COVID\_19

A.Pavlenko

2022-06-06

# Coronavirus Disease 2019 (COVID-19)

# Coronavirus disease (COVID-19) is an infectious disease caused by the SARS-CoV-2 virus.

COVID-19 affects different people in different ways. Infected people have had a wide range of symptoms reported – from mild symptoms to severe illness.

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──

## ✔ ggplot2 3.3.6 ✔ purrr 0.3.4  
## ✔ tibble 3.1.7 ✔ dplyr 1.0.9  
## ✔ tidyr 1.2.0 ✔ stringr 1.4.0  
## ✔ readr 2.1.2 ✔ forcats 0.5.1

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

## Download data in 4 files:  
url\_in <- "C:/Users/pavle/Desktop/COVID-19/"  
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)  
urls

## [1] "C:/Users/pavle/Desktop/COVID-19/time\_series\_covid19\_confirmed\_global.csv"  
## [2] "C:/Users/pavle/Desktop/COVID-19/time\_series\_covid19\_deaths\_global.csv"   
## [3] "C:/Users/pavle/Desktop/COVID-19/time\_series\_covid19\_confirmed\_US.csv"   
## [4] "C:/Users/pavle/Desktop/COVID-19/time\_series\_covid19\_deaths\_US.csv"

Reading Data:

global\_cases <- read\_csv(urls[1])

## Rows: 285 Columns: 871  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (2): Province/State, Country/Region  
## dbl (869): 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: 285 Columns: 871  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (2): Province/State, Country/Region  
## dbl (869): 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: 878  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (6): iso2, iso3, Admin2, Province\_State, Country\_Region, Combined\_Key  
## dbl (872): 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: 879  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (6): iso2, iso3, Admin2, Province\_State, Country\_Region, Combined\_Key  
## dbl (873): UID, code3, FIPS, Lat, Long\_, Population, 1/22/20, 1/23/20, 1/24/...  
##   
## 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.

Cleanup Data:

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\_cases

## # A tibble: 247,095 × 4  
## `Province/State` `Country/Region` date cases  
## <chr> <chr> <chr> <dbl>  
## 1 <NA> Afghanistan 1/22/20 0  
## 2 <NA> Afghanistan 1/23/20 0  
## 3 <NA> Afghanistan 1/24/20 0  
## 4 <NA> Afghanistan 1/25/20 0  
## 5 <NA> Afghanistan 1/26/20 0  
## 6 <NA> Afghanistan 1/27/20 0  
## 7 <NA> Afghanistan 1/28/20 0  
## 8 <NA> Afghanistan 1/29/20 0  
## 9 <NA> Afghanistan 1/30/20 0  
## 10 <NA> Afghanistan 1/31/20 0  
## # … with 247,085 more rows

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\_deaths

## # A tibble: 247,095 × 4  
## `Province/State` `Country/Region` date deaths  
## <chr> <chr> <chr> <dbl>  
## 1 <NA> Afghanistan 1/22/20 0  
## 2 <NA> Afghanistan 1/23/20 0  
## 3 <NA> Afghanistan 1/24/20 0  
## 4 <NA> Afghanistan 1/25/20 0  
## 5 <NA> Afghanistan 1/26/20 0  
## 6 <NA> Afghanistan 1/27/20 0  
## 7 <NA> Afghanistan 1/28/20 0  
## 8 <NA> Afghanistan 1/29/20 0  
## 9 <NA> Afghanistan 1/30/20 0  
## 10 <NA> Afghanistan 1/31/20 0  
## # … with 247,085 more rows

library(lubridate)

##   
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':  
##   
## date, intersect, setdiff, union

global <- global\_cases %>%  
 full\_join(global\_deaths) %>%  
 rename(Country\_Region = `Country/Region`,   
 Province\_State = `Province/State`) %>%  
 mutate(date = mdy(date))

## Joining, by = c("Province/State", "Country/Region", "date")

global

## # A tibble: 247,095 × 5  
## Province\_State Country\_Region date cases deaths  
## <chr> <chr> <date> <dbl> <dbl>  
## 1 <NA> Afghanistan 2020-01-22 0 0  
## 2 <NA> Afghanistan 2020-01-23 0 0  
## 3 <NA> Afghanistan 2020-01-24 0 0  
## 4 <NA> Afghanistan 2020-01-25 0 0  
## 5 <NA> Afghanistan 2020-01-26 0 0  
## 6 <NA> Afghanistan 2020-01-27 0 0  
## 7 <NA> Afghanistan 2020-01-28 0 0  
## 8 <NA> Afghanistan 2020-01-29 0 0  
## 9 <NA> Afghanistan 2020-01-30 0 0  
## 10 <NA> Afghanistan 2020-01-31 0 0  
## # … with 247,085 more rows

summary(global)

## Province\_State Country\_Region date cases   
## Length:247095 Length:247095 Min. :2020-01-22 Min. : 0   
## Class :character Class :character 1st Qu.:2020-08-25 1st Qu.: 299   
## Mode :character Mode :character Median :2021-03-30 Median : 6912   
## Mean :2021-03-30 Mean : 588646   
## 3rd Qu.:2021-11-02 3rd Qu.: 129326   
## Max. :2022-06-06 Max. :84882287   
## deaths   
## Min. : 0   
## 1st Qu.: 2   
## Median : 86   
## Mean : 10452   
## 3rd Qu.: 1967   
## Max. :1008857

global <- global %>% filter(cases > 0)  
  
global %>% filter(cases > 28000000)

## # A tibble: 990 × 5  
## Province\_State Country\_Region date cases deaths  
## <chr> <chr> <date> <dbl> <dbl>  
## 1 <NA> Brazil 2022-02-18 28072238 643340  
## 2 <NA> Brazil 2022-02-19 28177367 644195  
## 3 <NA> Brazil 2022-02-20 28218180 644592  
## 4 <NA> Brazil 2022-02-21 28258458 644918  
## 5 <NA> Brazil 2022-02-22 28361951 645735  
## 6 <NA> Brazil 2022-02-23 28493336 646714  
## 7 <NA> Brazil 2022-02-24 28589235 647703  
## 8 <NA> Brazil 2022-02-25 28679671 648496  
## 9 <NA> Brazil 2022-02-26 28749552 649184  
## 10 <NA> Brazil 2022-02-27 28776794 649437  
## # … with 980 more rows

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\_cases

## # A tibble: 2,897,514 × 6  
## Admin2 Province\_State Country\_Region Combined\_Key date cases  
## <chr> <chr> <chr> <chr> <date> <dbl>  
## 1 Autauga Alabama US Autauga, Alabama, US 2020-01-22 0  
## 2 Autauga Alabama US Autauga, Alabama, US 2020-01-23 0  
## 3 Autauga Alabama US Autauga, Alabama, US 2020-01-24 0  
## 4 Autauga Alabama US Autauga, Alabama, US 2020-01-25 0  
## 5 Autauga Alabama US Autauga, Alabama, US 2020-01-26 0  
## 6 Autauga Alabama US Autauga, Alabama, US 2020-01-27 0  
## 7 Autauga Alabama US Autauga, Alabama, US 2020-01-28 0  
## 8 Autauga Alabama US Autauga, Alabama, US 2020-01-29 0  
## 9 Autauga Alabama US Autauga, Alabama, US 2020-01-30 0  
## 10 Autauga Alabama US Autauga, Alabama, US 2020-01-31 0  
## # … with 2,897,504 more rows

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\_deaths

## # A tibble: 2,897,514 × 7  
## Admin2 Province\_State Country\_Region Combined\_Key Population date   
## <chr> <chr> <chr> <chr> <dbl> <date>   
## 1 Autauga Alabama US Autauga, Alabama… 55869 2020-01-22  
## 2 Autauga Alabama US Autauga, Alabama… 55869 2020-01-23  
## 3 Autauga Alabama US Autauga, Alabama… 55869 2020-01-24  
## 4 Autauga Alabama US Autauga, Alabama… 55869 2020-01-25  
## 5 Autauga Alabama US Autauga, Alabama… 55869 2020-01-26  
## 6 Autauga Alabama US Autauga, Alabama… 55869 2020-01-27  
## 7 Autauga Alabama US Autauga, Alabama… 55869 2020-01-28  
## 8 Autauga Alabama US Autauga, Alabama… 55869 2020-01-29  
## 9 Autauga Alabama US Autauga, Alabama… 55869 2020-01-30  
## 10 Autauga Alabama US Autauga, Alabama… 55869 2020-01-31  
## # … with 2,897,504 more rows, and 1 more variable: deaths <dbl>

US <- US\_cases %>%  
 full\_join(US\_deaths)

## Joining, by = c("Admin2", "Province\_State", "Country\_Region", "Combined\_Key",  
## "date")

US

## # A tibble: 2,897,514 × 8  
## Admin2 Province\_State Country\_Region Combined\_Key date cases Population  
## <chr> <chr> <chr> <chr> <date> <dbl> <dbl>  
## 1 Autau… Alabama US Autauga, Al… 2020-01-22 0 55869  
## 2 Autau… Alabama US Autauga, Al… 2020-01-23 0 55869  
## 3 Autau… Alabama US Autauga, Al… 2020-01-24 0 55869  
## 4 Autau… Alabama US Autauga, Al… 2020-01-25 0 55869  
## 5 Autau… Alabama US Autauga, Al… 2020-01-26 0 55869  
## 6 Autau… Alabama US Autauga, Al… 2020-01-27 0 55869  
## 7 Autau… Alabama US Autauga, Al… 2020-01-28 0 55869  
## 8 Autau… Alabama US Autauga, Al… 2020-01-29 0 55869  
## 9 Autau… Alabama US Autauga, Al… 2020-01-30 0 55869  
## 10 Autau… Alabama US Autauga, Al… 2020-01-31 0 55869  
## # … with 2,897,504 more rows, and 1 more variable: deaths <dbl>

global <- global %>%  
 unite("Combined\_Key",  
 c(Province\_State, Country\_Region),  
 sep = ", ",  
 na.rm = TRUE,  
 remove = FALSE)  
global

## # A tibble: 227,505 × 6  
## Combined\_Key Province\_State Country\_Region date cases deaths  
## <chr> <chr> <chr> <date> <dbl> <dbl>  
## 1 Afghanistan <NA> Afghanistan 2020-02-24 5 0  
## 2 Afghanistan <NA> Afghanistan 2020-02-25 5 0  
## 3 Afghanistan <NA> Afghanistan 2020-02-26 5 0  
## 4 Afghanistan <NA> Afghanistan 2020-02-27 5 0  
## 5 Afghanistan <NA> Afghanistan 2020-02-28 5 0  
## 6 Afghanistan <NA> Afghanistan 2020-02-29 5 0  
## 7 Afghanistan <NA> Afghanistan 2020-03-01 5 0  
## 8 Afghanistan <NA> Afghanistan 2020-03-02 5 0  
## 9 Afghanistan <NA> Afghanistan 2020-03-03 5 0  
## 10 Afghanistan <NA> Afghanistan 2020-03-04 5 0  
## # … with 227,495 more rows

uid\_lookup\_url <- "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse\_covid\_19\_data/UID\_ISO\_FIPS\_LookUp\_Table.csv"  
  
uid <- read\_csv(uid\_lookup\_url) %>%  
 select(-c(Lat, Long\_, Combined\_Key, code3, iso2, iso3, Admin2))

## Rows: 4317 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.

url\_in <- "https://github.com/CSSEGISandData/COVID-19/tree/master/csse\_covid\_19\_data/csse\_covid\_19\_time\_series"  
uid <- read\_csv(uid\_lookup\_url) %>%  
 select(-c(Lat, Long\_, Combined\_Key, code3, iso2, iso3, Admin2))

## Rows: 4317 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, Population,  
 Combined\_Key)  
global

## # A tibble: 227,505 × 6  
## Province\_State Country\_Region date cases Population Combined\_Key  
## <chr> <chr> <date> <dbl> <dbl> <chr>   
## 1 <NA> Afghanistan 2020-02-24 5 38928341 Afghanistan   
## 2 <NA> Afghanistan 2020-02-25 5 38928341 Afghanistan   
## 3 <NA> Afghanistan 2020-02-26 5 38928341 Afghanistan   
## 4 <NA> Afghanistan 2020-02-27 5 38928341 Afghanistan   
## 5 <NA> Afghanistan 2020-02-28 5 38928341 Afghanistan   
## 6 <NA> Afghanistan 2020-02-29 5 38928341 Afghanistan   
## 7 <NA> Afghanistan 2020-03-01 5 38928341 Afghanistan   
## 8 <NA> Afghanistan 2020-03-02 5 38928341 Afghanistan   
## 9 <NA> Afghanistan 2020-03-03 5 38928341 Afghanistan   
## 10 <NA> Afghanistan 2020-03-04 5 38928341 Afghanistan   
## # … with 227,495 more rows

US\_by\_state <- US %>%  
 group\_by(Province\_State, Country\_Region, date) %>%  
 summarise(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\_by\_state

## # A tibble: 50,286 × 7  
## Province\_State Country\_Region date cases deaths deaths\_per\_mill  
## <chr> <chr> <date> <dbl> <dbl> <dbl>  
## 1 Alabama US 2020-01-22 0 0 0  
## 2 Alabama US 2020-01-23 0 0 0  
## 3 Alabama US 2020-01-24 0 0 0  
## 4 Alabama US 2020-01-25 0 0 0  
## 5 Alabama US 2020-01-26 0 0 0  
## 6 Alabama US 2020-01-27 0 0 0  
## 7 Alabama US 2020-01-28 0 0 0  
## 8 Alabama US 2020-01-29 0 0 0  
## 9 Alabama US 2020-01-30 0 0 0  
## 10 Alabama US 2020-01-31 0 0 0  
## # … with 50,276 more rows, and 1 more variable: Population <dbl>

US\_totals <- US\_by\_state %>%  
 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.

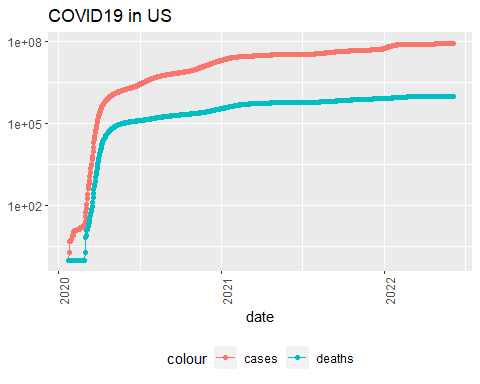
US\_totals

## # A tibble: 867 × 6  
## Country\_Region date cases deaths deaths\_per\_mill Population  
## <chr> <date> <dbl> <dbl> <dbl> <dbl>  
## 1 US 2020-01-22 1 1 0.00300 332875137  
## 2 US 2020-01-23 1 1 0.00300 332875137  
## 3 US 2020-01-24 2 1 0.00300 332875137  
## 4 US 2020-01-25 2 1 0.00300 332875137  
## 5 US 2020-01-26 5 1 0.00300 332875137  
## 6 US 2020-01-27 5 1 0.00300 332875137  
## 7 US 2020-01-28 5 1 0.00300 332875137  
## 8 US 2020-01-29 6 1 0.00300 332875137  
## 9 US 2020-01-30 6 1 0.00300 332875137  
## 10 US 2020-01-31 8 1 0.00300 332875137  
## # … with 857 more rows

tail(US\_totals)

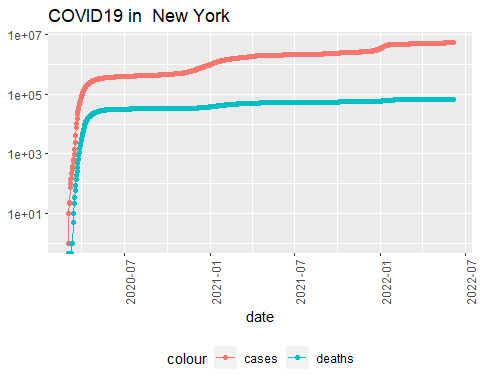
## # A tibble: 6 × 6  
## Country\_Region date cases deaths deaths\_per\_mill Population  
## <chr> <date> <dbl> <dbl> <dbl> <dbl>  
## 1 US 2022-06-01 84451901 1007714 3027. 332875137  
## 2 US 2022-06-02 84570325 1008031 3028. 332875137  
## 3 US 2022-06-03 84724329 1008422 3029. 332875137  
## 4 US 2022-06-04 84748884 1008567 3030. 332875137  
## 5 US 2022-06-05 84762022 1008585 3030. 332875137  
## 6 US 2022-06-06 84882287 1008857 3031. 332875137

## Include Plots:  
  
###COVID19 in US  
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)



###COVID19 in New-York  
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: Transformation introduced infinite values in continuous y-axis  
## Transformation introduced infinite values in continuous y-axis

 ### Analysing Data: max(US\_totals$date)

max(US\_totals$deaths)

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)

tail(US\_totals %>% select(new\_cases, new\_deaths, everything()))

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)

state <- “New York” 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 = str\_c(“COVID19 in”, state), y = NULL)

US\_state\_totals <- US\_by\_state %>% group\_by(Province\_State) %>% summarize(deaths = max(deaths), cases = max(cases), population = max(Population), cases\_per\_thou = 1000 \* cases / population, deaths\_per\_thou = 1000 \* deaths / population) %>% filter(cases > 0, population > 0)

US\_state\_totals %>% slice\_min(deaths\_per\_thou, n = 10)

US\_state\_totals %>% slice\_min(deaths\_per\_thou, n = 10) %>% select(deaths\_per\_thou, cases\_per\_thou, everything())

US\_state\_totals %>% slice\_max(deaths\_per\_thou, n = 10) %>% select(deaths\_per\_thou, cases\_per\_thou, everything())

## Modeling Data:

mod <- lm(deaths\_per\_thou ~ cases\_per\_thou, data = US\_state\_totals) summary(mod)

US\_state\_totals %>% slice\_min(cases\_per\_thou) US\_state\_totals %>% slice\_max(cases\_per\_thou)

x\_grid <- seq(1, 151) new\_df <- tibble(cases\_per\_thou = x\_grid) US\_state\_totals %>% mutate(pred = predict(mod))

US\_tot\_w\_pred <- US\_state\_totals %>% mutate(pred = predict(mod)) US\_tot\_w\_pred

US\_tot\_w\_pred %>% ggplot() + geom\_point(aes(x = cases\_per\_thou, y = deaths\_per\_thou), color = “blue”) + geom\_point(aes(x = cases\_per\_thou, y = pred), color = “red”)