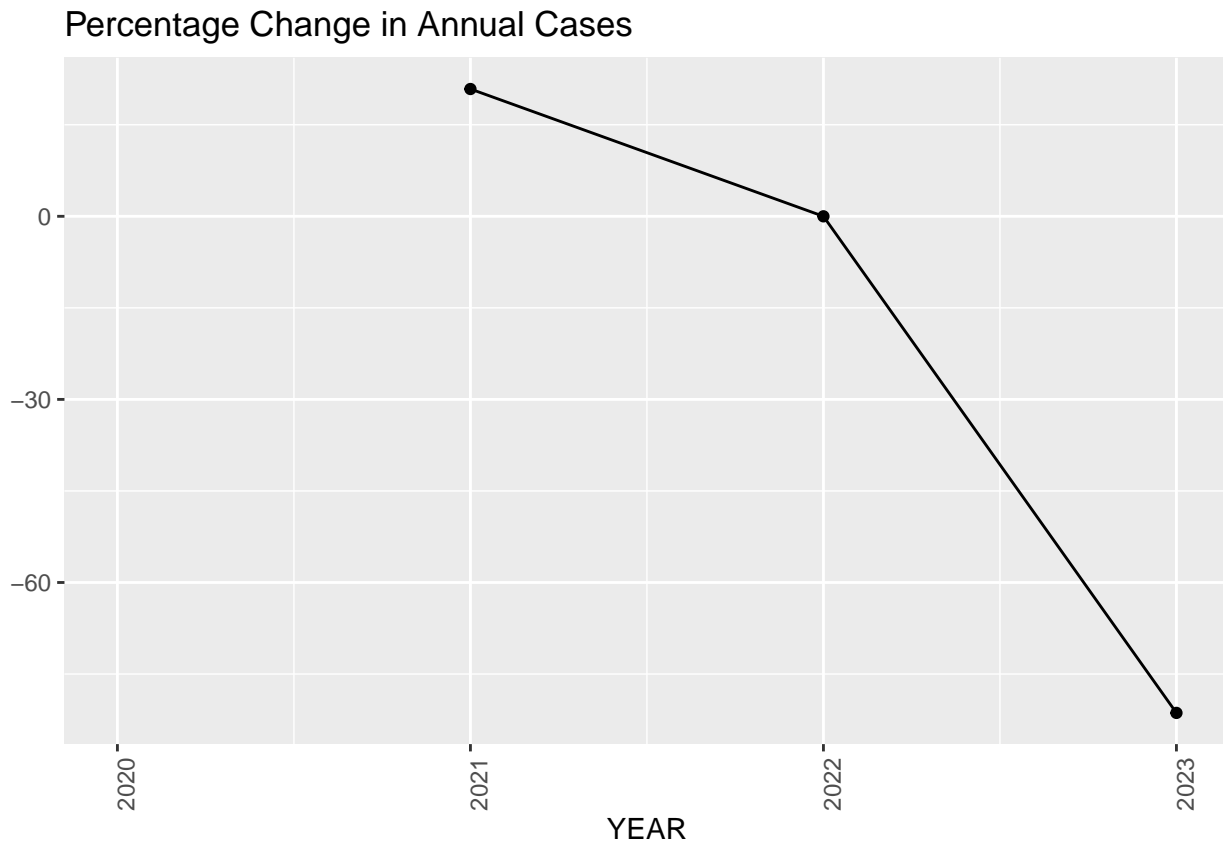
COVID19 in US



```
colo <- colo %>%
  mutate(YEAR = year(date))
colo %>%
  count(YEAR) %>%
  mutate(pct_change = (n - lag(n)) / lag(n) * 100) %>%
  ggplot(aes(x = YEAR, y = pct_change)) +
  geom_line() +
  geom_point() +
  theme(legend.position = "bottom",
        axis.text.x = element_text(angle=90)) +
  labs(title = "Percentage Change in Annual Cases", y = NULL)
```

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

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



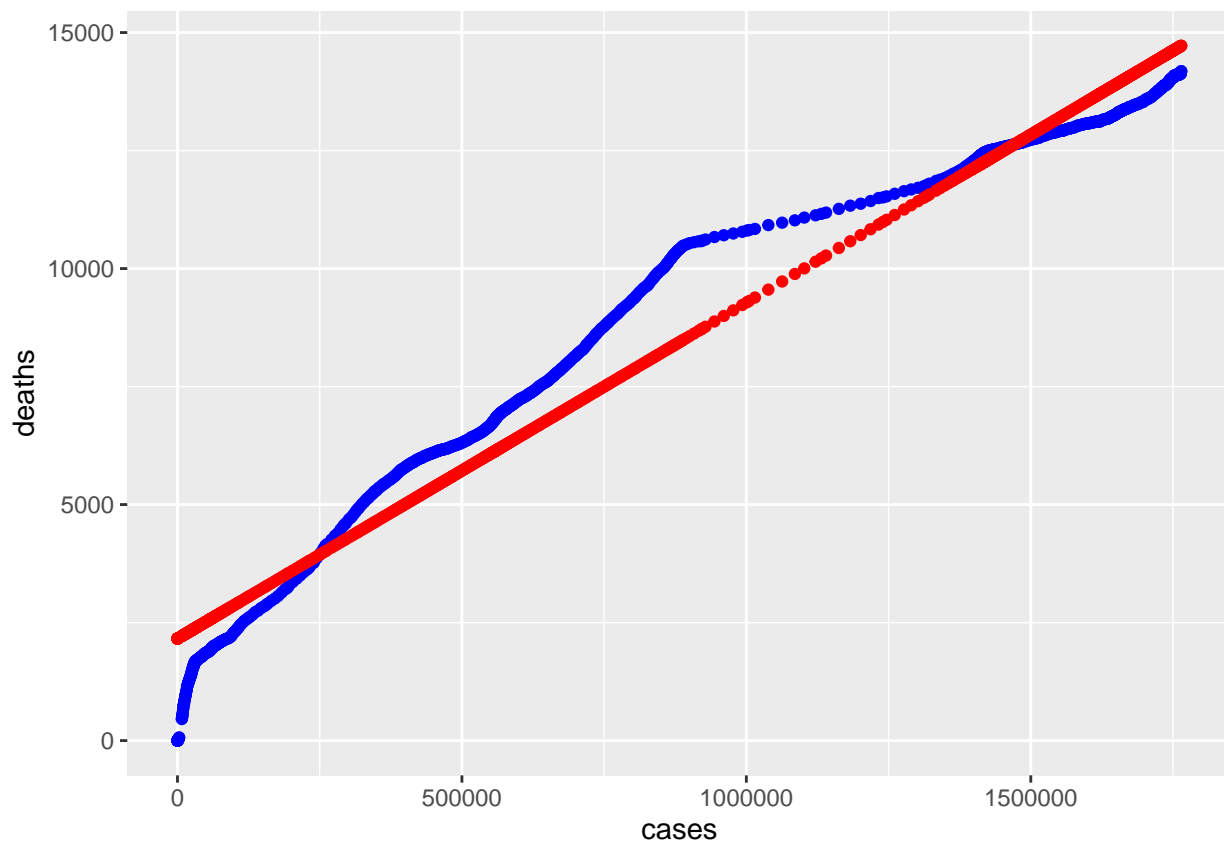
COVID-19 Case rates in Colorado dropped significantly since being tracked.

Modeling

```
mod <- lm(deaths ~ cases, data = colo)
colo %>% mutate(pred = predict(mod))
```

```
## # A tibble: 1,100 x 9
##   Province_State Country_Region date       cases deaths deaths_per_mill
##   <chr>          <chr>      <date>     <dbl>  <dbl>      <dbl>
## 1 Colorado      US        2020-03-05      2      0          0
## 2 Colorado      US        2020-03-06      8      0          0
## 3 Colorado      US        2020-03-07      8      0          0
## 4 Colorado      US        2020-03-08      8      0          0
## 5 Colorado      US        2020-03-09     12      0          0
## 6 Colorado      US        2020-03-10     17      0          0
## 7 Colorado      US        2020-03-11     34      0          0
## 8 Colorado      US        2020-03-12     49      1        0.174
## 9 Colorado      US        2020-03-13     78      2        0.347
## 10 Colorado     US        2020-03-14    103      2        0.347
## # i 1,090 more rows
## # i 3 more variables: Population <dbl>, YEAR <dbl>, pred <dbl>
```

```
colo_w_pred <- colo %>% mutate(pred = predict(mod))
colo_w_pred %>% ggplot() +
  geom_point(aes(x = cases, y = deaths), color = "blue") +
  geom_point(aes(x = cases, y = pred), color = "red")
```



There's an incredibly strong relationship between cases and deaths.

Additional Questions

How did vaccine distribution affect case numbers?

Bias

COVID was pervasive, world-changing, and devastating. Some people were more affected than others. Those who were greatly affected, maybe lost a loved one, may attempt to extract more insight than there is. Whereas someone who was less affected may opt to take a more surface level approach.

Bias Mitigation Techniques

A fundamental starting point is to engage in self-reflection to develop an understanding of personal biases. Approach the problem from a differing point of view, seeking a greater level of understanding and objectivity.

Conclusion

Colorado COVID-19 cases and deaths have an high correlation. Even post-vaccine distribution.

Session Info

```
sessionInfo()
```

```
## R version 4.4.2 (2024-10-31)
## Platform: aarch64-apple-darwin20
## Running under: macOS Sequoia 15.5
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/lib/libRlapack.dylib; LAPACK v
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## time zone: America/Denver
## tzcode source: internal
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] forcats_1.0.0 stringr_1.5.1 dplyr_1.1.4 purrr_1.0.2
## [5] readr_2.1.5   tidyr_1.3.1   tibble_3.3.0 ggplot2_3.5.2
## [9] tidyverse_2.0.0 lubridate_1.9.4
##
## loaded via a namespace (and not attached):
## [1] bit_4.6.0      gtable_0.3.6   jsonlite_1.8.9 crayon_1.5.3
## [5] compiler_4.4.2 tidyselect_1.2.1 parallel_4.4.2 jquerylib_0.1.4
## [9] scales_1.4.0   yaml_2.3.10    fastmap_1.2.0   R6_2.5.1
## [13] labeling_0.4.3 generics_0.1.4  curl_6.4.0      knitr_1.50
## [17] bslib_0.9.0    pillar_1.10.1  RColorBrewer_1.1-3 tzdb_0.5.0
## [21] rlang_1.1.6     utf8_1.2.4     stringi_1.8.4   cachem_1.1.0
```

## [25]	xfun_0.52	sass_0.4.10	bit64_4.6.0-1	timechange_0.3.0
## [29]	cli_3.6.5	withr_3.0.2	magrittr_2.0.3	digest_0.6.37
## [33]	grid_4.4.2	vroom_1.6.5	hms_1.1.3	lifecycle_1.0.4
## [37]	vctrs_0.6.5	evaluate_1.0.3	glue_1.8.0	farver_2.1.2
## [41]	rmarkdown_2.29	tools_4.4.2	pkgconfig_2.0.3	htmltools_0.5.8.1